Poverty Targeting with Imperfect Information

Juan C. Yamin*

Brown University

September 22, 2026

Abstract

How should antipoverty programs allocate transfers when household income is known only through noisy predictions? I formulate this as a statistical decision problem in which a policymaker chooses nonnegative transfers within a fixed budget to minimize squared deviations of post-transfer income from the poverty line. I show that the standard plug-in rule, which treats predictions as exact, is inadmissible. I then develop a nonparametric empirical Bayes allocation rule that replaces the estimated gaps with posterior mean poverty gaps. Its Bayes regret is bounded by the mean squared difference between its posterior mean gaps and the oracle's, so the budget and nonnegativity constraints do not slow its convergence to the oracle. The approach extends, with weaker guarantees, to a planner who cares only about how far households remain below the poverty line after transfers and to programs that pay a fixed set of benefit amounts. In simulations using household surveys from nine African countries, the allocations under the empirical Bayes rule reach about 1.8 times as many poor people as those under plug-in OLS with the same budget and achieve the same average poverty gap reduction with 6.7% less spending.

Keywords: Poverty targeting, Empirical Bayes, Statistical decision theory.
JEL: I32, H53, C44, D81

*Brown University, Department of Economics, Robinson Hall, 64 Waterman Street, Providence, RI 02912. Email: juan_yamin_silva@brown.edu. I am grateful for the generous advice and support of Toru Kitagawa, Soonwoo Kwon, and Jonathan Roth. I thank Andrew Chesher, Tim Christensen, Andrew Foster, Peter Hull, Patrick Kline, Eric D. Mbakop, Paul Niehaus, Kirill Ponomarev, Susanne Schennach, Jann Spiess, Stefan Wager, and participants at the 2025 World Congress of the Econometric Society and Brown's Econometric Coffee and Development Tea for their helpful comments. Caitlin Brown, Paul Corral, and Heath Henderson kindly shared their data for the simulation exercises.

1. Introduction

Targeted cash transfers are a central antipoverty instrument in developing countries, yet the governments that run them rarely observe the incomes of the households the transfers are meant to reach (Banerjee, Hanna, Olken, & Sverdlin Lisker, 2024). Targeting these transfers well requires knowing who is poor and how far below the poverty line each household falls. But because widespread informality and limited administrative data make income or consumption difficult to measure reliably for every potential beneficiary, governments predict it from information they can observe. A widely used method, proxy means testing (PMT), predicts consumption from household characteristics such as housing quality, education, and asset ownership. Under geographic targeting, transfers are assigned according to estimated poverty in the area where a household lives (Baker & Grosh, 1994; Elbers, Fujii, Lanjouw, Özler, & Yin, 2007).[1] In both cases, the policymaker chooses recipients and transfer amounts from imperfect estimates of who is poor and by how much.

Targeting performance depends both on how accurately household income is predicted and on how the predictions are used to decide who receives a transfer and how much. Predictions from a basic PMT, for example, miss most of the poor and misclassify many nonpoor households, with average inclusion and exclusion error rates of about 37% and 72% across nine African countries (Brown, Ravallion, & Van de Walle, 2018). Most efforts to reduce these errors have focused on improving the predictions. More flexible machine learning methods can predict more accurately, but the gains in targeting are often modest and vary across settings and performance measures (Baez, Kshirsagar, & Skoufias, 2019; Areias & Wai-Poi, 2022; Corral, Henderson, & Segovia, 2025). The allocation step, by contrast, is usually taken as given, with transfers assigned by a plug-in rule that treats the estimates as if they were exact.

This paper asks how a policymaker should assign transfers when household income is known only through noisy predictions. I formulate the allocation step as a statistical decision problem. The policymaker runs an unconditional cash transfer program, observes an income estimate and its standard error for each household, and chooses

[1] PMT underlies Colombia's Sisbén registry, Pakistan's BISP, and Mexico's Progresa/Oportunidades. Geographic targeting is used by the Productive Safety Net in Ethiopia, Challenging the Frontiers of Poverty Reduction in Bangladesh, and Chile Solidario (Barrientos, 2013).

how much each household receives. Transfers cannot be negative, so the program cannot tax, and total spending cannot exceed a fixed budget. I evaluate rules by the expected average squared distance between post-transfer income and the poverty line. This criterion weights larger remaining poverty gaps more heavily and penalizes transfers that raise income above the line, allowing for administrative, social, or political costs of mistargeting (Devereux et al., 2017; Della Guardia, Lake, & Schnitzer, 2022; Cameron & Shah, 2014). I use it as a tractable benchmark for targeting quality and also study a one-sided objective that counts only the squared poverty gaps remaining after transfers.

I show that the plug-in rule is inadmissible under the quadratic criterion. With Gaussian prediction errors and conditional mean incomes in a bounded range, another feasible rule has expected loss no greater than the plug-in rule's for every income configuration within that range and strictly lower for some. Because the plug-in rule treats each estimate as exact, differences in estimated poverty gaps among recipient households pass directly into differences in transfers, regardless of how precise each estimate is. I show that this behavior is incompatible with a Bayes rule under any proper prior on the bounded parameter space, and a complete-class argument then yields inadmissibility.

To move beyond this negative result, I develop a nonparametric empirical Bayes targeting rule that pools information across households while accounting for the precision of each household's income estimate. The rule approximates an oracle Bayes allocation, which would be optimal if the distribution of conditional mean incomes across households were known. Under the quadratic criterion, the oracle computes each household's posterior mean poverty gap from its income estimate and standard error and then selects the feasible allocation closest to the vector of posterior mean gaps. The allocation step is the same as under the plug-in rule, with posterior mean gaps in place of raw estimates.[2] Households with the same income estimate can therefore receive different transfers when their standard errors differ. In practice, the distribution of conditional mean incomes is unknown, so I estimate it by nonparametric maximum likelihood.

My main theoretical result is that the empirical Bayes rule approaches the oracle's

[2]The same approach applies to other policy problems in which a fixed budget is divided across many units using noisy estimates, such as health care funding distributed across areas according to estimated local demand for care (Smith, Rice, & Carr-Hill, 2001), school funding distributed across districts according to estimated counts of poor children (Zaslavsky & Schirm, 2002), and medical supplies allocated across regions from demand forecasts (X. Chen, Chong, Feng, & Zhang, 2021).

performance as the number of households grows. I evaluate the rule by its Bayes regret, the expected loss incurred under it beyond the loss under the oracle Bayes allocation, which measures the cost of estimating the distribution of conditional mean incomes rather than knowing it. This regret is bounded by the expected average squared difference between the estimated and oracle posterior mean gaps, even though the constraints make the allocation a nonsmooth function of those gaps. For n households, combining this inequality with the heteroskedastic nonparametric maximum likelihood bound of Jiang (2020) gives a regret bound of order $n^{-p/(2(p+1))}$ up to logarithmic factors when the distribution of conditional mean incomes has a finite p-th moment with $p \geq 2$. The same argument converts sharper bounds on posterior mean estimation error into sharper targeting regret bounds, including a bound of order $1/n$ up to logarithmic factors when that distribution has compact support (Jiang & Zhang, 2009; Soloff, Guntuboyina, & Sen, 2025).

In simulations built from household survey data for nine African countries, the empirical Bayes rule reaches about 1.8 times as many poor people as plug-in ordinary least squares (OLS) under the same budget ceiling. Both rules use the same out-of-sample consumption predictions from an OLS proxy means test, so the comparison isolates the allocation step. The empirical Bayes rule also achieves the same average poverty gap reduction as plug-in OLS with 6.7% less transfer spending. It recovers 24.5% of the reduction in quadratic loss that a policymaker who observed consumption would attain relative to no transfers, compared with 14.6% for plug-in OLS and 20.7% for a plug-in rule built on XGBoost, a flexible machine learning predictor. Plug-in OLS concentrates the budget on the households with the largest predicted gaps, which often partly reflect noise, whereas empirical Bayes trims those transfers, extends assistance to more poor households, and spends less on transfers beyond recipients' poverty gaps.

The empirical Bayes approach also applies to the one-sided objective, which does not penalize transfers that raise households' incomes above the poverty line. Up to normalization, this objective is the Foster–Greer–Thorbecke index with parameter $\alpha = 2$, evaluated at conditional mean income in the theoretical analysis. The perfect-information allocation remains optimal, and the plug-in rule is still inadmissible when the total budget exceeds the poverty line but can fall short of the aggregate poverty gap. The oracle Bayes rule, however, is no longer a projection of the posterior mean gaps. It equalizes the posterior expected remaining gaps across recipients, so the empirical Bayes rule must estimate how these gaps vary with the transfer amount. Under bounded support and strict budget scarcity, I establish a weaker regret bound, of order $1/\log n$. In the simulations, the two empirical Bayes rules perform nearly identically in the seven countries where the quadratic rule exhausts the budget. Material differences in the two rules' performance arise in Ethiopia and Mali, where the one-sided rule spends funds that the quadratic rule leaves unallocated and attains a higher gain under the one-sided criterion.

Under the quadratic loss, the empirical Bayes approach also extends to programs that restrict transfers to a fixed menu of benefit amounts. The oracle again selects the feasible allocation closest to the vector of posterior mean poverty gaps, and the empirical Bayes rule substitutes estimated posterior means in the same assignment problem. The discrete feasible set is nonconvex, and assignments can jump between payment levels. With a fixed menu and a budget proportional to the population, I establish a regret upper bound of order $n^{-p/(4(p+1))}$ up to logarithmic factors under the same finite p-th moment condition, slower than the $n^{-p/(2(p+1))}$ rate of the continuous problem.

In the poverty targeting literature, most work predicts income and then applies a plug-in allocation rule (Banerjee et al., 2024). Follett and Henderson (2023) bring community information into beneficiary selection, and Sahoo, Blumenstock, Niehaus, Selker, and Wager (2025) derive poverty-gap-minimizing transfers from conditional consumption quantiles and implement them with estimated quantiles. This paper takes income predictions and their standard errors as given and studies how sampling uncertainty in the predictions should change the transfers.

Statistical decision theory has entered policy design mainly through treatment choice, where the policymaker decides whom to treat, or at what dose, on the basis of estimated treatment effects (Manski, 2004; Hirano & Porter, 2009; Stoye, 2009; Tetenov, 2012; Kitagawa & Tetenov, 2018; Athey & Wager, 2021; Ai, Fang, & Xie, 2024; Kitagawa, Lee, & Qiu, 2026), including under budget constraints (Bhattacharya & Dupas, 2012; Sun, 2026) and with equality-minded welfare criteria (Kitagawa & Tetenov, 2021). In the targeting problem, by contrast, the effect of a transfer on income is known, the uncertainty is about income itself, and the loss measures squared deviations of post-transfer income from the poverty line. The inadmissibility result relates to the statistical literature on estimation under restricted parameter spaces (Charras & van Eeden, 1992; Marchand & Perron, 2001; Hartigan, 2004; Fourdrinier, Strawderman, & Wells, 2018), where the unknown parameter is known to lie in a constrained set and estimators are compared on that set. In the targeting problem, the constraint falls on the action space instead, since the budget and nonnegativity constraints define the set of feasible transfers.

The rule itself builds on the empirical Bayes literature on decisions with noisy estimates, where these methods have served denoising, ranking, and selection problems (Gu & Koenker, 2023; Kline, Rose, & Walters, 2022; Montiel Olea, O'Flaherty, & Sethi, 2021; Chetty, Friedman, & Rockoff, 2014; Chetty & Hendren, 2018; Kwon, 2026). Moon (2025) and Kline, Rose, and Walters (2024) apply them to downstream decisions. J. Chen (2026) studies binary selection rules based on posterior means, including selection of a fixed number of units, and J. Chen, Lei, Sudijono, Sun, and Xie (2025) choose selection rules by maximizing a Stein-type unbiased estimate of their welfare. The single-grant case of the fixed-menu extension is a budget-capped relative of these problems. The regret results draw on the Bayes regret literature for empirical Bayes estimators (Kiefer & Wolfowitz, 1956; Zhang, 2009; Jiang & Zhang, 2009; Saha & Guntuboyina, 2020; Jiang, 2020; Soloff et al., 2025; J. Chen, 2026).

The paper is organized as follows. Section 2 introduces the poverty targeting model, starting from a perfect-information benchmark and extending it to noisy income estimates. Section 3 establishes the inadmissibility of the plug-in rule. Section 4 presents the empirical Bayes framework, characterizes the oracle and feasible rules, and derives regret bounds. Section 5 reports the simulation evidence. Section 6 concludes. Appendix B develops the one-sided extension and Appendix C the fixed-menu extension.

2. Poverty Targeting Model

When household incomes are observed, the optimal allocation raises the poorest households' incomes to a common post-transfer level at or below the poverty line and gives nothing to the others. When income predictions are noisy, however, the policymaker is uncertain about both who should receive the transfers and how far each household falls below the line. Targeting therefore becomes a statistical decision problem.

2.1 Perfect Information

I consider a policymaker who allocates a fixed antipoverty budget across households through nonnegative transfers to reduce poverty, giving priority to the households furthest below the poverty line. In this benchmark, the policymaker observes household incomes without error.[3]

2.1.1 Setup

The policymaker observes household incomes $y = (y_1, \ldots, y_n)^\top$, where $y_i \geq 0$ is the realized income of household i. Throughout, income stands for the scalar resource measure that the program targets, such as per capita income or consumption. The budget $B > 0$ is taken as given, its financing is not modeled, and it may be insufficient to eliminate poverty. Transfers $\tau = (\tau_1, \ldots, \tau_n)^\top$ satisfy $\tau_i \geq 0$ for all i, so the policymaker cannot tax, and $\sum_{i=1}^{n} \tau_i \leq B$. Let $z > 0$ denote a common poverty line. In the benchmark, I treat each household as one unit and abstract from differences in household size, so household i's post-transfer income is $y_i + \tau_i$.

Notation. Vectors are columns, $\mathbf{1}$ is the n-vector of ones, and $\|\cdot\|$ is the Euclidean norm. For a scalar x, $(x)_+ := \max\{x, 0\}$, and for a vector, the positive part is taken coordinatewise. Hats denote objects estimated from the data. I use uppercase letters for observed random vectors, such as the noisy signals $\hat{Y}$, and lowercase letters for their realizations. Latent household quantities are also written in lowercase.

[3] Measuring household resources is challenging in practice. Survey-based income and consumption measures are sensitive to design choices, and poverty status can vary within the year (Beegle, De Weerdt, Friedman, & Gibson, 2012; Merfeld & Morduch, 2023). I abstract from these issues by taking a scalar income measure, defined by a fixed reference period and accounting protocol, as the latent policy object and studying how noisy signals of it should be used for targeting.

2.1.2 Targeting Objective

The policymaker ranks feasible transfer schedules τ given the pre-transfer income profile y, the poverty line z, and the budget B. The budget constraint incorporates fiscal scarcity, since a dollar transferred to one household is a dollar unavailable for others. The main analysis uses the quadratic loss to account for remaining poverty gaps and nonfiscal costs of mistargeting. For household i,

$$\ell_i(\tau_i, y_i) := (z - y_i - \tau_i)^2 = (z - y_i - \tau_i)_+^2 + (y_i + \tau_i - z)_+^2. \tag{1}$$

The first term is the squared poverty gap that remains, and on its own, it is the one-sided loss $\ell_i^+(\tau_i, y_i) := (z - y_i - \tau_i)_+^2$. Divided by z^2, it is household i's contribution to the Foster–Greer–Thorbecke (FGT) index with parameter $\alpha = 2$ (Foster, Greer, & Thorbecke, 1984), a standard measure of poverty severity in empirical work (Ravallion, 2016). The index is convex in the transfer, places greater weight on larger shortfalls, and satisfies focus, monotonicity, and the transfer principle.[4]

The second term is a penalty on post-transfer income above the poverty line, a reduced-form representation of the nonfiscal targeting costs reviewed by Devereux et al. (2017). These costs are distinct from the opportunity cost of funds, which the budget constraint already captures. The loss is a program objective for comparing transfer schedules and not a social welfare function. Pre-transfer incomes are fixed and transfers are nonnegative, so no feasible allocation lowers any household's income below its pre-transfer level. The second term therefore only discourages transfers that push a household above the poverty line or further above it.[5]

In Indonesia, Cameron and Shah (2014) find that leakage of a cash transfer to ineligible households is associated with lower participation in community groups and higher crime, while undercoverage of eligible households shows no significant associ-

[4] In the FGT class, focus follows from $(\cdot)_+$, monotonicity holds for any positive exponent, and the transfer principle for any exponent above one; see Sen (1976) and Foster et al. (1984). In the perfect-information problem, the optimal allocation is the same for every exponent above one, so the choice of the squared gap is without loss for that benchmark.

[5] A household's being above the poverty line before the program is not itself a targeting error. The loss it contributes when it receives nothing, $(y_i - z)^2$, is the same under every transfer schedule. Subtracting this baseline leaves comparisons between allocations and between decision rules unchanged. What matters for targeting is the additional loss caused by a transfer. Larger payments incur larger penalties, especially when the recipient is already far above the poverty line.

ation with either outcome. This asymmetry matches the structure of (1), which prices nonfiscal costs on transfers above the poverty line but not on gaps left unclosed. Beneficiary selection more broadly affects satisfaction, perceived legitimacy, and social cohesion (Alatas, Banerjee, Hanna, Olken, & Tobias, 2012; Premand & Schnitzer, 2021; Della Guardia et al., 2022). This evidence motivates pricing nonfiscal costs of mistargeting but does not identify their functional form. The quadratic loss is a tractable benchmark that delivers the paper's main ideas in their simplest setting.

Under perfect information, the allocation that minimizes the quadratic loss gives no household more than its poverty gap, and it also minimizes the one-sided loss. Decisions based on noisy estimates can instead exclude households that would receive transfers under perfect information or include nonpoor households. Both errors leave more deprivation unaddressed for a given budget, but only the quadratic loss penalizes a transfer to a nonpoor household beyond the budget it absorbs, so the two losses can rank transfer rules differently. In the main text, I focus on the quadratic loss, for which I derive an empirical Bayes rule whose Bayes regret converges to zero at a polynomial rate. In Appendix B, I consider an extension to the one-sided loss and show that Bayes regret still converges to zero, but at a slower rate.

2.1.3 Optimal Allocation

Let $\mathcal{A} := \{\tau \in \mathbb{R}^n_+ : \mathbf{1}^\top \tau \leq B\}$ denote the set of feasible transfer schedules, and let $P_{\mathcal{A}}(u) := \arg\min_{\tau \in \mathcal{A}} \|u - \tau\|^2$ denote the Euclidean projection onto $\mathcal{A}$. Under perfect information, the policymaker chooses transfers to minimize the average program loss,

$$\tau^{\mathrm{PI}}(y) := \arg\min_{\tau \in \mathcal{A}} \frac{1}{n} \sum_{i=1}^{n} \ell_i(\tau_i, y_i). \tag{2}$$

In many applications, the budget is not large enough to eliminate all pre-transfer poverty gaps, that is, $B < \sum_{i=1}^{n} (z - y_i)_+$, so the policymaker must prioritize who receives transfers and in what amounts.

Proposition 2.1 gives three equivalent representations of the perfect-information benchmark. The leveling-up allocation follows Bourguignon and Fields (1990, 1997), and the Karush–Kuhn–Tucker (KKT) conditions give its threshold form. The projection representation is the one used in the rest of the paper, where the poverty gaps are replaced by estimated or posterior mean gaps. Appendix A collects the proofs of the main-text propositions and the supporting lemmas.

Proposition 2.1 (Perfect-Information Allocation): *Problem* (2) *has a unique solution,*

$$\tau^{\mathrm{PI}}(y) = P_{\mathcal{A}}(z\mathbf{1} - y).$$

Its household-level transfers are

$$\tau_i^{\mathrm{PI}}(y) = \big(z - y_i - \lambda^{\mathrm{PI}}\big)_+, \qquad i = 1, \ldots, n,$$

where $\lambda^{\mathrm{PI}} \geq 0$*. If* $\sum_{i=1}^{n}(z - y_i)_+ \leq B$*, then* $\lambda^{\mathrm{PI}} = 0$*. Otherwise,* $\lambda^{\mathrm{PI}} > 0$ *is the unique solution of* $\sum_{i=1}^{n}\big(z - y_i - \lambda^{\mathrm{PI}}\big)_+ = B$*. This allocation raises all households with incomes below* $z - \lambda^{\mathrm{PI}}$ *to this common income level and gives nothing to the others.*

The optimal rule applies a common cutoff λ^{PI} to households' poverty gaps.[6] Households with gaps at or below the cutoff receive nothing, while those with larger gaps receive the excess. All recipients therefore reach the same post-transfer income, $z - \lambda^{\mathrm{PI}}$. This allocation levels the bottom of the income distribution upward. It raises the poorest household toward the income of the next poorest, then raises both together, and continues until the budget is exhausted or the poverty line is reached.

The budget determines how high this common income level can rise. If it suffices to close every poverty gap, $\lambda^{\mathrm{PI}} = 0$, and every poor household reaches the poverty line. Otherwise, the cutoff is positive and measures the gap left to each recipient. Increasing the budget lowers the cutoff and raises the common income level until the poverty line is reached.

The projection representation expresses the same allocation as the feasible schedule closest to $z\mathbf{1} - y$ in Euclidean distance. Setting transfers equal to this vector would bring every household exactly to the poverty line, but the schedule is infeasible whenever some household is above the line or closing every gap costs more than B. Projection onto $\mathcal{A}$ sets to zero the entries of $z\mathbf{1} - y$ below the cutoff and subtracts the cutoff from the rest.

[6] Cutoffs are expressed in income units. For the average-loss objective with total budget B, the budget multiplier equals $2\lambda^{\mathrm{PI}}/n$.

Figure D.1 shows this geometry for two households. Under the incomes y, the households are similarly poor, and the budget is large enough to close the difference between them and then raise both together. Each receives a transfer, and both are left with the same remaining gap λ^{PI}. Under y', household 1 is so much poorer that even the whole budget would leave it below household 2. It therefore receives everything, the projection lands on the vertex $(B, 0)$, and household 2 receives nothing despite being poor.

2.2 Noisy Measurements

The allocation of Proposition 2.1 requires the income vector y, which the policymaker typically does not observe. What is available instead is a prediction of each household's income from observable characteristics W_i, such as assets, housing quality, or location, together with its standard error. The predictions come from a model estimated on separate survey data, as in PMT or geographic targeting. I take them as given and study how the policymaker should use them to allocate the budget.

The policymaker observes the prediction vector $\hat{Y} = (\hat{Y}_1, \ldots, \hat{Y}_n)^\top$ and its standard errors $\sigma = (\sigma_1, \ldots, \sigma_n)^\top$. Each $\hat{Y}_i$ is a noisy estimate of the conditional mean income $\mu_i = \mathbb{E}[y_i \mid W_i]$, the part of household i's income that its characteristics predict. I assume that the estimates are normally distributed around μ_i, with known variances bounded away from zero and infinity.[7]

Assumption 2.1 (Normality with Bounded Variance) *Conditional on (μ, σ), the income estimates $\hat{Y}$ are independent across households and satisfy*

$$\hat{Y}_i \mid \mu_i, \sigma_i \overset{\text{ind.}}{\sim} \mathcal{N}(\mu_i, \sigma_i^2), \qquad i = 1, \ldots, n,$$

where $\mu_i = \mathbb{E}[y_i \mid W_i] \in \mathbb{R}_+$ and $\sigma_i^2 \in \mathbb{R}_{++}$ is known and treated as fixed. Furthermore, for all $i = 1, \ldots, n$, $0 < \sigma_{\min}^2 \leq \sigma_i^2 \leq \sigma_{\max}^2$.

[7]The model abstracts from systematic prediction bias, which can arise, for example, when the predictions and the income measure refer to different seasons. For allocations obtained by projecting predicted gaps, bias changes transfers by at most its Euclidean norm. A bias common to all households leaves transfers unchanged if the budget binds both with and without it. The simulations in Section 5 do not impose the signal model and still show gains from empirical Bayes over the plug-in rule.

Remarks 2.1 and 2.2 discuss why normality is a reasonable approximation under geographic targeting and PMT.

Remark 2.1 Geographic Targeting. Geographic targeting allocates the budget across areas, such as villages or municipalities, rather than across households. In the benchmark case, every household in area j receives the same transfer τ_j. Let N_j denote the number of households in area j and μ_j their mean income. The policymaker chooses $\tau_1, \ldots, \tau_J$ to minimize $\sum_j N_j(z - \mu_j - \tau_j)^2$ subject to $\tau_j \geq 0$ and $\sum_j N_j \tau_j \leq B$. The weights appear in both the loss and the budget because a transfer to a larger area reaches more households and costs more. With areas in place of households and these weights, the problem has the same structure as the household-level one. In particular, the optimal allocation keeps the threshold form of Proposition 2.1, $\tau_j = (z - \mu_j - \lambda)_+$, where $\lambda = 0$ if $\sum_j N_j(z - \mu_j)_+ \leq B$ and otherwise $\lambda > 0$ is such that $\sum_j N_j \tau_j = B$.

The policymaker does not observe μ_j but estimates it from a survey of the area. If $\hat{Y}_j$ is the survey mean, the central limit theorem makes it approximately normal around μ_j for large area samples under standard sampling conditions, and independent sampling across areas makes the estimates independent. Small-area estimates that combine survey and census data, as in Elbers, Lanjouw, and Lanjouw (2003), rely on a shared fitted model, so their errors can be correlated across areas.

Remark 2.2 Proxy Means Testing. Under PMT, the policymaker estimates an income regression on an auxiliary survey $\{(\tilde{y}_j, \tilde{W}_j)\}_{j=1}^m$ and applies the estimated coefficients to the target households,

$$\tilde{y}_j = \tilde{W}_j'\beta + \tilde{\epsilon}_j, \qquad \hat{Y}_i = W_i'\hat{\beta}, \tag{3}$$

where $\hat{\beta}$ is the OLS estimator. If the covariates take finitely many values and the regression is saturated, $\hat{Y}_i$ is the survey mean income of households with the same observable type as i. Targeting then proceeds by type, the normal approximation of Remark 2.1 applies to each type mean, and the means of different types are independent because they use disjoint survey observations.

In general, the approximation comes from the asymptotic normality of $\hat{\beta}$. If the linear model is correctly specified for the population from which both samples are drawn, then $\mu_i = W_i'\beta$, and under standard moment and rank conditions, $\hat{Y}_i$ is approximately normal around μ_i with variance $\sigma_i^2 = W_i' \operatorname{Var}(\hat{\beta} \mid \tilde{W}_1, \ldots, \tilde{W}_m) W_i$. This variance measures the uncertainty in the estimated conditional mean. Because every prediction shares the estimation error in $\hat{\beta}$, prediction errors are correlated across households, and the correlations do not vanish as the auxiliary survey grows, since variances and covariances are both of order $1/m$. Independence in Assumption 2.1 is therefore a simplification rather than a consequence of OLS asymptotics.

The unknown state is the vector of conditional mean incomes $\mu = (\mu_1, \ldots, \mu_n)^\top$, with parameter space $\mathcal{M} := \mathbb{R}^n_+$. After observing $\hat{y}$, the policymaker chooses a transfer schedule $\tau \in \mathcal{A}$. The program objective, however, is defined at realized incomes, which differ from the state by the residuals $\varepsilon_i := y_i - \mu_i$. Assume that these residuals have finite second moments and mean zero conditional on the policymaker's information. Remark 2.3 shows that, under quadratic loss, the expected loss at realized incomes then differs from that at conditional mean incomes by a term that does not depend on the transfer rule, so the two criteria rank rules identically. I therefore evaluate transfers at state μ using the loss

$$L(\tau, \mu) := \frac{1}{n} \sum_{i=1}^{n} \big(z - \mu_i - \tau_i\big)^2.$$

A statistical decision rule is a measurable mapping $\delta : \mathbb{R}^n \to \mathcal{A}$ that assigns a feasible transfer schedule $\delta(\hat{y})$ to each observed prediction vector $\hat{y}$. Unlike in a standard estimation problem, the action space differs from the parameter space, since the policymaker chooses transfers in $\mathcal{A}$ while the unknown state is a vector of incomes in $\mathcal{M}$. Let $\mathcal{D}$ denote the class of nonrandomized rules, which is essentially complete because the loss and $\mathcal{A}$ are convex (Ferguson, 1967). The frequentist risk of $\delta \in \mathcal{D}$ at state μ is

$$R(\delta, \mu) := \mathbb{E}_\mu\big[L\big(\delta(\hat{Y}), \mu\big)\big],$$

where the expectation is over the predictions $\hat{Y}$ under state μ.

Remark 2.3 Realized versus Predictable Income. Fix (μ, σ), and let $\mathbb{E}_{\mu,\varepsilon}$ denote the expectation over prediction noise and residual income variation at that state. Let $\mathcal{I}$ denote the pre-transfer information generated by the target-household covariates and the independent auxiliary survey used to estimate the prediction rule. The objects $\hat{Y}$, σ, μ, and hence $\delta(\hat{Y})$, are $\mathcal{I}$-measurable. At each state, the assumption on the residuals gives $\mathbb{E}_{\mu,\varepsilon}[\varepsilon_i \mid \mathcal{I}] = 0$ and $\mathbb{E}_{\mu,\varepsilon}[\varepsilon_i^2] < \infty$, where $\varepsilon_i = y_i - \mu_i$.

Since the cross term vanishes after conditioning on $\mathcal{I}$,

$$\mathbb{E}_{\mu,\varepsilon}\Big[(z - y_i - \delta_i(\hat{Y}))^2\Big] = \mathbb{E}_{\mu}\Big[(z - \mu_i - \delta_i(\hat{Y}))^2\Big] + \mathbb{E}_{\mu,\varepsilon}[\varepsilon_i^2].$$

Averaging over households yields

$$\mathbb{E}_{\mu,\varepsilon}\Big[L\big(\delta(\hat{Y}), y\big)\Big] = R(\delta,\mu) + \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\mu,\varepsilon}[\varepsilon_i^2]. \tag{4}$$

The additional term in (4) does not depend on the transfer rule. Thus, expected loss at realized incomes and $R(\delta,\mu)$ rank transfer rules identically at every state.[8]

3. Inadmissibility of Plug-In Rule

The plug-in rule allocates transfers as if the estimated poverty gaps were exact. This section shows that, on a compact set of conditional mean incomes, this rule is inadmissible under quadratic loss. Some other feasible rule has risk no greater at every state in that set and strictly lower risk at some state.

3.1 Poverty Gaps and the Plug-In Rule

Expressing the problem in poverty gaps gives a heteroskedastic Gaussian location model. Define the observed poverty gap vector $X := z\mathbf{1} - \hat{Y}$ and the predictable poverty gap vector $\theta := z\mathbf{1} - \mu$.[9] Under Assumption 2.1,

$$X_i \mid \theta, \sigma \overset{\text{ind.}}{\sim} N(\theta_i, \sigma_i^2), \qquad i = 1, \ldots, n. \tag{5}$$

Since conditional mean incomes are nonnegative, the parameter space becomes $\Theta := (-\infty, z]^n$. For a household with predictable income below the poverty line, θ_i is its shortfall from the line, and for one above the line, $-\theta_i$ is its excess over it.

[8] For nonquadratic losses, averaging over residual income variation generally produces a correction that depends on the transfer, and expected loss at realized and at conditional mean incomes can then rank rules differently.

[9] The predictable gap θ_i differs from the realized gap $z - y_i$ by the residual ε_i. By Remark 2.3, this difference does not affect risk comparisons.

Transfers remain in $\mathcal{A}$, and the loss becomes $L(\tau, \mu) = \frac{1}{n} \sum_{i=1}^{n} (\theta_i - \tau_i)^2$, the average squared difference between transfers and gaps. Since $\hat{Y} = z\mathbf{1} - X$, observing the gap vector is equivalent to observing the income predictions. Every rule $\delta \in \mathcal{D}$ therefore has a counterpart $x \mapsto \delta(z\mathbf{1} - x)$ in the gap formulation that assigns exactly the same transfers. The two rules have identical risk at the respective states μ and $\theta = z\mathbf{1} - \mu$. The remainder of this section works in the gap formulation.

A plug-in rule replaces unknown parameters by their estimates when choosing an action (Hirano & Porter, 2020). If the predictable gap vector θ were known, the optimal allocation would be $P_{\mathcal{A}}(\theta)$ by Proposition 2.1. The plug-in rule applies the same projection to the observed gap vector X, treating it as if it were the true gap vector, $\tau^{\text{plug}}(X) := P_{\mathcal{A}}(X)$. Like the perfect-information allocation, it has a threshold form, $\tau_i^{\text{plug}}(X) = (X_i - \lambda(X))_+$ for $i = 1, \ldots, n$, where $\lambda(X) = 0$ if $\mathbf{1}^\top X_+ \leq B$ and otherwise $\lambda(X) > 0$ solves $\sum_{i=1}^{n} (X_i - \lambda(X))_+ = B$. Households whose predicted income exceeds the poverty line have $X_i < 0$ and receive nothing. The cutoff now depends on the data, and it is positive exactly when the observed poverty gaps cost more than the budget.

When the observed gaps cost more than the budget, the plug-in rule spends all of it, and its threshold form determines how the budget is divided. Each recipient receives a transfer equal to its observed gap minus the common cutoff $\lambda(X)$, so for any two recipients i and j,

$$\tau_i^{\text{plug}}(X) - \tau_j^{\text{plug}}(X) = X_i - X_j.$$

Differences in observed gaps pass through to transfers one for one: a recipient whose gap is estimated to be one dollar larger than another's receives one dollar more. The rule does not use the standard errors σ, so this holds whether the gaps are estimated precisely or not.

3.2 Bayes Rules

The inadmissibility argument compares the plug-in rule with Bayes rules, which minimize expected risk under a prior over the gap vector. Consider a proper prior π on Θ with finite second moment, $\int_\Theta \|\theta\|^2 \, d\pi(\theta) < \infty$, with the noise levels σ known and held fixed. The Bayes risk of a rule $\delta \in \mathcal{D}$ is the expectation of its risk when θ is drawn from π, $r(\pi, \delta) := \int_\Theta R(\delta, \theta) \, d\pi(\theta)$. A rule that minimizes Bayes risk is a Bayes rule (Berger, 1985).

Given an observation x, the prior and the Gaussian model (5) determine the posterior distribution of the gap vector. Under quadratic loss, its mean, $m_\pi(x) := \mathbb{E}_\pi[\theta \mid X = x]$, determines the optimal allocation.

Lemma 3.1 (Bayes rules): *Under Assumption 2.1, for a proper prior π on Θ with finite second moment, the rule $\delta_\pi(x) := P_{\mathcal{A}}(m_\pi(x))$ is a Bayes rule in $\mathcal{D}$ and uniquely minimizes the posterior expected loss at every $x \in \mathbb{R}^n$. Every Bayes rule in $\mathcal{D}$ agrees with δ_π for Lebesgue-almost every x.*

The lemma keeps the projection step of the plug-in rule and changes its input, from the observed gaps x to the posterior mean gaps $m_\pi(x)$. The reason is that, under quadratic loss, the posterior expected loss equals the loss evaluated at these means plus the average posterior variance of the gaps across the n households. The variance term does not depend on the transfers, so it cannot affect which feasible allocation minimizes loss. By Proposition 2.1, the allocation again has the cutoff form $\delta_{\pi,i}(x) = (m_{\pi,i}(x) - \lambda_\pi(x))_+$, with $\lambda_\pi(x) \geq 0$ determined by the same budget condition applied to $m_\pi(x)$. Consequently, for any two recipients i and j, differences in posterior mean gaps pass through to transfers one for one, $\delta_{\pi,i}(x) - \delta_{\pi,j}(x) = m_{\pi,i}(x) - m_{\pi,j}(x)$.

3.3 Dominance Results

So far, the only restriction on a household's predictable income is that it be nonnegative. For the inadmissibility result, I also impose a finite upper bound, which can be chosen arbitrarily large. With an upper bound $z + M$, where $M > 0$, predictable incomes in $[0, z + M]$ correspond to gaps in $[-M, z]$, giving the compact parameter space in Assumption 3.1.

Assumption 3.1 (Bounded parameter space) *Let $M \in (0, \infty)$ be fixed, and define the compact parameter space $\Theta_M := [-M, z]^n \subset \Theta$.*

The next proposition shows that, on this parameter space, the plug-in rule is dominated by another feasible rule.

Proposition 3.1 (Inadmissibility of the plug-in rule): *Assume $n \geq 2$. Under Assumptions 2.1 and 3.1, the plug-in rule $\tau^{\text{plug}}(x) = P_{\mathcal{A}}(x)$ is inadmissible in the class of measurable nonrandomized decision rules $\mathcal{D}$. Equivalently, there exists a rule $\tilde{\delta} \in \mathcal{D}$ such that*

$$R(\tilde{\delta}, \theta) \leq R(\tau^{\text{plug}}, \theta) \qquad \textit{for all } \theta \in \Theta_M,$$

with strict inequality for at least one $\theta \in \Theta_M$.

The proposition identifies an avoidable cost of the plug-in rule. With the same predictions and budget, another feasible rule lowers risk at some states without raising it anywhere on Θ_M. The improvement comes from using the available information differently.

The proof asks whether any prior over the gap vector makes the plug-in allocation a Bayes rule. On the compact set Θ_M, a complete class theorem makes this a necessary condition for admissibility (Wald, 1950; Berger, 1985), and Lemma A.2 verifies its conditions for the present problem. By Lemma 3.1, a prior π can do so only if projecting its posterior mean gaps reproduces the plug-in allocation. Among recipients, the plug-in rule preserves differences in observed gaps, while a Bayes rule preserves differences in posterior mean gaps. On an open region where the budget binds and households i and j both receive transfers, which exists by Lemma A.1, the two allocations can agree only if

$$m_{\pi,i}(x) - m_{\pi,j}(x) = x_i - x_j.$$

Under the Gaussian model, posterior means are real analytic functions of the observations, so an identity that holds on an open region holds on the whole sample space (Krantz & Parks, 2002). But a prior supported on Θ_M keeps each posterior mean gap between $-M$ and z, so the left side is bounded by $M + z$ in absolute value, while $x_i - x_j$ is not. Lemma A.3 concludes that no proper prior on Θ_M reproduces the plug-in allocation, and the rule is inadmissible.[10]

Remark 3.1 Relation to Other Inadmissibility Results. The James–Stein theorem provides an explicit estimator that improves on the observation vector under squared error

[10]The lower bound $-M$ makes the contradiction immediate by bounding posterior gap differences, but it is not essential to the non-Bayes conclusion. With p_π the marginal density of X, the displayed identity and Tweedie's formula (Efron, 2011) give $\sigma_i^2 \partial_i \log p_\pi = \sigma_j^2 \partial_j \log p_\pi$ for every x, which forces the positive density p_π to be constant along a fixed direction and contradicts its integrability.

loss in the homoskedastic Gaussian location model when $n \geq 3$ (Stein, 1956; James & Stein, 1961), and Efron and Morris (1973) give this shrinkage an empirical Bayes interpretation. In the targeting problem, estimates must be converted into transfers that satisfy the budget and nonnegativity constraints. Lower mean squared error in estimating poverty gaps does not automatically imply lower expected loss for the resulting allocation. Projection changes transfer amounts and can change which households receive them, so a dominance argument must be made for the allocation rules themselves. Proposition 3.1 establishes that an improving feasible rule exists without constructing it.

Related work studies Bayes optimality and admissibility in restricted parameter spaces (Charras & van Eeden, 1991; Fourdrinier et al., 2018). In particular, Charras and van Eeden (1991) give conditions under which estimators that take values on the boundary of a convex parameter space are not Bayes and hence are inadmissible. That argument does not apply here, because the parameter space Θ_M and the feasible set $\mathcal{A}$ impose different restrictions. Even when posterior mean gaps lie in the interior of Θ_M, their projection can lie on the boundary of $\mathcal{A}$, so exhausting the budget or assigning zero transfers is compatible with Bayes optimality.

4. Empirical Bayes Rules

Section 3 evaluated transfer rules at each vector of conditional mean incomes and established the plug-in rule's inadmissibility without constructing an improving rule. I now develop a feasible alternative with its own performance guarantee using an empirical Bayes approach. I model the coordinates of μ as draws from a common but unknown distribution G and estimate it from the income predictions. The benchmark is the oracle Bayes rule, the Bayes rule of Lemma 3.1 when G is known. The empirical Bayes rule projects estimates of the oracle's posterior mean gaps instead. Its Bayes regret, its expected loss beyond the oracle's under this model, is bounded by the mean squared difference between the estimated and the oracle posterior mean gaps.

4.1 Prior Distribution

Throughout this section, I condition on the noise levels $\sigma = (\sigma_1, \dots, \sigma_n)^\top$ and treat them as known, and I assume that, given σ, the conditional mean incomes are independent draws from a common distribution G.

Assumption 4.1 (Cross-sectional prior) *Conditional on $\sigma = (\sigma_1, \dots, \sigma_n)^\top$, the conditional mean incomes satisfy*

$$\mu_1, \dots, \mu_n \mid \sigma \overset{\text{iid}}{\sim} G, \qquad \int u^2 \, dG(u) < \infty.$$

The second-moment condition and the boundedness of $\mathcal{A}$ ensure that every rule in $\mathcal{D}$ has finite Bayes risk. The assumption also imposes precision independence, since the distribution of μ_i does not vary with σ_i. This restriction is common in heteroskedastic empirical Bayes models, and Remark 4.1 discusses how to relax it with the conditional location-scale empirical Bayes (CLOSE) approach of J. Chen (2026).

4.2 Oracle Bayes Rule

The oracle Bayes rule, available only when G is known, is the benchmark for targeting under uncertainty and the object that the empirical Bayes rule approximates. For each household i, define the posterior mean poverty gap by $g_{G,i}(\hat{y}_i, \sigma_i) := \mathbb{E}_G[z - \mu_i \mid \hat{y}_i, \sigma_i]$, and let $g_G(\hat{y}, \sigma) := (g_{G,1}(\hat{y}_1, \sigma_1), \dots, g_{G,n}(\hat{y}_n, \sigma_n))^\top$. Under Assumption 4.1, $g_G(\hat{y}, \sigma)$ is the posterior mean of the gap vector, so by Lemma 3.1, the oracle applies the perfect-information allocation of Proposition 2.1 to these posterior mean gaps.

Proposition 4.1 (Oracle Bayes Rule): *Under Assumptions 2.1 and 4.1, the oracle Bayes rule is*

$$\tau_G^\star(\hat{y}) = P_{\mathcal{A}}\big(g_G(\hat{y}, \sigma)\big).$$

Equivalently, $\tau_{G,i}^\star(\hat{y}) = \big(g_{G,i}(\hat{y}_i, \sigma_i) - \lambda_G(\hat{y})\big)_+$ for $i = 1, \dots, n$, where $\lambda_G(\hat{y}) = 0$ if $\mathbf{1}^\top g_G(\hat{y}, \sigma)_+ \le B$ and otherwise $\lambda_G(\hat{y}) > 0$ solves $\mathbf{1}^\top \tau_G^\star(\hat{y}) = B$. The rule uniquely minimizes the posterior expected loss at every $\hat{y}$, and every Bayes rule in $\mathcal{D}$ agrees with it for Lebesgue-almost every $\hat{y}$.

Proposition 4.1 differs from Proposition 2.1 only in what is projected. The full-information rule projects the vector of poverty gaps onto the feasible transfer set, and the oracle Bayes rule projects the vector of their posterior means. This is why shrinkage matters for targeting under uncertainty. The oracle replaces each observed gap with its posterior mean under G before allocating, so the estimated poverty gaps of households with noisy or unusually extreme signals are pulled toward values that are more plausible given the cross section. Relative to the plug-in rule, the oracle is therefore less willing to concentrate the budget on a few households that look especially poor in the noisy data alone, and it typically reaches more households with smaller transfers.

4.3 Estimating the Prior Distribution

To make the oracle rule feasible, I estimate G by the nonparametric maximum likelihood estimator (NPMLE) of Kiefer and Wolfowitz (1956), which maximizes the marginal likelihood of the observed signals under the Gaussian mixture model induced by Assumptions 2.1 and 4.1 without imposing a parametric form on G. The shape of G determines how noisy signals are translated into posterior mean poverty gaps and hence into transfers. A parametric family would restrict that shape in advance, whereas the NPMLE learns it from the cross section, so the feasible rule can aim at the oracle for the actual G rather than at the best rule within a prespecified family (Jiang & Zhang, 2009).

Let $\mathcal{G}$ denote the class of all probability distributions on $\mathbb{R}$. For $H \in \mathcal{G}$, define

$$f_{H,\sigma_i}(\hat{y}_i) = \int \frac{1}{\sigma_i}\varphi\left(\frac{\hat{y}_i - \mu}{\sigma_i}\right)\, dH(\mu),$$

where φ is the standard normal density. The NPMLE is any maximizer of the marginal log-likelihood,

$$\hat{G} \in \arg\max_{H \in \mathcal{G}} \frac{1}{n}\sum_{i=1}^{n} \log f_{H,\sigma_i}(\hat{y}_i).$$

Two features of the NPMLE matter for computation. First, the optimization ranges over an infinite-dimensional class, but the log-likelihood is concave in H, so the problem can be solved by convex programming over probability weights on a fixed grid of income values (Koenker & Mizera, 2014). Second, a maximizer can be chosen to be discrete with at most n support points (Koenker & Gu, 2017), whether or not G itself is discrete.

4.4 Feasible Empirical Bayes Rule

Given $\hat{G}$, the feasible rule is obtained by substituting $\hat{G}$ for G in the oracle construction. For each household i, define the feasible posterior mean poverty gap by $\hat{g}_i(\hat{y}_i, \sigma_i) := \mathbb{E}_{\hat{G}}[z - \mu_i \mid \hat{y}_i, \sigma_i]$, and let $\hat{g}(\hat{y}, \sigma) := (\hat{g}_1(\hat{y}_1, \sigma_1), \ldots, \hat{g}_n(\hat{y}_n, \sigma_n))^\top$. The feasible empirical Bayes rule is then

$$\hat{\tau}^{\mathrm{EB}}(\hat{y}) = P_{\mathcal{A}}\big(\hat{g}(\hat{y}, \sigma)\big).$$

As in Proposition 4.1, this rule admits an equivalent thresholding representation, with $\mathbb{E}_{\hat{G}}$ in place of $\mathbb{E}_G$ and a data-dependent cutoff $\hat{\lambda}^{\mathrm{EB}}(\hat{y})$. Both rules apply shrinkage and then projection. Under the oracle, household i's posterior mean gap depends only on its own signal and noise level, and other households affect its transfer only through the common budget cutoff. Under the feasible rule, $\hat{G}$ is estimated from all n signals, so other households also enter through the shrinkage step.

4.5 Statistical Properties

The Bayes regret of the feasible rule, its expected loss in excess of the oracle's, is bounded by the mean squared error of the estimated posterior mean gaps relative to the oracle's. Under two assumptions taken from Jiang (2020), his finite-sample oracle inequality bounds that error, so the feasible rule inherits the convergence rate of NPMLE posterior means.

4.5.1 Assumptions

The maximizer of the marginal likelihood over all distributions is not available in closed form, so the first assumption allows $\hat{G}$ to be an approximate maximizer of the kind that numerical implementations of the NPMLE return.

Assumption 4.2 (Approximate NPMLE) *Let f_{G,σ_i} be the marginal density defined in Section 4.3. The computed estimator $\hat{G}$ satisfies*

$$\prod_{i=1}^{n} f_{\hat{G},\sigma_i}(\hat{y}_i) \geq q_n \sup_{\tilde{G}\in\mathcal{G}} \prod_{i=1}^{n} f_{\tilde{G},\sigma_i}(\hat{y}_i), \qquad q_n = \min\{e\sqrt{2\pi}/n^2, 1\}.$$

The assumption requires the likelihood at $\hat{G}$ to be at least a fraction q_n of the maximum, which is the tolerance in equation (2.1) of Jiang (2020), under which his Theorem 1 holds. Since q_n is of order n^{-2}, the average log-likelihood at $\hat{G}$ may fall below the maximum by a term of order $(\log n)/n$.

The second assumption controls the tails of G. A finite absolute moment of order p limits how much mass G places in its tails, and a larger p limits it more tightly.

Assumption 4.3 (Moment condition) *Let*

$$\mathcal{M}_p(G) := \left(\int |u|^p \, dG(u)\right)^{1/p}$$

denote the p-th root of the p-th absolute moment of the prior G. Fix $p \geq 2$ such that $\mathcal{M}_p(G) < \infty$.

Assumption 4.1 already guarantees the condition for $p = 2$, so the assumption adds a requirement only when the bound is applied with $p > 2$. Even then, only a finite moment of some order is required, so G may have polynomial tails. A log-normal distribution, for example, has every finite moment and satisfies the assumption for every finite p.

4.5.2 Bayes Regret

Following the empirical Bayes literature, I evaluate the feasible rule by comparing it with the oracle Bayes rule under the true prior distribution G. Conditional on σ, let $\mathbb{E}_{G,\hat{Y}}$ denote the expectation under the joint distribution induced by $\mu_1, \ldots, \mu_n \overset{\text{iid}}{\sim} G$ and $\hat{Y}_i \mid (\mu_i, \sigma_i) \sim N(\mu_i, \sigma_i^2)$. The Bayes regret of $\hat{\tau}^{\text{EB}}$ relative to $\tau_G^\star$ is

$$\text{BR}(\hat{\tau}^{\text{EB}}, \tau_G^\star \mid \sigma) := \mathbb{E}_{G,\hat{Y}}\left[L\Big(\hat{\tau}^{\text{EB}}(\hat{Y}), \mu\Big) - L\Big(\tau_G^\star(\hat{Y}), \mu\Big)\right].$$

Bayes regret measures the statistical cost of estimating G. When it is close to zero, the feasible rule performs, on average, almost as well as the oracle that knows the prior. Proposition 4.2 bounds it for all n large enough.

Proposition 4.2 (Finite-Sample Bayes Regret Bound): *Let Assumptions 2.1, 4.1, 4.2, and 4.3 hold, and condition on σ. Let $p \geq 2$ be as in Assumption 4.3. Define*

$$\varepsilon(n, G, p) := \max\left\{\sqrt{2\log n}, \left(n^{1/p}\sqrt{\log n}\,\mathcal{M}_p(G)\right)^{\frac{p}{2+2p}}\right\}\sqrt{\frac{\log n}{n}},$$

and let $\Delta_n := M_0\,\varepsilon(n, G, p)\,(\log n)^{3/2}$, where M_0 is a constant depending only on p and the noise bounds $\sigma_{\min}$ and $\sigma_{\max}$. Then, for all n large enough such that $\log n > 1/p$,

$$\mathrm{BR}(\hat{\tau}^{\mathrm{EB}}, \tau_G^{\star} \mid \sigma) \leq 2\bar{\sigma}\,\Delta_n + \Delta_n^2,$$

where $\bar{\sigma}^2 = 1/n\sum_{i=1}^n \sigma_i^2$.

For fixed G and p, Δ_n converges to zero, so the cost of estimating the prior vanishes as the cross section grows. The leading term is $2\bar{\sigma}\,\Delta_n$, since Δ_n^2 is of smaller order. Up to logarithmic factors, the bound is of order $n^{-p/(2(p+1))}$, so $p = 2$ gives $n^{-1/3}$ and the exponent approaches $1/2$ as p grows. A smaller p permits heavier tails at the cost of a slower rate, and a larger p imposes tighter control of the tails and yields a faster one.

The proof has a geometric step and a statistical step. The former bounds the cost of allocating with estimated rather than oracle posterior mean gaps. Both rules project their posterior mean gaps onto the same feasible set, and this projection has kinks where the recipient set changes or the budget switches between slack and binding. The allocation is therefore a nonsmooth function of the estimated posterior mean gaps. A small error in these gaps changes recipient status only for households near the cutoff, and for these households, little is at stake. Their transfers are small because a recipient's transfer falls continuously to zero as its posterior mean gap approaches the cutoff. Each unit they receive because of the error also costs little, because their marginal reduction in loss is almost that of a recipient. Similarly, near the switch between a binding and a slack budget, the marginal reduction in loss from additional spending is close to zero, and a small error in how much to spend costs little. The loss from these mistakes is the product of the amount misallocated and its cost per unit, and both are at most of the order of the error. Lemma A.4 makes this precise, using only that the loss is quadratic and the feasible set is convex. The regret is at most $\frac{1}{n}\mathbb{E}_{G,\hat{Y}}\|\hat{g}(\hat{Y},\sigma) - g_G(\hat{Y},\sigma)\|^2$, the mean squared error of the estimated posterior mean gaps relative to the oracle's.

The statistical step is Theorem 1 of Jiang (2020), which bounds the excess root mean squared error of the NPMLE posterior means relative to the oracle's by Δ_n. The oracle's root mean squared error is at most $\bar{\sigma}$, because its posterior means have no greater mean squared error than the raw signals. Jiang's bound therefore controls how accurately the feasible rule learns the oracle's posterior mean gaps, and the projection inequality shows that this is enough to allocate almost as well as the oracle.

The moment condition is chosen for generality, and the rate it delivers need not be the sharpest possible.[11] The argument is modular, since any bound on the mean squared error of the estimated posterior means yields a Bayes regret bound of the same order through Lemma A.4. Under compact support, for example, the bounds of Jiang and Zhang (2009) and Soloff et al. (2025) give an error of order $1/n$ up to logarithmic factors, so the Bayes regret has an upper bound of the same order.

Remark 4.1 Relaxing Independence Assumptions. The regret bound in Proposition 4.2 is derived under two independence restrictions, precision independence and cross-household independence. Neither restriction enters the projection argument.

The first restriction fails when the conditional distribution of μ_i given σ_i depends on σ_i. In that case, one can instead use the CLOSE approach of J. Chen (2026). CLOSE models the conditional distribution of μ_i given σ_i as a location-scale family whose location and scale vary with σ_i, while the common shape is estimated nonparametrically from residualized data. By the modularity of the argument, any bound on the discrepancy between the CLOSE posterior means and their oracle counterparts yields a regret bound for the feasible rule.

The second restriction fails when the signals share estimation error, as in the PMT setting of Remark 2.2. In the saturated case, however, collapsing to types restores independence, and the approach applies with the weights of Remark 2.1 and one signal

[11] Lower bounds for posterior mean estimation, such as those of Polyanskiy and Wu (2021), do not by themselves establish targeting lower bounds, because projection can leave transfers unchanged despite estimation error. Lemma A.5 bounds the Bayes regret below by $n^{-1}\mathbb{E}\|\delta(\hat{Y}) - \tau^{\star}_G(\hat{Y})\|^2$. A matching minimax lower bound on this allocation error, over the same model class, would establish optimality of the rate in Proposition 4.2.

per type. The precision of a type mean reflects how common the type is, and CLOSE handles the case in which mean income varies with type frequency. In general, the criterion in Section 4.3 is a composite marginal likelihood rather than the joint likelihood, studied by Han and Zhang (2026) under correlated Gaussian noise and by J. Chen, Deb, and Ignatiadis (2026) under approximate normality and estimated variances. Neither result establishes vanishing regret for PMT with a fixed number of coefficients. All predictions share the estimation error of those coefficients, and the effective sample size stays bounded as the target population grows. Extending Proposition 4.2 to this case remains an open question. The simulations of Section 5 do not impose normality or independence, and there the empirical Bayes rule still shows gains over the plug-in rule.

Appendix B.5 studies regret rates under the one-sided poverty gap loss ℓ_+. Relative to the squared-loss case, the key difference is that performance depends on posterior tail behavior near the poverty line rather than on posterior means alone. The resulting problem is therefore more difficult because it requires the procedure to recover the parts of the latent income distribution that govern those tail probabilities, and the associated regret rates are slower.

5. Empirical Simulation Evidence

This section evaluates the allocation rules in an empirical simulation built from the household consumption data for nine African countries assembled by Brown et al. (2018). In each replication, the policymaker observes only consumption predicted from PMT models estimated on a small training sample, as in the setting studied by Corral et al. (2025), and allocates a fixed budget to the held-out population. I compare the plug-in OLS rule, a plug-in XGBoost rule, and the empirical Bayes rule of Section 4. The empirical Bayes rule reaches 1.77 times as many poor people as plug-in OLS under the same budget ceiling and achieves the same poverty gap reduction with 6.7% less transfer spending. Under quadratic loss, it outperforms plug-in OLS in all 9 countries and XGBoost in 8 of the 9. I also evaluate the rules under the one-sided poverty gap loss developed in Appendix B.

5.1 Setup

The simulation uses the household survey data assembled by Brown et al. (2018) for nine African countries: Burkina Faso, Ethiopia, Ghana, Malawi, Mali, Niger, Nigeria, Tanzania, and Uganda. For each country, I set the poverty line at the person-weighted 40th percentile of per capita consumption and let y_{ic} denote the consumption of household i in country c as a share of that line, so the normalized poverty line is $z_c = 1$ for every country.[12] The losses, budgets, and allocation rules use person weights ω_{ic}, obtained by multiplying each household's survey weight by its household size and rescaling to average one within each replication's target population, so the simulated population is representative of people.

The prediction exercise is repeated 500 times for each country. In each replication, I draw a training sample of 500 households, stratified by urban status and region, and estimate two PMT models for normalized consumption on the same "basic PMT" covariate set from Brown et al. (2018).[13] The target population consists of the remaining households with complete PMT covariates, for which both models predict consumption. Appendix Table D.1 reports country-level descriptives and prediction quality.

I compare three feasible allocation rules. The plug-in OLS rule treats the OLS prediction as true consumption. The plug-in XGBoost rule does the same with the XGBoost prediction. The empirical Bayes rule uses the OLS prediction and its heteroskedasticity-robust standard error, computed with the HC1 degrees-of-freedom correction. This standard error measures sampling uncertainty in the estimated conditional mean. The rule estimates the prior distribution of conditional mean consumption by nonparametric maximum likelihood using the person weights ω_{ic} on a grid restricted to nonnegative values, computes each household's posterior mean poverty gap, and projects the vector of posterior mean gaps onto the feasible transfer set.[14] The projection imposes the

[12] Consumption is winsorized at the person-weighted 95th percentile before normalization. Both percentiles are computed on all households with valid consumption and survey weights, before the sample is restricted to households with complete PMT covariates, so the target-population poverty rates are close to but not exactly 40 percent.

[13] OLS is estimated without survey weights. XGBoost is tuned within each training sample by fivefold cross-validation over a prespecified grid of tree depth and minimum child weight, with early stopping selecting the number of boosting rounds.

[14] The grid has 600 equally spaced points from zero to the largest OLS prediction. Appendix Table D.2 evaluates the standard GLmix support, which starts at the smallest prediction. The prior-estimation solver reports an optimal solution in all 500 replications in each country, and all three headline rules are evaluated on these same replications.

budget as an inequality, so the rule leaves part of the budget unspent when the positive posterior mean gaps sum to less than the budget.

The design is an empirical simulation rather than a fully model-generated Monte Carlo. Household consumption is not drawn from a known prior, and the prediction error is not generated by adding synthetic normal noise to survey consumption. Instead, the empirical distribution of consumption, the relationship between the PMT covariates and consumption, and the resulting prediction errors all arise from the observed data and from repeated finite-sample estimation of the PMT model. Thus, the simulation does not impose the assumptions used in the theoretical analysis, including normality of the signals, conditional independence across households, and correct specification of the linear conditional mean. The empirical Bayes rule is nevertheless implemented as if the Gaussian signal model were exact, so the exercise asks whether it remains useful in a realistic environment where the conditions that motivate it do not hold exactly.

Let B_c denote the total transfer budget for country c's target population. Each country has the same transfer budget per represented person, $b_c := B_c / \sum_i \omega_{ic} = 0.005$, or 0.5 percent of its poverty line. The budget therefore has the same fiscal cost as paying a transfer equal to 10 percent of the poverty line to 5 percent of the population. A common value keeps program size comparable across countries. On this per-person basis, the benchmark is generous relative to the program scenarios documented in Appendix Table D.3, exceeding 10 of the 11 reported budgets. Even so, the budget amounts to only 3.29%–5.19% of the aggregate poverty gap, depending on the country. The results are broadly the same when the budget varies over a sixteen-fold range or matches each documented program scenario, as Appendix Table D.2 shows.

The main performance measure is gain under quadratic loss. For any feasible allocation $\hat{\tau}_c$, define

$$\mathrm{Gain}_{L_{2,c}}(\hat{\tau}_c) = \frac{L_{2,c}(0) - L_{2,c}(\hat{\tau}_c)}{L_{2,c}(0) - L_{2,c}(\tau_c^{\mathrm{PI}})},$$

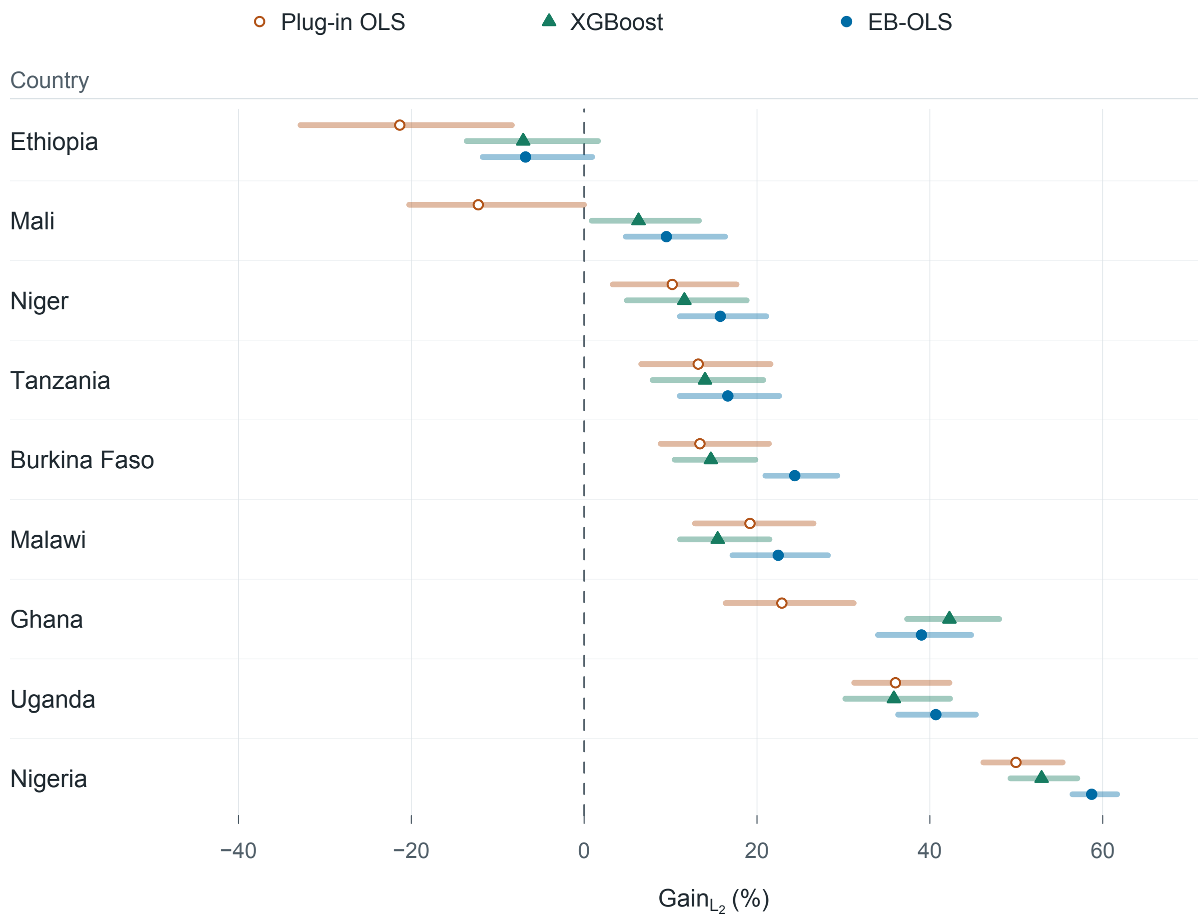
where $L_{2,c}(\tau)$ is the person-weighted quadratic loss of allocation τ evaluated at realized consumption in country c, $L_{2,c}(0)$ is the loss without transfers, and τ_c^{PI} is the perfect-information allocation of Proposition 2.1. I report this ratio as a percentage of the perfect-information improvement over no transfers. A gain of 0% means no improvement over no transfers, 100% means that the allocation attains the same improvement as the perfect-information allocation, and a negative gain means that it increases quadratic loss relative to no transfers. Differences in gains are reported in percentage points. I also report the analogous measure, Gain_{L_+}, under the one-sided poverty gap loss.

5.2 Main Results

Relative to the plug-in rules, the empirical Bayes rule based on OLS predictions (EB-OLS) reaches more poor people at a given budget and requires less transfer spending to achieve a given reduction in poverty. With the same budget, EB-OLS gives a positive transfer to 12.3% of initially poor people, 1.77 times the 7.0% reached by plug-in OLS, while XGBoost reaches 11.0%. The same advantage appears as lower spending when poverty reduction is held fixed. In Appendix Figure D.2, I lower the budget available to EB-OLS until its mean realized poverty gap reduction across countries and replications equals that of a plug-in rule at the headline budget. Comparing the transfers actually paid out, EB-OLS matches the poverty gap reduction of plug-in OLS with 6.7% less spending and that of XGBoost with 3.4% less. For a program that already runs a PMT regression, these savings come at no additional financial cost. The survey data, the regression, and the predictions remain the same. Only the allocation step changes, with transfers assigned using posterior mean poverty gaps rather than raw predictions.

EB-OLS also delivers the highest average gain under quadratic loss. On average, it recovers 24.5% of the reduction in quadratic loss attainable under perfect information relative to no transfers, compared with 14.6% for plug-in OLS and 20.7% for XGBoost. Figure 1 reports the comparison country by country. EB-OLS yields higher gains than plug-in OLS in all 9 countries and XGBoost in 8 of the 9. When the same allocations are instead evaluated under the one-sided loss, their cross-country averages retain the same ordering, with Gain_{L_+} equal to 40.4% for EB-OLS, 38.9% for XGBoost, and 36.3% for plug-in OLS.

Figure 1: Targeting performance under quadratic loss



Notes. Symbols report country-specific mean gains under quadratic loss over 500 simulation replications at the common budget $b = 0.005$. Gains are reported as percentages of the perfect-information improvement over no transfers. A gain of 0% matches the no-transfer benchmark, and 100% matches the perfect-information benchmark, which is outside the displayed range. Pale intervals span the 25th to 75th percentiles across replications. Negative gains indicate higher quadratic loss than with no transfers. Countries are ordered by mean plug-in OLS gain, from lowest to highest.

The comparisons of EB-OLS with plug-in OLS and with XGBoost isolate different changes. Plug-in OLS and EB-OLS start from the same predictions, but EB-OLS uses their uncertainty to replace the predicted poverty gaps with posterior mean poverty gaps. This replacement raises the gain in every country, with increases ranging from 3.3 to 21.8 percentage points. XGBoost instead changes the prediction model while retaining the plug-in allocation. EB-OLS attains a higher average gain than XGBoost in every country except Ghana, although its margin in Ethiopia is negligible.[15]

[15]Across 4,500 country–replication pairs, EB-OLS outperforms plug-in OLS in 85.6% of cases, with an average quadratic-gain advantage of 12.2 percentage points when it wins and a disadvantage of 3.9 percentage points when it loses. Against XGBoost, it wins in 65.4% of cases; the corresponding advantage

The absolute level of the gain also varies widely across countries, and it is associated with two features of each country. The first is how much the predictions reveal about who is poor. Although 40.1% of people are poor, OLS predicts only 18.6%–32.4% of people to be below the line and correctly identifies only 27.4%–60.7% of poor people as poor. Shrinkage uses an imperfect prediction more cautiously, and a flexible model extracts different patterns from the same covariates, but neither can recover information that the covariates do not contain, so all three feasible rules tend to yield lower gains where the predictions reveal less. The second feature is the share of people near the poverty line, where small prediction errors turn intended gap closure into overshoot or leakage. In countries with a share of people within 10 percent of the line at or above the median, the mean gain of EB-OLS is 19.3%, compared with 31.0% for the other countries, and the plug-in rules show the same pattern.

Ethiopia and Nigeria mark the two ends of this range. In Ethiopia, the out-of-sample OLS regression has $R^2 = 0.119$ and predicts only 18.6% of people to be poor. All three rules perform worse than making no transfers under the quadratic loss. The households with the largest predicted gaps are often not the poorest, so much of the budget overshoots or leaks to the nonpoor, and the quadratic criterion counts both against the rule. EB-OLS increases quadratic loss much less than plug-in OLS and about as much as XGBoost, because it shrinks the largest predicted gaps and spends only where posterior mean gaps remain positive, which leaves about a third of the budget unspent. Its gain is -6.8%, compared with -21.3% for plug-in OLS and -7.0% for XGBoost. By contrast, in Nigeria, the out-of-sample regression has $R^2 = 0.481$, and plug-in OLS already attains a gain of 50.0%. EB-OLS raises it to 58.7%, above the 52.9% attained by XGBoost. Shrinkage therefore improves on plug-in OLS whether the predictions are weak or strong, but it can only use the information the covariates contain more cautiously, not replace what they lack.[16]

Appendix Table D.2 varies the budget and the design. Over the sixteen-fold budget range, the advantage of EB-OLS over plug-in OLS grows with the budget, while its advantage over XGBoost narrows. EB-OLS remains the best-performing rule under ev-

and disadvantage are 10.0 and 7.8 percentage points.

[16] The empirical Bayes rule is not built on XGBoost because the signal model requires approximately unbiased predictions with standard errors for their sampling variation. OLS supplies both when the linear model is correctly specified, as Remark 2.2 shows, whereas XGBoost predictions carry regularization bias that bootstrap standard errors would not measure.

ery prespecified change to the design. Its advantage over plug-in OLS is largest with 250 training households—where the predictions are noisiest and plug-in OLS does no better than making no transfers—and shrinks as the training sample grows to 1,000. Clustering the OLS covariance matrix by enumeration area leaves the results unchanged, and the CLOSE-NPMLE rule of J. Chen (2026), which lets the prior depend on the standard error, performs somewhat below EB-OLS and above both plug-in rules. At the documented program budgets, EB-OLS has the highest gain in 10 of the 11 scenarios.[17]

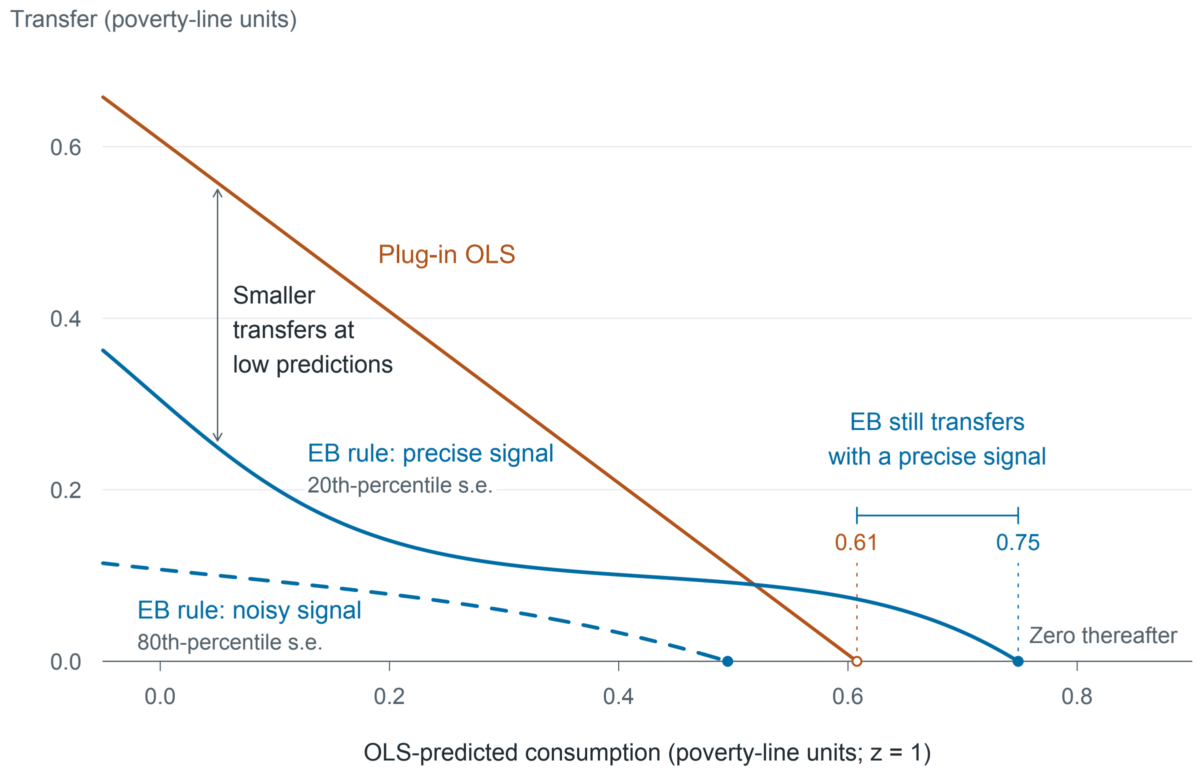
5.3 Allocation Mechanisms

Figure 2 plots the transfer assigned by plug-in OLS and by EB-OLS against predicted consumption for Mali in one replication. Under plug-in, the transfer equals the predicted poverty gap minus a common cutoff. It falls linearly as predicted consumption rises and reaches zero at the cutoff. Under EB-OLS, the transfer depends on the prediction and on its standard error, and the figure draws the schedule for a precise and for a noisy prediction. Both curves are nonlinear in predicted consumption because the estimated posterior mean is. The noisy curve lies below the precise curve for all low predictions, so a household whose consumption is estimated less precisely receives a smaller transfer at the same low prediction. At the lowest predictions, the EB curves also lie below the plug-in line because EB-OLS trims the largest transfers. The resources trimmed there help finance transfers to households with precise signals and predicted consumption above the plug-in cutoff. The EB allocation is less concentrated, with smaller transfers at the lowest predictions and positive transfers for a broader set of households.

Averaged across countries and replications, EB-OLS reaches more people with smaller transfers. Table 1 reports transfer coverage and budget use for the three rules, Appendix Table D.4 repeats it by country, and Appendix Table D.5 reports transfer amounts among recipients. Plug-in OLS treats 3.6% of people, with a 90th-percentile transfer among recipients of 0.33 poverty-line units. EB-OLS treats 6.6% of people and lowers the 90th-percentile transfer to 0.17. EB-OLS gives each recipient less and ex-

[17] The exception is Malawi at the smallest budget evaluated, where plug-in OLS transfers only to the households with the most extreme predicted gaps, a group it targets very accurately, whereas shrinkage spreads the same budget over more than twice as many households and admits more nonpoor recipients. At the larger Malawi budget in the table, EB-OLS leads by 3.0 percentage points.

Figure 2: How the empirical Bayes rule reallocates transfers under prediction uncertainty



Notes. Mali, replication 221, selected mechanically from the allocation schedules. Predictions and transfers are expressed in poverty-line units. The EB curves hold the prediction standard error at its person-weighted 20th percentile (precise) and 80th percentile (noisy), with all other components of the fitted EB allocation rule held fixed. Actual household allocations use each household's own standard error. Plug-in OLS and EB-OLS each exhaust the same budget, $b = 0.005$. The marked schedules remain zero beyond their respective cutoffs.

tends assistance to substantially more poor people, instead of making large transfers to a small group selected on noisy predictions.

Table 1 also shows how the rules use the budget. For every 100 units of available budget, overshoot—spending on poor recipients beyond their initial poverty gaps—falls from 10.62 under plug-in OLS to 4.69 under EB-OLS, while leakage to initially nonpoor recipients changes little, from 19.28 to 19.20. The principal change in budget use is therefore lower overshoot rather than lower payments to the nonpoor. XGBoost also reduces overshoot, to 4.85, but raises leakage to 21.75. EB-OLS and XGBoost leave 4.28 and 2.95 per 100 units unspent on average, mostly in Ethiopia and Mali.[18]

[18]Appendix Table D.2 shows that requiring every rule to spend the full budget leaves the gain of EB-OLS essentially unchanged, so the unspent budget does not explain its advantage.

Table 1: Targeting outcomes across nine countries

	Gain (%)		Transfer coverage		Use of available budget (%)			
Rule	Gain_{L_2}	Gain_{L_+}	Population treated (%)	Poor people reached (%)	Poverty gaps closed	Overshoot	Leakage	Unspent
Plug-in OLS	14.6	36.3	3.6	7.0	70.1	10.6	19.3	0.0
XGBoost	20.7	38.9	5.9	11.0	70.4	4.9	21.8	3.0
EB-OLS	24.5	40.4	6.6	12.3	71.8	4.7	19.2	4.3

Notes: Equal-weighted averages of country-specific simulation means are reported over 500 draws. Outcomes are evaluated at realized consumption in the held-out target population using person weights. The common budget $b = 0.005$ equals 0.5% of the poverty line per represented person. Gains are reported as percentages of the perfect-information improvement over no transfers. A gain of 0% matches the no-transfer benchmark, and 100% matches the perfect-information benchmark. Gain_{L_+} evaluates the same allocations under L_+. Population treated is the share of persons receiving a positive transfer. Poor people reached is the corresponding share among those initially poor. The final four columns divide 100 budget units among poverty gaps closed, overshoot (spending on poor recipients beyond their initial gaps), leakage to initially nonpoor recipients, and unspent resources.

The broader transfer coverage extends to the poorest people. Appendix Table D.5 shows that the share of extremely poor people—those with consumption below $0.5z$—who receive a transfer rises from 12.1% under plug-in OLS to 19.6% under EB-OLS. Despite reaching more of them, EB-OLS closes 2.8% of their aggregate pre-transfer poverty gap—a share similar to the 3.0% under plug-in OLS and 2.6% under XGBoost. Under EB-OLS, fewer extremely poor people receive nothing, but the average reduction in the poverty gap among those reached is smaller. Appendix Figure D.4 shows the consequence for headcounts. EB-OLS and XGBoost lift more people out of extreme poverty, while plug-in OLS lifts more people out of poverty. The small transfers of EB-OLS and XGBoost raise the consumption of deeply poor households but usually leave them below the line, and the large transfers of plug-in OLS more often carry their recipients all the way across it.

The empirical Bayes rule for the one-sided loss, developed in Appendix B, uses the same estimated posterior distribution but does not penalize overshoot and therefore exhausts the budget. Evaluated under the one-sided loss, it attains a cross-country mean gain of 41.5%, compared with 36.3% for plug-in OLS and 38.9% for XGBoost. It yields higher gains than plug-in OLS in 9 of the 9 countries and XGBoost in 8 of the 9. Appendix Figure D.3 reports the comparison by country, and Appendix Table D.6 evaluates both EB rules under both losses. In the seven countries where the quadratic rule spends the full budget, the two EB rules perform nearly identically, with gains within 0.1 percentage points of each other under the one-sided loss and within 0.5 percentage points under the quadratic loss. Material differences are confined to Ethiopia and Mali, where the quadratic rule leaves part of the budget unspent and the one-sided rule allocates it, raising the gain under the one-sided loss. The choice of objective therefore matters empirically mainly when the penalty on overshoot leads the quadratic rule to stop spending.

6. Conclusion

Targeting programs invest in better data and better prediction models and then convert the predictions into transfers with a rule that treats them as if they were true. This paper shows that the conversion step discards information the budget could use. The plug-in rule is inadmissible. The empirical Bayes rule pools information across households, shrinking each estimated poverty gap toward what the distribution of incomes in the population makes plausible and taking the precision of each estimate into account, before imposing the budget. Its regret relative to the oracle Bayes allocation is bounded by the estimation error in the posterior mean gaps, so the budget and nonnegativity constraints do not slow the rate at which it approaches the oracle. The one-sided objective and fixed benefit menus show that the same approach carries over to program designs closer to practice, with weaker guarantees.

The practical implication is that antipoverty programs can improve targeting with the information they already have. The empirical Bayes rule needs the PMT predictions and the standard errors of the estimated conditional means, which can be computed from the fitted regression. It changes nothing about data collection, the prediction model, or the eligibility survey. In the nine-country simulations, that change alone reaches about 1.8 times as many poor people as plug-in OLS under the same budget, or attains the same average poverty gap reduction with 6.7% less transfer spending. Tanzania has the smallest estimated saving in the simulations, 1.9 percent. If that rate carried over to its Productive Social Safety Net at the program's 2015 scale, the estimated annualized transfer outlay of about 140 million dollars would imply savings of about 2.6

million dollars a year. Better targeting can come from the predictions programs already have, once the step that converts them into transfers is properly designed.

This paper takes the prediction model as given and optimizes the allocation rule. Designing the prediction and the allocation together is a natural direction for future work. In the simulations, OLS and XGBoost predict consumption about equally well, yet their plug-in allocations perform very differently, so overall prediction accuracy does not determine targeting performance. Programs should evaluate a prediction method together with the allocation rule built on it, under the targeting loss studied in this paper. Extending the analysis to joint design would require accounting for local bias in machine learning predictors and estimation error shared across households, both of which lie outside the regret bound.

References

Ai, C., Fang, Y., & Xie, H. (2024). *Data-driven policy learning for continuous treatments.* (arXiv:2402.02535)

Alatas, V., Banerjee, A., Hanna, R., Olken, B. A., & Tobias, J. (2012). Targeting the poor: evidence from a field experiment in indonesia. *American Economic Review*, *102*(4), 1206–1240.

Areias, A., & Wai-Poi, M. (2022). Machine learning and prediction of beneficiary eligibility for social protection programs. In *Revisiting targeting in social assistance: A new look at old dilemmas* (pp. 507–539). World Bank. doi: 10.1596/978-1-4648-1814-1_ch8

Athey, S., & Wager, S. (2021). Policy learning with observational data. *Econometrica*, *89*(1), 133–161.

Baez, J. E., Kshirsagar, V., & Skoufias, E. (2019). *Adaptive safety nets for rural Africa: Drought-sensitive targeting with sparse data* (Policy Research Working Paper No. 9071). World Bank. doi: 10.1596/1813-9450-9071

Baker, J. L., & Grosh, M. E. (1994). Poverty reduction through geographic targeting: How well does it work? *World Development*, *22*(7), 983–995.

Banerjee, A., Hanna, R., Olken, B. A., & Sverdlin Lisker, D. (2024). Social protection in the developing world. *Journal of Economic Literature*, *62*(4), 1349–1421. doi: 10.1257/jel.20241646

Barrientos, A. (2013). *Social assistance in developing countries.* Cambridge University Press.

Beegle, K., De Weerdt, J., Friedman, J., & Gibson, J. (2012). Methods of household consumption measurement through surveys: Experimental results from Tanzania. *Journal of Development Economics*, *98*(1), 3–18. doi: 10.1016/j.jdeveco.2011.11.001

Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis* (2nd ed.). Springer. doi: 10.1007/978-1-4757-4286-2

Bhattacharya, D., & Dupas, P. (2012). Inferring welfare maximizing treatment assignment under budget constraints. *Journal of Econometrics*, *167*(1), 168–196.

Bourguignon, F., & Fields, G. S. (1990). Poverty measures and anti-poverty policy. *Recherches Économiques de Louvain/Louvain Economic Review*, *56*(3-4), 409–427. doi: 10.1017/S0770451800043979

Bourguignon, F., & Fields, G. S. (1997). Discontinuous losses from poverty, generalized Pα measures, and optimal transfers to the poor. *Journal of Public Economics*, *63*(2), 155–175. doi: 10.1016/S0047-2727(96)01589-7

Boyd, S. P., & Vandenberghe, L. (2004). *Convex optimization.* Cambridge University Press.

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

Cameron, L., & Shah, M. (2014). Can mistargeting destroy social capital and stimulate crime? Evidence from a cash transfer program in Indonesia. *Economic Development and Cultural Change*, *62*(2), 381–415. doi: 10.1086/674102

Charras, A., & van Eeden, C. (1991). Bayes and admissibility properties of estimators in truncated parameter spaces. *The Canadian Journal of Statistics*, *19*(2), 121–134. doi: 10.2307/3315792

Charras, A., & van Eeden, C. (1992). Bayes properties of estimators of location parameters in truncated parameter spaces. *Statistics & Risk Modeling*, *10*(1-2), 81–86. doi: 10.1524/strm.1992.10.12.81

Chen, J. (2026). Empirical Bayes when estimation precision predicts parameters. *Econometrica*, *94*(2), 305–340. doi: 10.3982/ECTA22935

Chen, J., Deb, N., & Ignatiadis, N. (2026). Normal approximations in nonparametric empirical Bayes. *arXiv preprint arXiv:2605.31599*.

Chen, J., Lei, L., Sudijono, T., Sun, L., & Xie, T. (2025). *Compound selection decisions: An almost SURE approach.* (arXiv:2511.11862)

Chen, X., Chong, W. F., Feng, R., & Zhang, L. (2021). Pandemic risk management: Resources contingency planning and allocation. *Insurance: Mathematics and Economics*, *101*, 359–383.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review*, *104*(9), 2593–2632.

Chetty, R., & Hendren, N. (2018). The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects. *The Quarterly Journal of Economics*, *133*(3), 1107–1162.

Combettes, P. L. (2018). Monotone operator theory in convex optimization. *Mathematical Programming*, *170*(1), 177–206. doi: 10.1007/s10107-018-1303-3

Corral, P., Henderson, H., & Segovia, S. (2025). Poverty mapping in the age of machine learning. *Journal of Development Economics*, *172*, 103377. doi: 10.1016/j.jdeveco.2024.103377

Della Guardia, A., Lake, M., & Schnitzer, P. (2022). Selective inclusion in cash transfer programs: Unintended consequences for social cohesion. *World Development*, *157*, 105922. doi: 10.1016/j.worlddev.2022.105922

Devereux, S., Masset, E., Sabates-Wheeler, R., Samson, M., Rivas, A.-M., & Te Lintelo, D. (2017). The targeting effectiveness of social transfers. *Journal of Development Effectiveness*, *9*(2), 162–211. doi: 10.1080/19439342.2017.1305981

Efron, B. (2011). Tweedie's formula and selection bias. *Journal of the American Statistical Association*, *106*(496), 1602–1614.

Efron, B., & Morris, C. (1973). Stein's estimation rule and its competitors. An empirical Bayes approach. *Journal of the American Statistical Association*, *68*(341), 117–130.

Elbers, C., Fujii, T., Lanjouw, P., Özler, B., & Yin, W. (2007). Poverty alleviation through geographic targeting: How much does disaggregation help? *Journal of Development Economics*, *83*(1), 198–213.

Elbers, C., Lanjouw, J. O., & Lanjouw, P. (2003). Micro-level estimation of poverty and inequality. *Econometrica*, *71*(1), 355–364. doi: 10.1111/1468-0262.00399

Ferguson, T. S. (1967). *Mathematical statistics: A decision theoretic approach.* Academic Press.

Follett, L., & Henderson, H. (2023). A hybrid approach to targeting social assistance. *Journal of Development Economics*, *160*, 103002. doi: 10.1016/j.jdeveco.2022.103002

Foster, J., Greer, J., & Thorbecke, E. (1984). A class of decomposable poverty measures. *Econometrica*, *52*(3), 761–766. doi: 10.2307/1913475

Fourdrinier, D., Strawderman, W. E., & Wells, M. T. (2018). *Shrinkage estimation.* Springer.

Gu, J., & Koenker, R. (2023). Invidious comparisons: Ranking and selection as compound decisions. *Econometrica*, *91*(1), 1–41.

Han, Q., & Zhang, C.-H. (2026). Empirical Bayes for correlated Gaussian sequence model. *arXiv preprint arXiv:2607.03596*.

Hartigan, J. A. (2004). Uniform priors on convex sets improve risk. *Statistics & Probability Letters*, *67*(4), 285–288.

Hirano, K., & Porter, J. R. (2009). Asymptotics for statistical treatment rules. *Econometrica*, *77*(5), 1683–1701.

Hirano, K., & Porter, J. R. (2020). Asymptotic analysis of statistical decision rules in econometrics. In S. N. Durlauf, L. P. Hansen, J. J. Heckman, & R. L. Matzkin (Eds.), *Handbook of econometrics* (Vol. 7A, pp. 283–354). Elsevier. doi: 10.1016/bs.hoe.2020.09.001

James, W., & Stein, C. (1961). Estimation with quadratic loss. In *Proc. 4th berkeley symp. on math. statist. and prob., 1961* (pp. 361–379). University of California Press.

Jiang, W. (2020). On general maximum likelihood empirical Bayes estimation of heteroscedastic IID normal means. *Electronic Journal of Statistics*, *14*(1), 2272–2297. doi: 10.1214/20-EJS1717

Jiang, W., & Zhang, C.-H. (2009). General maximum likelihood empirical Bayes estimation of normal means. *The Annals of Statistics*, *37*(4), 1647–1684. doi: 10.1214/08-AOS638

Kiefer, J., & Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *The Annals of Mathematical Statistics*, *27*(4), 887–906. doi: 10.1214/aoms/1177728066

Kitagawa, T., Lee, S., & Qiu, C. (2026). Treatment choice with nonlinear regret. *Biometrika*, *113*(2). (asag008)

Kitagawa, T., & Tetenov, A. (2018). Who should be treated? Empirical welfare maximization methods for treatment choice. *Econometrica*, *86*(2), 591–616.

Kitagawa, T., & Tetenov, A. (2021). Equality-minded treatment choice. *Journal of Business & Economic Statistics*, *39*(2), 561–574.

Kline, P., Rose, E. K., & Walters, C. R. (2022). Systemic discrimination among large US employers. *The Quarterly Journal of Economics*, *137*(4), 1963–2036.

Kline, P., Rose, E. K., & Walters, C. R. (2024). A discrimination report card. *American Economic Review*, *114*(8), 2472–2525.

Koenker, R., & Gu, J. (2017). REBayes: an R package for empirical Bayes mixture methods. *Journal of Statistical Software*, *82*(8), 1–26. doi: 10.18637/jss.v082.i08

Koenker, R., & Mizera, I. (2014). Convex optimization, shape constraints, compound decisions, and empirical Bayes rules. *Journal of the American Statistical Association*, *109*(506), 674–685. doi: 10.1080/01621459.2013.869224

Krantz, S. G., & Parks, H. R. (2002). *A primer of real analytic functions.* Springer Science & Business Media.

Kwon, S. (2026). Optimal shrinkage estimation of fixed effects in linear panel data models. *Econometrica*, *94*(2), 663–677.

Manski, C. F. (2004). Statistical treatment rules for heterogeneous populations. *Econometrica*, *72*(4), 1221–1246.

Marchand, É., & Perron, F. (2001). Improving on the MLE of a bounded normal mean. *The Annals of Statistics*, *29*(4), 1078–1093. doi: 10.1214/aos/1013699994

Merfeld, J. D., & Morduch, J. (2023). Poverty at higher frequency. *KDI School of Pub Policy & Management Paper No. DP23-03.*

Montiel Olea, J. L., O'Flaherty, B., & Sethi, R. (2021). Empirical Bayes counterfactuals in Poisson regression with an application to police use of deadly force. *Available at SSRN 3857213.*

Moon, S. (2025). Optimal policy choices under uncertainty. *arXiv preprint arXiv:2503.03910.*

Polyanskiy, Y., & Wu, Y. (2021). Sharp regret bounds for empirical bayes and compound decision problems. *arXiv preprint arXiv:2109.03943*.

Premand, P., & Schnitzer, P. (2021). Efficiency, legitimacy, and impacts of targeting methods: Evidence from an experiment in Niger. *The World Bank Economic Review*, *35*(4), 892–920. doi: 10.1093/wber/lhaa019

Ravallion, M. (2016). *The economics of poverty: History, measurement, and policy*. Oxford University Press. doi: 10.1093/acprof:oso/9780190212766.001.0001

Saha, S., & Guntuboyina, A. (2020). On the nonparametric maximum likelihood estimator for Gaussian location mixture densities with application to Gaussian denoising. *The Annals of Statistics*, *48*(2), 738–762. doi: 10.1214/19-AOS1817

Sahoo, R., Blumenstock, J., Niehaus, P., Selker, L., & Wager, S. (2025, December). *What would it cost to end extreme poverty?* (Working Paper No. 34583). National Bureau of Economic Research.

Sen, A. (1976). Poverty: An ordinal approach to measurement. *Econometrica*, *44*(2), 219–231. doi: 10.2307/1912718

Smith, P. C., Rice, N., & Carr-Hill, R. (2001). Capitation funding in the public sector. *Journal of the Royal Statistical Society: Series A*, *164*(2), 217–257.

Soloff, J. A., Guntuboyina, A., & Sen, B. (2025). Multivariate, heteroscedastic empirical Bayes via nonparametric maximum likelihood. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *87*(1), 1–32. doi: 10.1093/jrsssb/qkae040

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the third berkeley symposium on mathematical statistics and probability* (Vol. 1, pp. 197–206). University of California Press.

Stoye, J. (2009). Minimax regret treatment choice with finite samples. *Journal of Econometrics*, *151*(1), 70–81.

Sun, L. (2026). Empirical welfare maximization with constraints. *Journal of Econometrics*, *253*, 106169.

Tetenov, A. (2012). Statistical treatment choice based on asymmetric minimax regret criteria. *Journal of Econometrics*, *166*(1), 157–165.

Wald, A. (1950). *Statistical decision functions*. New York: Wiley.

Zaslavsky, A. M., & Schirm, A. L. (2002). Interactions between survey estimates and federal funding formulas. *Journal of Official Statistics*, *18*(3), 371–391.

Zhang, C.-H. (2009). Generalized maximum likelihood estimation of normal mixture densities. *Statistica Sinica*, *19*(3), 1297–1318.

A. Proofs and Supporting Results for the Main Text

This appendix collects the proofs of the main-text results and the supporting lemmas.

A.1 Proofs of Main-Text Results

Proof of Proposition 2.1. The objective is

$$\frac{1}{n}\sum_{i=1}^{n}(z-y_i-\tau_i)^2 = \frac{1}{n}\|\tau-(z\mathbf{1}-y)\|^2.$$

Because $\mathcal{A}$ is nonempty, closed, and convex, the projection of $z\mathbf{1}-y$ onto $\mathcal{A}$ exists and is unique, so the unique minimizer is $P_{\mathcal{A}}(z\mathbf{1}-y)$.

To obtain the threshold representation, multiply the objective by n. Since $B>0$, the allocation $\tau_i = B/(2n)$ is strictly feasible, so the KKT conditions are necessary and sufficient. With nonnegative multipliers ν for the budget constraint and ξ_i for nonnegativity, stationarity gives $2(\tau_i - z + y_i) + \nu - \xi_i = 0$. Together with $\tau_i \geq 0$, $\xi_i \geq 0$, and $\xi_i\tau_i = 0$, this implies $\tau_i = \big(z-y_i-\nu/2\big)_+$. Write $\lambda^{\mathrm{PI}} := \nu/2$, so that $\tau_i = \big(z-y_i-\lambda^{\mathrm{PI}}\big)_+$. Budget feasibility and complementary slackness require

$$T(\lambda^{\mathrm{PI}}) \leq B, \qquad \lambda^{\mathrm{PI}}\big(T(\lambda^{\mathrm{PI}})-B\big) = 0, \qquad \text{where} \quad T(a) := \sum_{i=1}^{n}\big(z-y_i-a\big)_+.$$

If $T(0)\leq B$, setting $\lambda^{\mathrm{PI}}=0$ satisfies these conditions. Otherwise, $\lambda^{\mathrm{PI}}>0$ and $T(\lambda^{\mathrm{PI}})=B$. The function T is continuous, strictly decreasing wherever positive, and eventually zero, so this equation has a unique positive solution.

Finally, the threshold formula implies $y_i + \tau_i^{\mathrm{PI}}(y) = z - \lambda^{\mathrm{PI}}$ for every recipient and zero transfers otherwise, which is the leveling representation. QED.

Proof of Lemma 3.1. Fix x. The Gaussian likelihood is bounded and the marginal density of X is strictly positive everywhere, so the posterior defined by Bayes' formula exists at every x and, by the second-moment condition on π, has finite second moments. Expanding the square, $\mathbb{E}_\pi[L(\tau,\theta)\mid X=x] = n^{-1}\|\tau-m_\pi(x)\|^2 + n^{-1}\mathbb{E}_\pi[\|\theta-m_\pi(x)\|^2 \mid X=x]$. The first term is strictly convex in τ, so $P_{\mathcal{A}}(m_\pi(x))$ is its unique minimizer over $\mathcal{A}$. The map m_π is measurable and $P_{\mathcal{A}}$ is 1-Lipschitz (Combettes, 2018), hence continuous, so $\delta_\pi \in \mathcal{D}$. Since $\mathcal{A}$ is bounded and π has a finite second moment, every rule in $\mathcal{D}$ has finite Bayes risk, which equals the integral of posterior expected loss against the marginal law of X. Hence δ_π minimizes Bayes risk, and any other minimizer attains the posterior minimum for marginal-almost every x and therefore agrees with δ_π there. The marginal has a strictly positive Lebesgue density, so the exceptional set is Lebesgue-null. QED.

Proof of Proposition 3.1. Suppose, toward a contradiction, that τ^{plug} is admissible in $\mathcal{D}$. By Lemma A.2, there exists a proper Borel prior π on Θ_M and a Bayes rule δ_π under π such that $\tau^{\text{plug}}(X) = \delta_\pi(X) \qquad P_\theta$-a.s. for every $\theta \in \Theta_M$. Thus τ^{plug} is Bayes, up to the usual almost-sure equivalence of decision rules, with respect to a proper prior on Θ_M.

This contradicts Lemma A.3, which shows that no proper prior on Θ_M can make $\tau^{\text{plug}}(x) = P_{\mathcal{A}}(x)$ Bayes, even up to P_θ-almost-sure equivalence for every $\theta \in \Theta_M$, under squared-error loss. Therefore τ^{plug} is inadmissible in $\mathcal{D}$. QED.

Proof of Proposition 4.1. Let π be the prior that Assumption 4.1 induces on the gap vector θ. It is proper with finite second moment, and conditional on σ, Assumptions 2.1 and 4.1 imply that the posterior factorizes across households, so $m_\pi(z\mathbf{1} - \hat{y}) = g_G(\hat{y}, \sigma)$ for every $\hat{y}$. Lemma 3.1 then shows that $P_{\mathcal{A}}(g_G(\hat{y}, \sigma))$ uniquely minimizes posterior expected loss over $\mathcal{A}$ at every $\hat{y}$, minimizes Bayes risk over $\mathcal{D}$, and agrees with every other Bayes rule in $\mathcal{D}$ for Lebesgue-almost every $\hat{y}$. The threshold form is Proposition 2.1 applied with the posterior mean incomes $z\mathbf{1} - g_G(\hat{y}, \sigma)$ in place of y, with $\lambda_G(\hat{y})$ the cutoff determined by its budget condition. QED.

Proof of Proposition 4.2. Throughout, condition on σ, and write $\mathbb{E}[\cdot]$ for expectation under the joint distribution of $(\mu, \hat{Y})$ induced by Assumptions 2.1–4.1.

Let $\tilde{\mu}_{G,i} := \mathbb{E}_G[\mu_i \mid \hat{Y}_i, \sigma_i]$ and $\tilde{\mu}_G := (\tilde{\mu}_{G,1}, \ldots, \tilde{\mu}_{G,n})^\top$. Since $g_G(\hat{Y}, \sigma) = z\mathbf{1} - \tilde{\mu}_G$, expanding the squared loss, taking conditional expectation given $(\hat{Y}, \sigma)$, and canceling the posterior-variance term gives

$$\mathrm{BR}(\hat{\tau}^{\mathrm{EB}}, \tau_G^\star \mid \sigma) = \frac{1}{n}\mathbb{E}\Big[\big\|z\mathbf{1} - \tilde{\mu}_G - \hat{\tau}^{\mathrm{EB}}\big\|^2 - \big\|z\mathbf{1} - \tilde{\mu}_G - \tau_G^\star\big\|^2\Big].$$

Now set $u := g_G(\hat{Y}, \sigma)$ and $v := \hat{g}(\hat{Y}, \sigma)$. By definition, $\tau_G^\star = P_{\mathcal{A}}(u)$ and $\hat{\tau}^{\mathrm{EB}} = P_{\mathcal{A}}(v)$, so the integrand is $\|u - P_{\mathcal{A}}(v)\|^2 - \|u - P_{\mathcal{A}}(u)\|^2$. Applying Lemma A.4 pointwise yields

$$\mathrm{BR}(\hat{\tau}^{\mathrm{EB}}, \tau_G^\star \mid \sigma) \le \frac{1}{n}\,\mathbb{E}\Big[\|\hat{g}(\hat{Y}, \sigma) - g_G(\hat{Y}, \sigma)\|^2\Big].$$

It remains to relate the posterior mean poverty-gap discrepancy to a posterior-mean risk difference in the heteroskedastic normal-means problem. Define $\tilde{\mu}_{\hat{G},i} := \mathbb{E}_{\hat{G}}[\mu_i \mid \hat{Y}_i, \sigma_i]$. Since $\hat{g}_i(\hat{Y}_i, \sigma_i) = z - \tilde{\mu}_{\hat{G},i}$ and $g_{G,i}(\hat{Y}_i, \sigma_i) = z - \tilde{\mu}_{G,i}$, we have $\hat{g}_i(\hat{Y}_i, \sigma_i) - g_{G,i}(\hat{Y}_i, \sigma_i) =$

$\tilde{\mu}_{G,i} - \tilde{\mu}_{\hat{G},i}$. Also, $\tilde{\mu}_{\hat{G},i} - \mu_i = (\tilde{\mu}_{\hat{G},i} - \tilde{\mu}_{G,i}) + (\tilde{\mu}_{G,i} - \mu_i)$. The estimated prior $\hat{G}$ is computed from the full sample, so $\tilde{\mu}_{\hat{G},i}$ is generally not fixed after conditioning only on $(\hat{Y}_i, \sigma_i)$. I therefore condition on the full observed data vector $(\hat{Y}, \sigma)$. Under Assumptions 2.1–4.1, conditional on σ, the pairs $(\mu_i, \hat{Y}_i)$ are independent across i. Hence $\mathbb{E}[\mu_i \mid \hat{Y}, \sigma] = \mathbb{E}[\mu_i \mid \hat{Y}_i, \sigma_i] = \tilde{\mu}_{G,i}$. Therefore $\mathbb{E}[\tilde{\mu}_{G,i} - \mu_i \mid \hat{Y}, \sigma] = 0$. Since $\tilde{\mu}_{\hat{G},i} - \tilde{\mu}_{G,i}$ is fixed after conditioning on $(\hat{Y}, \sigma)$, the cross-term vanishes:

$$\mathbb{E}\Big[(\tilde{\mu}_{\hat{G},i} - \tilde{\mu}_{G,i})(\tilde{\mu}_{G,i} - \mu_i)\Big] = \mathbb{E}\Big[(\tilde{\mu}_{\hat{G},i} - \tilde{\mu}_{G,i})\mathbb{E}[\tilde{\mu}_{G,i} - \mu_i \mid \hat{Y}, \sigma]\Big] = 0.$$

Expanding the square, summing over i, and rearranging gives

$$\frac{1}{n}\mathbb{E}\Big[\|\hat{g}(\hat{Y}, \sigma) - g_G(\hat{Y}, \sigma)\|^2\Big] = \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n}(\tilde{\mu}_{\hat{G},i} - \mu_i)^2\right] - \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n}(\tilde{\mu}_{G,i} - \mu_i)^2\right].$$

In words, the expected posterior-mean poverty-gap discrepancy equals the excess average squared-error risk of $\tilde{\mu}_{\hat{G}}$ relative to the oracle posterior mean $\tilde{\mu}_G$.

By Jiang (2020), Theorem 1,

$$\left\{\frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n}(\tilde{\mu}_{\hat{G},i} - \mu_i)^2\right]\right\}^{1/2} - \left\{\frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n}(\tilde{\mu}_{G,i} - \mu_i)^2\right]\right\}^{1/2} \leq \Delta_n.$$

Writing $R^* := \frac{1}{n}\mathbb{E}[\sum_{i=1}^{n}(\tilde{\mu}_{G,i} - \mu_i)^2]$, this implies $\frac{1}{n}\mathbb{E}[\|\hat{g} - g_G\|^2] \leq 2\Delta_n\sqrt{R^*} + \Delta_n^2$.

To bound R^*, note that under squared loss the posterior mean minimizes conditional risk. Taking the competitor $\delta_i = \hat{Y}_i$ therefore gives $\mathbb{E}[(\tilde{\mu}_{G,i} - \mu_i)^2] \leq \mathbb{E}[(\hat{Y}_i - \mu_i)^2] = \sigma_i^2$, where the last equality uses Assumption 2.1. Hence $R^* \leq n^{-1}\sum_{i=1}^{n}\sigma_i^2 =: \bar{\sigma}^2$, so $\sqrt{R^*} \leq \bar{\sigma}$.

Combining the preceding bounds yields $\mathrm{BR}(\hat{\tau}^{\mathrm{EB}}, \tau_{\hat{G}}^{\star} \mid \sigma) \leq 2\bar{\sigma}\,\Delta_n + \Delta_n^2$. This proves the proposition. QED.

A.2 Supporting Lemmas

Lemma A.1 (Local form of the plug-in rule): *For every nonempty $K \subseteq \{1, \ldots, n\}$ there is a nonempty open set $\mathcal{U}_K \subseteq \mathbb{R}^n$ on which the plug-in rule exhausts the budget, has active set K, and satisfies*

$$\tau_i^{\mathrm{plug}}(x) = \begin{cases} x_i - \lambda_K(x), & i \in K, \\ 0, & i \notin K, \end{cases} \qquad \lambda_K(x) := \frac{\sum_{\ell\in K} x_\ell - B}{|K|},$$

with $\lambda_K(x) > 0$ on $\mathcal{U}_K$. The rule is therefore affine on $\mathcal{U}_K$, and $\tau_i^{\text{plug}}(x) - \tau_j^{\text{plug}}(x) = x_i - x_j$ for all $i, j \in K$ and $x \in \mathcal{U}_K$.

Proof of Lemma A.1. Let x^0 have $x_i^0 = 2B/|K|$ for $i \in K$ and $x_j^0 = 0$ for $j \notin K$. Then $\lambda_K(x^0) = B/|K| > 0$, $x_i^0 > \lambda_K(x^0)$ for $i \in K$, and $x_j^0 < \lambda_K(x^0)$ for $j \notin K$. Since λ_K is continuous, these strict inequalities hold on an open neighborhood $\mathcal{U}_K$ of x^0. Fix $x \in \mathcal{U}_K$. The vector $t := (x - \lambda_K(x)\mathbf{1})_+$ has positive coordinates exactly on K and satisfies $\mathbf{1}^\top t = B$, and $\mathbf{1}^\top x_+ \geq \sum_{i \in K} x_i = B + |K|\lambda_K(x) > B$. Thus t satisfies the KKT characterization of the projection established in the proof of Proposition 2.1, so $t = \tau^{\text{plug}}(x)$. The affine form and the difference identity follow. QED.

Lemma A.2 (Admissible rules are Bayes): *Under Assumptions 2.1 and 3.1, every rule $\delta \in \mathcal{D}$ that is admissible on Θ_M agrees P_θ-almost surely, for every $\theta \in \Theta_M$, with a Bayes rule under some proper prior on Θ_M.*

Proof of Lemma A.2. The parameter space Θ_M and the action space $\mathcal{A}$ are compact, the loss is continuous and bounded on $\mathcal{A} \times \Theta_M$, and the Gaussian density is jointly continuous in (x, θ). Every rule in $\mathcal{D}$ has continuous risk, because the Gaussian location family is continuous in total variation and the loss is uniformly continuous on $\mathcal{A} \times \Theta_M$. Since $\mathcal{A}$ is convex and the loss is convex in the action, replacing a randomized action by its conditional mean preserves feasibility and weakly lowers risk, so a rule admissible in $\mathcal{D}$ is admissible among randomized rules as well. The complete-class theorem of Berger (1985, Theorem 12, Section 8.8) therefore implies that δ is Bayes under some proper prior π on Θ_M.

A prior on Θ_M has finite second moments, so by Lemma 3.1 δ agrees with $\delta_\pi(x) = P_{\mathcal{A}}(m_\pi(x))$ Lebesgue-almost everywhere and hence P_θ-almost surely for every $\theta \in \Theta_M$. QED.

Lemma A.3 (The plug-in rule is not Bayes): *Assume $n \geq 2$. Under Assumptions 2.1 and 3.1, the plug-in rule is not Bayes under any proper prior on Θ_M, even up to P_θ-almost-sure equivalence for every $\theta \in \Theta_M$.*

Proof of Lemma A.3. Suppose otherwise, and let π be the corresponding proper prior, with posterior mean $m_\pi(x) := \mathbb{E}_\pi[\theta \mid X = x]$. By Lemma 3.1 and the strict positivity of the Gaussian densities, $\tau^{\text{plug}}(x) = P_{\mathcal{A}}(m_\pi(x))$ for Lebesgue-almost every x. Each coordinate of m_π is real analytic on $\mathbb{R}^n$, being a ratio of Gaussian integrals over the compact set Θ_M with strictly positive denominator, so both sides of this equality are continuous and it holds everywhere on any open set on which it holds almost everywhere.

Fix distinct indices $i \neq j$ and let $K = \{i, j\}$. By Lemma A.1, there is a nonempty open set $\mathcal{U}_K$ on which the plug-in rule exhausts the budget and has active set K. On $\mathcal{U}_K$ the Bayes projection has the same two positive coordinates, and its common threshold cancels in their difference, so

$$m_{\pi,i}(x) - m_{\pi,j}(x) = \tau_i^{\text{plug}}(x) - \tau_j^{\text{plug}}(x) = x_i - x_j \qquad \text{for all } x \in \mathcal{U}_K.$$

The real analytic function $x \mapsto m_{\pi,i}(x) - m_{\pi,j}(x) - (x_i - x_j)$ vanishes on the nonempty open set $\mathcal{U}_K$, so by the identity theorem it vanishes on all of $\mathbb{R}^n$. This is impossible, because $|m_{\pi,i}(x) - m_{\pi,j}(x)| \leq M + z$ for every x while $|x_i - x_j|$ is unbounded. QED.

Lemma A.4 (Projection mismatch inequality): *Let $\mathcal{A} \subset \mathbb{R}^n$ be nonempty, closed, and convex, and let $P_{\mathcal{A}} : \mathbb{R}^n \to \mathcal{A}$ denote Euclidean projection. Then, for any $u, v \in \mathbb{R}^n$,*

$$\|u - P_{\mathcal{A}}(v)\|^2 - \|u - P_{\mathcal{A}}(u)\|^2 \leq \|u - v\|^2. \tag{6}$$

The constant 1 is sharp whenever $\mathcal{A}$ contains two distinct points.

Proof of Lemma A.4. Write $p := P_{\mathcal{A}}(u)$ and $q := P_{\mathcal{A}}(v)$. By convexity, $q + t(p - q) \in \mathcal{A}$ for $t \in [0, 1]$. Since q minimizes the squared distance to v over $\mathcal{A}$, the right derivative of $\|v - q - t(p - q)\|^2$ at $t = 0$ is nonnegative, which gives $\langle v - q,\, p - q\rangle \leq 0$. Hence

$$\|v - q\|^2 - \|v - p\|^2 = 2\langle v - q,\, p - q\rangle - \|p - q\|^2 \leq -\|p - q\|^2.$$

Expanding the squared distances and applying this inequality,

$$\begin{aligned}
\|u - q\|^2 - \|u - p\|^2 &= 2\langle u - v,\, p - q\rangle + \|v - q\|^2 - \|v - p\|^2 \\
&\leq 2\langle u - v,\, p - q\rangle - \|p - q\|^2 \\
&= \|u - v\|^2 - \|(u - v) - (p - q)\|^2 \\
&\leq \|u - v\|^2.
\end{aligned}$$

If $u, v \in \mathcal{A}$ are distinct, then $p = u$ and $q = v$, so (6) holds with equality and the constant cannot be reduced. QED.

Lemma A.5 (Projection mismatch lower bound): *Let $\mathcal{A} \subset \mathbb{R}^n$ be nonempty, closed, and convex, and let $P_{\mathcal{A}} : \mathbb{R}^n \to \mathcal{A}$ denote Euclidean projection. Then, for any $g \in \mathbb{R}^n$ and any $\tau \in \mathcal{A}$,*

$$\|\tau - g\|^2 - \|P_{\mathcal{A}}(g) - g\|^2 \geq \|\tau - P_{\mathcal{A}}(g)\|^2. \tag{7}$$

Proof of Lemma A.5. Write $\tau^\star := P_{\mathcal{A}}(g)$. Expanding,

$$\|\tau - g\|^2 = \|\tau - \tau^\star\|^2 + 2\langle \tau - \tau^\star,\ \tau^\star - g\rangle + \|\tau^\star - g\|^2.$$

Since $\tau^\star$ minimizes the distance to g over $\mathcal{A}$ and $(1-t)\tau^\star + t\tau \in \mathcal{A}$ for all $t \in [0,1]$, expanding $\|g - (1-t)\tau^\star - t\tau\|^2 \geq \|g - \tau^\star\|^2$ and letting $t \to 0$ gives $\langle g - \tau^\star,\ \tau - \tau^\star\rangle \leq 0$. The middle term above is therefore nonnegative, and (7) follows. QED.

Taking $g = g_G(\hat{y}, \sigma)$ and $\tau = \delta(\hat{y})$ for a rule $\delta \in \mathcal{D}$, and using that posterior expected loss equals $\|\tau - g\|^2/n$ plus a term that does not depend on τ, Lemma A.5 implies

$$\mathrm{BR}(\delta, \tau_G^\star \mid \sigma) \ \geq \ \frac{1}{n}\,\mathbb{E}\big\|\delta(\hat{Y}) - \tau_G^\star(\hat{Y})\big\|^2 \qquad \text{for every } \delta \in \mathcal{D}.$$

Online Appendix

Poverty Targeting with Imperfect Information

Juan C. Yamin

September 22, 2026

Contents of the Online Appendix

B. One-Sided Poverty Gap Objective

This appendix replaces the quadratic loss of the main text with the one-sided poverty-gap objective

$$L_+(\tau, y) = \frac{1}{n}\sum_{i=1}^{n}(z - y_i - \tau_i)_+^2,$$

which equals z^2 times the Foster–Greer–Thorbecke index with parameter $\alpha = 2$. Transfers have value only insofar as they close poverty gaps, and inclusion errors carry no cost beyond the budget they use. Under imperfect information, I evaluate the objective at conditional mean income. Unlike in the quadratic loss case, this is not equivalent to minimizing expected realized poverty, because the residual term no longer separates from the transfer.

The perfect-information rule of Proposition 2.1 remains optimal, and the plug-in rule is again inadmissible when $B > z$. The oracle equalizes posterior expected remaining gaps across recipients, and the empirical Bayes rule uses estimated counterparts of these expectations. Posterior means alone no longer determine the allocation, since the expected remaining gap depends on the posterior distribution of the gap above the transfer received, not only on its mean. Under bounded support and strict fiscal scarcity, I obtain an $O(1/\log n)$ Bayes-regret bound for the feasible rule, a weaker guarantee than the polynomial bound in Proposition 4.2.

B.1 Perfect Information

Any transfer can be capped at the household's poverty gap $(z-y_i)_+$ without changing L_+ or violating the budget. On this restricted set, the one-sided and quadratic losses differ only by a constant, so they have the same minimizer.

Corollary B.1 (Perfect-Information Policy under L_+): *The policy $\tau^{\mathrm{PI}}(y)$ of Proposition 2.1 is optimal under L_+. It is the unique minimizer if $B \le \sum_{i=1}^{n}(z - y_i)_+$. If $B > \sum_{i=1}^{n}(z - y_i)_+$, the minimum loss is zero, and the minimizers are exactly the feasible allocations satisfying $\tau_i \ge (z - y_i)_+$ for every household.*

Transfers beyond the poverty line leave L_+ unchanged, which is why uniqueness fails once the budget covers every gap. Since the perfect-information policy is unchanged, so is the plug-in rule of Section 3.1, which applies τ^{PI} to estimated incomes.

B.2 Predictable versus Realized Poverty

Under imperfect information, the one-sided loss of a rule δ can be evaluated at realized income or at conditional mean income, and the two evaluations, the realized-income objective $\mathbb{E}[L_+(\delta(\hat{Y}), y)]$ and the predictable-income objective $\mathbb{E}[L_+(\delta(\hat{Y}), \mu)]$, can rank rules differently. Under the quadratic loss, the residual adds its variance to every household's expected loss whatever the transfer, which is what makes the two evaluations agree in Remark 2.3. Under the one-sided loss, the amount it adds depends on where the transfer leaves the household. A household brought exactly to the poverty line in conditional mean income has zero loss under the predictable-income objective but can have positive expected loss under the realized-income objective, because negative residual draws leave it poor. Whether this residual poverty enters the objective is a choice of targeting mandate.

Lemma B.1 (Bounds between realized and predictable one-sided risk): *Suppose* $\mathbb{E}[\varepsilon_i \mid \mu, \hat{Y}] = 0$ *and* $\mathbb{E}[\varepsilon_i^2] < \infty$ *for every* i*, as in Remark 2.3, and let* $\bar{v}_n := \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[\varepsilon_i^2]$*. Then, for every decision rule* $\delta \in \mathcal{D}$*,*

$$0 \le \mathbb{E}\big[L_+(\delta(\hat{Y}), y)\big] - \mathbb{E}\big[L_+(\delta(\hat{Y}), \mu)\big] \le \bar{v}_n, \tag{8}$$

where the expectation is over the signals and the residuals at fixed μ*.*

For every rule, the realized-income objective exceeds the predictable-income objective by at most the average residual variance $\bar{v}_n$. Switching objectives therefore changes the risk difference between any two rules by at most $\bar{v}_n$, so a rule that beats another by more than $\bar{v}_n$ under one objective also beats it under the other. The realized-income objective depends on the probability and depth of poverty generated by residual variation, and the maintained model does not specify the conditional residual distribution needed to evaluate it. The rest of this appendix therefore studies the predictable-income objective.

B.3 Inadmissibility of Plug-In Rule

Under L_+, the plug-in rule can misallocate budget even when it exhausts it, and a direct dominance argument shows this without the Bayes-risk and complete-class arguments of Section 3. In gap notation, $L_+(\tau, \theta) = \frac{1}{n}\sum_{i=1}^{n}(\theta_i - \tau_i)_+^2$ with $\theta_i = z - \mu_i$, and since $\mu_i \geq 0$, every predictable poverty gap is at most z.[19] A transfer above z is not penalized under L_+, but the excess spends budget without reducing poverty. The plug-in rule makes such transfers with positive probability. It projects the noisy gap vector X onto the budget set, and because the Gaussian signal has unbounded support, some component of $P_{\mathcal{A}}(X)$ exceeds z on a set of positive probability whenever $B > z$.

Capping every transfer at z and spreading the freed budget uniformly across households gives the cap-and-spread rule

$$\tau_i^{\text{cs}}(x) := \min\{\tau_i^{\text{plug}}(x),\ z\} + \frac{1}{n}\sum_{j=1}^{n}\big(\tau_j^{\text{plug}}(x) - z\big)_+, \qquad i = 1, \ldots, n,$$

which preserves total spending and uses no information beyond x. Capping leaves the loss of the affected households at zero, since a transfer of z already closes any gap, and spreading cannot raise the loss of anyone else, since L_+ is nonincreasing in transfers. When total poverty gaps exceed the budget, the plug-in allocation leaves some gap uncovered. On realizations where the cap releases funds, that household receives more, and its loss falls strictly.

Proposition B.1 (A simple domination of the plug-in rule under L_+): *Let $R_+(\delta, \theta) := \mathbb{E}_\theta[L_+(\delta(X), \theta)]$ denote the frequentist risk under the loss L_+. Under Assumption 2.1 and $B > z$,*

(i) For every $\theta \in (-\infty, z]^n$, $R_+(\tau^{\text{cs}}, \theta) \leq R_+(\tau^{\text{plug}}, \theta)$.

(ii) For every $\theta \in (-\infty, z]^n$ satisfying $\sum_{i=1}^{n}(\theta_i)_+ > B$, the inequality in (i) is strict.

In particular, the plug-in rule is inadmissible on any parameter space $\Theta \subseteq (-\infty, z]^n$ that contains at least one θ with $\sum_{i=1}^{n}(\theta_i)_+ > B$.

[19] The gap formulation of Section 3.1 applies verbatim under L_+, since $(z - \mu_i - \tau_i)_+^2 = (\theta_i - \tau_i)_+^2$.

Nothing recommends τ^{cs} beyond this comparison. It shows only that noisy signals lead the plug-in rule to spend budget on transfers no household could need, and that under L_+ those dollars have no poverty-reduction value.

B.4 Oracle and Empirical Bayes Rules under L_+

Under the quadratic loss, household i enters the oracle problem only through its posterior mean gap $g_{G,i}$, and the oracle of Proposition 4.1 projects the vector of posterior mean gaps onto the budget set. Under L_+, the posterior expected loss of household i at transfer t is $\mathbb{E}_G[(z-\mu_i-t)_+^2 \mid \hat{y}_i, \sigma_i]$, and an additional dollar lowers it only in the posterior states where the household is still below the poverty line after receiving t. Under the moment condition of Proposition B.2,

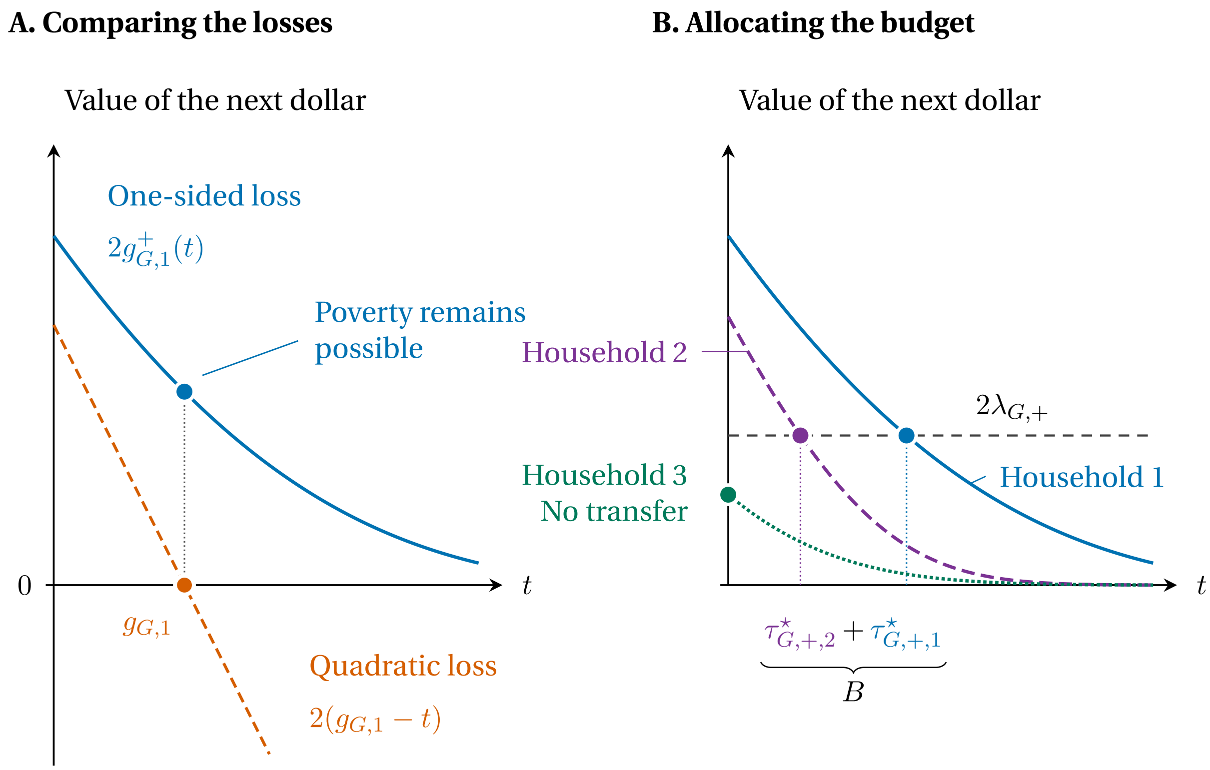
$$-\frac{\partial}{\partial t}\mathbb{E}_G\big[(z-\mu_i-t)_+^2 \mid \hat{y}_i, \sigma_i\big] = 2\, g^+_{G,i}(t \mid \hat{y}_i, \sigma_i), \qquad g^+_{G,i}(t \mid \hat{y}_i, \sigma_i) := \mathbb{E}_G[(z-\mu_i-t)_+ \mid \hat{y}_i, \sigma_i]\,,$$

so $2g^+_{G,i}(t \mid \hat{y}_i, \sigma_i)$ is the value of the next dollar to household i, and $g^+_{G,i}(t \mid \hat{y}_i, \sigma_i)$ is its posterior expected remaining gap after a transfer of t. The expected remaining gap is the posterior probability that the household remains poor after the transfer times its expected shortfall in that event,

$$g^+_{G,i}(t \mid \hat{y}_i, \sigma_i) = \Pr_G(z-\mu_i > t \mid \hat{y}_i, \sigma_i)\; \mathbb{E}_G[z-\mu_i-t \mid z-\mu_i > t, \hat{y}_i, \sigma_i]\,.$$

Panel A of Figure B.1 compares the marginal values for household 1 under the two losses. Under quadratic loss, $2(g_{G,1}-t)$ reaches zero at the posterior mean gap and becomes negative beyond it, because the loss also penalizes predictable income above the poverty line after the transfer. Under one-sided loss, $2g^+_{G,1}(t)$ lies weakly above the quadratic line. With a nondegenerate posterior, it remains positive at $t = g_{G,1}$, because states with gaps above the mean remain poor and an additional transfer still reduces expected loss in those states. The size of this expected shortfall depends on how the posterior distributes probability around its mean. In particular, because $(z-\mu_i-t)_+$ is convex in μ_i, a mean-preserving spread of the posterior weakly raises $g^+_{G,i}(t)$ at every t. Households with the same posterior mean gap can therefore value an additional dollar differently and, as the next result shows, receive different transfers under L_+.

Figure B.1: Marginal values and allocation under one-sided loss



Notes. Curves show marginal reductions in expected household loss; under L_+, these are $2g^+_{G,i}(t)$. A common bounded uniform prior and Gaussian signals yield the illustrated truncated-normal posteriors. Households 1 and 2 share the same posterior mean, but household 1's noisier signal produces a more dispersed posterior. Panel A uses household 1 from Panel B. Recipients reach $2\lambda_{G,+}$, while household 3 receives nothing because $2g^+_{G,3}(0) < 2\lambda_{G,+}$.

Because the posterior expected loss is nonincreasing in the transfer, an optimal allocation can always spend the whole budget, so the proposition imposes $\sum_i \tau_i = B$.

Proposition B.2 (Oracle Bayes Rule under L_+): *Under Assumptions 2.1 and 4.1, fix $(\hat{y}, \sigma)$, and assume that each household has finite posterior one-sided second moment, $\mathbb{E}_G[(z - \mu_i)^2_+ \mid \hat{y}_i, \sigma_i] < \infty$. Consider the equality-constrained oracle Bayes problem*

$$\min_{\tau \in \mathbb{R}^n_+} \frac{1}{n} \sum_{i=1}^n \mathbb{E}_G\left[(z - \mu_i - \tau_i)^2_+ \mid \hat{y}_i, \sigma_i\right] \quad \textit{subject to} \quad \sum_{i=1}^n \tau_i = B.$$

Then, there exist an optimal allocation $\tau^\star_{G,+}$ and a scalar $\lambda_{G,+} \geq 0$ such that $\sum_{i=1}^n \tau^\star_{G,+,i} = B$ and, for each household i, $\tau^\star_{G,+,i} = 0$ only if $g^+_{G,i}(0 \mid \hat{y}_i, \sigma_i) \leq \lambda_{G,+}$, while any household with $\tau^\star_{G,+,i} > 0$ satisfies $g^+_{G,i}(\tau^\star_{G,+,i} \mid \hat{y}_i, \sigma_i) = \lambda_{G,+}$.

Under L_+, the oracle equalizes the posterior expected remaining gaps across recipients, as the perfect-information rule equalizes the actual remaining gaps. Equivalently, recipients reach the common marginal value $2\lambda_{G,+}$. If two recipients had different marginal values, a small reallocation toward the one with the higher value would reduce the expected loss.

Panel B of Figure B.1 illustrates this allocation for three households, with the cutoff chosen so that the transfers exhaust the budget. Households 1 and 2 share the same posterior mean gap, so the quadratic oracle would give them equal transfers. Here, household 1's noisier signal produces a more dispersed posterior, so its marginal value reaches $2\lambda_{G,+}$ at a larger transfer. Household 3 receives nothing because its marginal value at a zero transfer is already below $2\lambda_{G,+}$, even though poverty remains possible.

The empirical Bayes rule replaces G with the Kiefer–Wolfowitz NPMLE $\hat{G}$. For each household i and transfer level $t \geq 0$, let $\hat{g}_i^+(t \mid \hat{y}_i, \sigma_i) := \mathbb{E}_{\hat{G}}[(z - \mu_i - t)_+ \mid \hat{y}_i, \sigma_i]$. Since $\hat{G}$ is discrete, $\hat{g}_i^+$ is a finite posterior-weighted sum of the lines $(z - u - t)_+$ over the support points u of $\hat{G}$, hence continuous and piecewise linear in t. In the binding-budget case, the empirical Bayes allocation $\hat{\tau}_+^{\mathrm{EB}}$ is defined by replacing $g_{G,i}^+$ with $\hat{g}_i^+$ and $\lambda_{G,+}$ with the budget-clearing cutoff $\hat{\lambda}_+$ in the oracle characterization.[20] Its input is a remaining-gap function for each household rather than a vector of posterior mean gaps.

B.5 Bayes Regret under L_+

Under the assumptions below, the Bayes regret of the feasible rule is $O(1/\log n)$. Two ingredients deliver the bound. Strict scarcity keeps the oracle's cutoff bounded away from zero with high probability, and a positive cutoff makes the posterior expected loss grow at least quadratically with the distance from the oracle allocation, so errors in the estimated remaining-gap functions translate into regret. Those errors are controlled on average across households by the Wasserstein accuracy of the NPMLE, which is where the slow rate of Gaussian deconvolution enters.

[20] If several allocations are optimal, $\hat{\tau}_+^{\mathrm{EB}}$ denotes a fixed measurable choice among those that spend the full budget.

B.5.1 Regret Criterion and Remaining-Gap Error

As in Section 4.5, the feasible rule is evaluated by its Bayes regret relative to the oracle rule under the true prior,

$$\mathrm{BR}_+\left(\hat{\tau}_+^{\mathrm{EB}}, \tau_{G,+}^\star \mid \sigma\right) := \mathbb{E}_{G,\hat{Y}}\left[L_+\left(\hat{\tau}_+^{\mathrm{EB}}(\hat{Y},\sigma),\mu\right) - L_+\left(\tau_{G,+}^\star(\hat{Y},\sigma),\mu\right) \,\middle|\, \sigma\right],$$

where the expectation is over $(\mu, \hat{Y})$ at fixed σ. The two rules solve the same allocation problem and differ only in the prior they use, so the regret is the cost of replacing G with $\hat{G}$.

Throughout, the true prior G and the fitted prior $\hat{G}$ are assumed to be supported on a bounded interval of nonnegative incomes, and $\hat{\tau}_+^{\mathrm{EB}}$ is the rule of Section B.4 computed from $\hat{G}$. Nonnegative incomes mean that no poverty gap exceeds z, so a transfer above z cannot reduce loss further. When $0 < B < nz$, any allocation that spends B can be capped at z per household and the amount removed redistributed within the caps without raising loss. Lemma E.1 therefore lets both rules be chosen with $0 \le \tau_i \le z$ and $\sum_i \tau_i = B$ without changing their optimal values. From here on, $\tau_{G,+}^\star$ and $\hat{\tau}_+^{\mathrm{EB}}$ denote these capped rules, and $\lambda_{G,+} \ge 0$ is the common posterior expected remaining gap at the oracle allocation.

The statistical input to regret is the discrepancy between the estimated and true remaining-gap functions, measured by its root mean square across households, $r_+(\hat{y},\sigma)$, where

$$r_+^2(\hat{y},\sigma) := \frac{1}{n}\sum_{i=1}^{n}\Big[\sup_{t\in[0,z]}\left|\hat{g}_i^+(t \mid \hat{y}_i,\sigma_i) - g_{G,i}^+(t \mid \hat{y}_i,\sigma_i)\right|\Big]^2.$$

The supremum over t is needed because each household's transfer is chosen after the data are observed, so the estimated function must be accurate at a data-dependent point, and $[0,z]$ covers every transfer the capped rules can choose. Averaging across households matches the normalization of the loss, and since both functions take values in $[0,z]$, each household contributes at most z^2/n.

B.5.2 Assumptions

The regret bound needs two conditions beyond the maintained assumptions, one on the estimated prior and one on the budget.

Assumption B.1 (Compact support and support-constrained approximate NPMLE) *For some $M < \infty$, the true and the estimated mixing distributions satisfy $G([0, M]) = 1$ and $\hat{G}([0, M]) = 1$. In addition, there is a constant $C_\ell < \infty$ such that*

$$\ell_n(G) - \ell_n(\hat{G}) \le C_\ell \frac{(\log n)^2}{n}, \qquad \textit{where} \quad \ell_n(H) := \frac{1}{n}\sum_{i=1}^{n} \log \int \phi_{\sigma_i}(\hat{Y}_i - u)\, dH(u). \tag{9}$$

Here, ϕ_σ denotes the $N(0, \sigma^2)$ density.

The support condition has two roles. Nonnegative incomes bound every poverty gap by z, which the capped rules use. The upper bound M keeps the posterior stable in the prior. A household's remaining-gap function is a ratio of integrals against the prior, and by Lemma E.3, its estimation error on a bounded support is at most the Wasserstein distance $W_2(G, \hat{G})$ times a factor that depends only on the household's own signal. Accuracy of $\hat{G}$ in Wasserstein distance is therefore what the allocation needs.

The likelihood condition (9) is the approximate-maximizer condition of Soloff, Guntuboyina, and Sen (2025). It requires $\hat{G}$ to explain the data nearly as well as the true prior does, and their deconvolution theory shows that this alone makes an approximate NPMLE close to the true prior. By Lemma E.5, their argument adapted to the bounded support gives $W_2^2(G, \hat{G}) \lesssim 1/\log n$ with probability at least $1 - 2/n^8$. The condition is easy to meet. An exact maximizer over distributions supported on $[0, M]$ satisfies it with zero gap, since G itself is a candidate. By Lemma E.4, a maximizer over a grid of mesh δ_n computed to a likelihood tolerance η_n satisfies it whenever $\delta_n^2 + \eta_n \lesssim (\log n)^2/n$.[21]

The deepest poverty gap the prior allows is $z - \inf \operatorname{supp}(G)$, the gap of a household at the lowest income in the support of G. Scarcity requires spending per household to stay below this gap by a margin.

Assumption B.2 (Strict fiscal scarcity) *There is a constant $\bar{b} < z - \inf \operatorname{supp}(G)$ such that $B/n \le \bar{b}$ for all sufficiently large n.*

A constant total budget satisfies the condition once the population is large enough, provided G puts positive mass below z, and a budget proportional to the population

[21] The empirical implementation estimates $\hat{G}$ on 600 equally spaced points over $[0, \max\{0, \max_{1\le i\le n} \hat{Y}_i\} + 10^{-4}]$, with a GLmix relative tolerance of 10^{-8}. The upper endpoint is data dependent rather than a known bound on the support of G, and the mesh and tolerance were not chosen to satisfy the rate condition of Lemma E.4, so Proposition B.3 does not formally cover the simulation implementation.

satisfies it whenever the proportion is below $z - \inf \operatorname{supp}(G)$. The threshold is the deepest gap the prior allows, so the condition can hold even when B exceeds the total poverty gap of the population at hand. If spending per household reached the threshold, the program could lift every household from the deepest possible poverty to the poverty line, and targeting would stop being a rationing problem.

The role of scarcity is to keep the oracle's cutoff, the common expected remaining gap at the oracle allocation, away from zero. A small cutoff means that every household is left with a small expected remaining gap, so every household with a signal in a moderate range must already have received a sizable transfer. Under scarcity, the budget cannot afford sizable transfers to that many households, and such signals occur for a stable share of the population, so a small cutoff is an event of exponentially small probability. By Lemma E.6, there are constants $\underline{\lambda} > 0$ and $c_\lambda > 0$ such that $P_{G,\hat{Y}}(\lambda_{G,+} < \underline{\lambda} \mid \sigma) \leq e^{-c_\lambda n}$ whenever $B/n \leq \bar{b}$. Together, the first assumption controls the average error in the remaining-gap functions, and the second controls the cutoff that converts that error into regret.

B.5.3 Bayes Regret

When the oracle's cutoff is positive, the additional posterior expected loss from using estimated remaining-gap functions in place of the true ones is at most a multiple of their average squared error.

Lemma B.2 (Global stability of the L_+ allocation): *Let Assumptions 2.1, 4.1, and B.1 hold, and let* $0 < B < nz$. *For the capped oracle and empirical Bayes rules defined above, any realization* $(\hat{y}, \sigma)$ *with* $\lambda_{G,+} > 0$ *satisfies*

$$\mathbb{E}_G\left[L_+(\hat{\tau}_+^{\mathrm{EB}}, \mu) - L_+(\tau_{G,+}^{\star}, \mu) \,\middle|\, \hat{y}, \sigma\right] \leq \frac{4z}{\lambda_{G,+}}\, r_+^2(\hat{y}, \sigma).$$

The bound comes from how the posterior expected loss rises when transfers move away from the oracle allocation. At the oracle allocation, every recipient has expected remaining gap $\lambda_{G,+}$, and every household that receives nothing has a gap no larger. A reallocation raises some transfers and lowers others while keeping total spending at B. Because the remaining-gap functions are convex and reach zero at z, raising a transfer by d puts its expected remaining gap at least $\lambda_{G,+} d/z$ below the common level, and lowering a transfer by d puts it at least that far above. The marginal reduction in household loss is twice the expected remaining gap. Valued at $\lambda_{G,+}$, the additions and the cuts cancel, since total spending is unchanged, so the loss from the reallocation is the accumulation of these departures, which grows with the square of the change. For every allocation τ with $0 \le \tau_i \le z$ and $\sum_{i=1}^n \tau_i = B$,

$$\mathbb{E}_G[L_+(\tau,\mu)\,|\,\hat{y},\sigma] - \mathbb{E}_G\big[L_+(\tau^\star_{G,+},\mu)\,\big|\,\hat{y},\sigma\big] \;\ge\; \frac{\lambda_{G,+}}{z}\,\frac{\|\tau-\tau^\star_{G,+}\|^2}{n}.$$

The empirical Bayes allocation minimizes the estimated objective, so its excess true loss is at most the estimation error at the coordinates it moves, and the quadratic growth converts that into the lemma.

The support and likelihood condition of Assumption B.1 bounds the average squared error itself.

Lemma B.3 (Average remaining-gap error): *Let Assumptions 2.1, 4.1, and B.1 hold. There are constants $C_r < \infty$ and $n_r < \infty$, depending only on z, M, $\sigma_{\min}$, $\sigma_{\max}$, and the constant in the approximate-likelihood condition, such that for all $n \ge n_r$,*

$$\mathbb{E}_{G,\hat{Y}}\left[r_+^2(\hat{Y},\sigma)\,\middle|\,\sigma\right] \le \frac{C_r}{\log n} + \frac{C_r}{n^8}.$$

A household's remaining-gap function moves at most in proportion to the Wasserstein distance between the estimated and the true prior, and by Lemma E.5, that distance is of order $1/\sqrt{\log n}$ except on an event of polynomially small probability. Squaring and averaging gives the first term, and the exceptional event gives the second. This is where the rate falls behind the main text. The quadratic-loss bound of Section 4.5 needs only accurate posterior means, which empirical Bayes theory delivers at polynomial rates without recovering G, whereas the remaining-gap functions depend on the whole posterior, and the proof recovers G itself in Wasserstein distance, a Gaussian deconvolution problem whose rates are logarithmic.

Under strict scarcity, the cutoff stays away from zero with high probability, and the two lemmas combine into the regret bound.

Proposition B.3 (Finite-Sample Bayes Regret Bound under L_+): *Let Assumptions 2.1, 4.1, B.1, and B.2 hold, and consider the capped oracle and empirical Bayes rules defined* *above. There are constants $C < \infty$, $c > 0$, and $n_0 < \infty$, none depending on n or on the particular vector σ, such that for all $n \geq n_0$,*

$$\mathrm{BR}_+\big(\hat{\tau}_+^{\mathrm{EB}}, \tau_{G,+}^\star \mid \sigma\big) \leq \frac{C}{\log n} + Ce^{-cn} + \frac{C}{n^8}.$$

In particular, $\mathrm{BR}_+(\hat{\tau}_+^{\mathrm{EB}}, \tau_{G,+}^\star \mid \sigma) = O(1/\log n)$ uniformly over noise-level vectors satisfying Assumption 2.1.

On the event that the cutoff is at least $\underline{\lambda}$, Lemma B.2 bounds conditional regret by $(4z/\underline{\lambda})\, r_+^2$, and Lemma B.3 bounds its expectation. Off the event, conditional regret is at most z^2, since incomes and transfers are nonnegative. By Lemma E.6, the event fails with exponentially small probability. The three terms are the price of Gaussian deconvolution, the probability of a small cutoff, and the failure probability of the Wasserstein guarantee. Without Assumption B.2, the comparison behind Lemma B.2 still bounds conditional regret by $2z\, r_+$, and the same argument gives regret of order $(\log n)^{-1/2}$, so scarcity improves the rate rather than enabling vanishing regret.

C. Fixed Benefit Menus

The main text lets the policymaker give every household its own amount, and each of its allocation rules projects a vector of poverty gaps onto the convex budget set. Many programs instead use a short list of benefit levels set in advance, which simplifies administration and makes costs easy to forecast. The targeting system decides who receives nothing, who receives the basic amount, and who receives more.[22] This appendix extends the quadratic targeting problem to such a menu. The policymaker observes the noisy income estimates and chooses which level, if any, each household receives, subject to the budget. The simplest case is a single common grant, for which only the beneficiary list is chosen, a budget-constrained relative of the selection problems of Chen (2026).

Replacing the budget set by the finite set of menu allocations changes the geometry but not the logic. Under perfect information, the optimal allocation is the menu allocation nearest to the vector of poverty gaps, the oracle is the menu allocation nearest to the vector of posterior mean poverty gaps, and the empirical Bayes rule is the menu allocation nearest to the estimated posterior means. What changes is the cost of estimation error. Projection onto the convex budget set does not increase distances, so in the main text, an error moves the allocation by no more than the error itself, and the regret bound is quadratic in the error. With a menu, a small error can switch a household from one level to another and move the allocation by a full menu step. The switch is cheap in loss, because it happens only when the two allocations are nearly tied, but the regret bound becomes linear in the error rather than quadratic. With a fixed menu and a budget proportional to the population, it is of the order of the square root of the bound in the main text, a weaker guarantee.

C.1 Setup

The decision problem is that of the main text with the budget set replaced by a menu. A menu of payment levels $\mathcal{C} = \{0, c_1, \ldots, c_J\}$ with $0 < c_1 < \cdots < c_J$ is announced before

[22] Georgia's Targeted Social Assistance pays a per-member amount graded by the household's PMT score (Gugushvili & Le Nestour, 2019), and South Africa's Child Support Grant pays a fixed amount per eligible child (Barrientos, 2013).

the signals are observed, zero denotes no benefit, and each household is assigned one level. The feasible allocations are $\mathcal{A}_{\mathcal{C}} := \{\tau \in \mathcal{C}^n \mid \mathbf{1}^\top \tau \leq B\}$, and I assume $c_1 \leq B$, so that some positive level is affordable. The set $\mathcal{A}_{\mathcal{C}}$ is a finite subset of the budget set $\mathcal{A}$ of the main text, so every menu allocation is feasible in the continuous problem, but it is not convex.

Every rule in this section has the form of its continuous counterpart, a nearest point of the feasible set to a vector of poverty gaps. For $v \in \mathbb{R}^n$, let $P_{\mathcal{A}_{\mathcal{C}}}(v)$ denote a menu allocation nearest to v, $P_{\mathcal{A}_{\mathcal{C}}}(v) \in \arg\min_{\tau \in \mathcal{A}_{\mathcal{C}}} \|\tau - v\|^2$. Without the budget, the nearest allocation rounds each coordinate of v to its nearest menu level. When that rounding is unaffordable, the budget couples the households, and the nearest allocation must be chosen jointly. Unlike the projection onto $\mathcal{A}$, this map can send two nearby vectors to allocations a full menu step apart.

C.2 Perfect Information

When the income vector y is observed, the objective is proportional to the squared distance between τ and the poverty gap vector $z\mathbf{1} - y$, as in Section 2.1, so an optimal allocation is $\tau_{\mathcal{C}}^{\mathrm{PI}}(y) = P_{\mathcal{A}_{\mathcal{C}}}(z\mathbf{1} - y)$. Without the budget, the projection rounds each poverty gap to its nearest menu level, and the midpoints between adjacent levels mark where the assignment switches. With the budget, this rounding remains optimal whenever it is affordable. When it is not, some households must be moved down the menu, and the projection chooses which ones jointly, but it never moves a poorer household below a richer one.

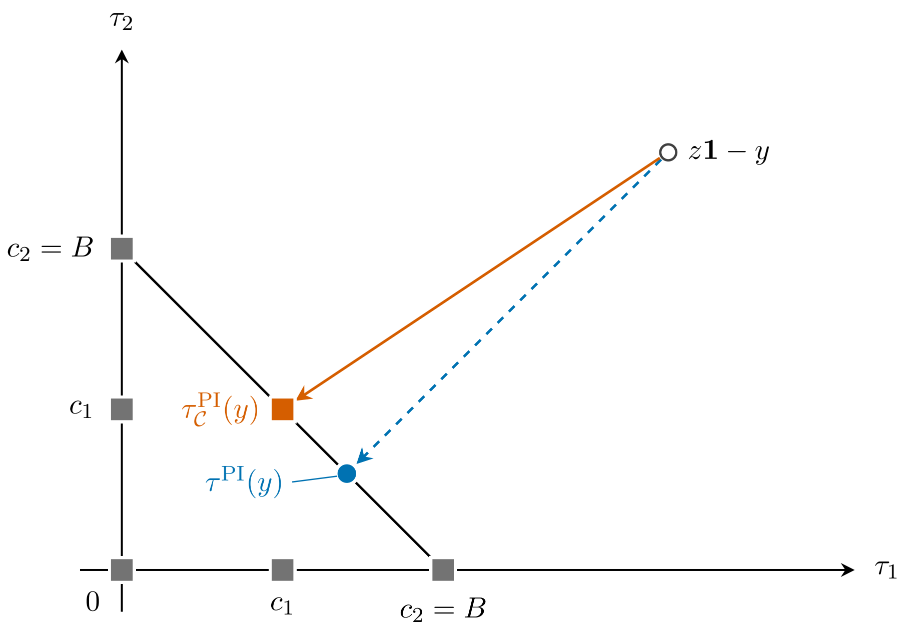
Proposition C.1 (Perfect-Information Menu Allocation): *Relabel households so that $y_1 \leq y_2 \leq \cdots \leq y_n$. An optimal allocation exists.*

(i) If some allocation assigning each household a menu level nearest to its poverty gap is feasible, the optimal allocations are exactly the feasible allocations of this form. In this case, the optimal allocation is unique when no poverty gap lies at a midpoint between adjacent menu levels.

(ii) Every optimal allocation satisfies $\tau_i \geq \tau_k$ whenever $y_i < y_k$, and an optimal allocation with $\tau_1 \geq \tau_2 \geq \cdots \geq \tau_n$ exists.

If a poorer household received less than a richer one, swapping the two payments would keep the cost and lower the loss, and swapping payments between households with equal incomes changes neither. The rationed problem therefore reduces to choosing how many households receive benefits at each level, with the highest level going to the poorest households, the next level to the next poorest, and nothing to the rest. What remains is a tradeoff between depth—raising the level paid to a poorer household—and breadth—extending benefits at a lower level to more households. Its resolution depends on the sizes of the poverty gaps and the spacing of the menu rather than on the income ranking alone.[23] Figure C.1 illustrates how the fixed menu changes the projection for two households.

Figure C.1: Fixed-menu allocation as a projection



Notes. The squares form the feasible menu set $\mathcal{A}_{\mathcal{C}}$; the triangle outlines the continuous feasible set $\mathcal{A}$. Under quadratic loss, the poverty gap vector projects to the blue point under continuous transfers and to the orange square under the fixed menu.

C.3 Oracle Bayes Rule

The oracle knows G and observes the estimates $\hat{y}$ with their standard errors σ. By the posterior loss decomposition in the proof of Lemma 3.1, the posterior expected loss

[23] Interestingly, tightening the budget can bring a household into the program. With menu $\{0, 2, 5\}$ and poverty gaps 5 and 1.1, a budget of 5 pays the poorer household 5, while a budget of 4 pays each household 2.

equals $n^{-1}\|\tau - g_G(\hat{y},\sigma)\|^2$ plus a term that does not depend on τ, so the oracle projects the posterior mean poverty gap vector onto the menu.[24]

Proposition C.2 (Oracle Bayes Menu Rule): *Under Assumptions 2.1 and 4.1, an oracle Bayes rule over $\mathcal{A}_{\mathcal{C}}$ is $\tau^{\star}_{G,\mathcal{C}}(\hat{y},\sigma) = P_{\mathcal{A}_{\mathcal{C}}}\big(g_G(\hat{y},\sigma)\big)$, and a measurable rule with values in $\mathcal{A}_{\mathcal{C}}$ is Bayes if and only if it selects a menu allocation nearest to $g_G(\hat{y},\sigma)$ at almost every realization of the data.*

As in the continuous problem, uncertainty changes only the vector that is projected. Proposition C.1 therefore applies with posterior mean poverty gaps in place of poverty gaps. When assigning each household the level nearest to its posterior mean gap is affordable, the oracle does exactly that, so it first shrinks the estimates and then assigns levels. Otherwise it rations, and households with larger posterior mean gaps receive weakly larger transfers. In either case, a household receives level c_j only if its posterior mean gap is at least $c_j/2$, since below that value, giving it nothing would lower the posterior expected loss and free budget. These comparisons use the magnitudes of the posterior mean gaps, not only their ranking, so shrinkage can change how many households receive benefits at each level even when it leaves the ranking unchanged.[25]

With a single positive benefit level, $\mathcal{C} = \{0, c\}$, only the beneficiary list remains to be chosen.

Corollary C.1 (One Common Grant): *Let $\mathcal{C} = \{0, c\}$. At each realization of the data, an oracle Bayes rule enrolls households in decreasing order of their posterior mean poverty gaps, stopping when the next household's posterior mean poverty gap falls below $c/2$ or when another grant is unaffordable.*

A grant lowers household i's posterior expected loss by $2c\, g_{G,i}(\hat{y}_i,\sigma_i) - c^2$, which is positive exactly when the posterior mean gap exceeds $c/2$, the cutoff in the corollary. Enrollment therefore proceeds down a ranked list and stops for one of two reasons.

[24] In Section 2.2, the convexity of $\mathcal{A}$ lets nonrandomized rules dominate randomized ones at every state. Here, randomization is unnecessary for the Bayes problem, since the posterior risk of a randomized rule averages the posterior risks of the allocations it mixes over.

[25] Under a common standard error, the posterior mean is the same nondecreasing function of every household's estimate, so the plug-in and oracle rules rank households in the same order and differ only in the number assigned to each level. With household-specific standard errors, rank reversals become possible.

Either the next household no longer benefits from the grant, in which case budget is left over although poverty remains, or the budget cannot fund another grant, in which case households who would benefit go unserved. The caseload is part of the solution rather than fixed in advance. The grant is also indivisible, so a recipient whose gap lies between $c/2$ and c receives more than its gap, even under perfect information. Shifting gaps by $c/2$ turns the single-grant case into a selection problem with a maximum caseload, closely related to the selection problems of Chen (2026).

C.4 Feasible Empirical Bayes Rule and Regret

The feasible rule projects the estimated posterior mean poverty gaps $\hat{g}(\hat{y}, \sigma)$ of Section 4, computed under the NPMLE $\hat{G}$ of Section 4.3, onto the menu, $\hat{\tau}^{\mathrm{EB}}_{\mathcal{C}}(\hat{y}, \sigma) = P_{\mathcal{A}_{\mathcal{C}}}\big(\hat{g}(\hat{y}, \sigma)\big)$. Estimating the posterior mean gaps is the problem of the main text, so what changes is how estimation error becomes regret. In Section 4, that step is the projection mismatch inequality of Lemma A.4, which uses the convexity of $\mathcal{A}$ to bound the excess loss by the squared error. The menu is not convex, and the danger comes from households near an assignment boundary. For example, with a single grant and two households whose posterior mean gaps are nearly equal, an arbitrarily small error reverses their order and moves the grant from one to the other, so the allocation jumps by a fixed distance however small the error. Such a switch occurs only when the two allocations are nearly tied under the true posterior mean gaps, which is why its cost stays small. Lemma C.1 makes this precise.

Lemma C.1 (Projection Mismatch Inequality without Convexity): *For any* $u, v \in \mathbb{R}^n$*,*

$$0 \leq \|u - P_{\mathcal{A}_{\mathcal{C}}}(v)\|^2 - \|u - P_{\mathcal{A}_{\mathcal{C}}}(u)\|^2 \leq \quad 2\,\|u - v\|\,\|P_{\mathcal{A}_{\mathcal{C}}}(u) - P_{\mathcal{A}_{\mathcal{C}}}(v)\|.$$

When $u = g_G(\hat{y}, \sigma)$ and $v = \hat{g}(\hat{y}, \sigma)$, $\tau^\star := P_{\mathcal{A}_{\mathcal{C}}}(u)$ and $\hat{\tau} := P_{\mathcal{A}_{\mathcal{C}}}(v)$ denote the oracle and feasible menu allocations, respectively. The middle expression is then n times the posterior expected loss of the feasible allocation in excess of the oracle allocation. The lemma bounds this excess by the estimation error times the change in the allocation. An error therefore costs nothing unless it changes the program, and then costs in proportion to how much it changes it. The menu and the budget limit that change. A menu pays at most c_J per household and serves at most $s_B := \min\{n, \lfloor B/c_1 \rfloor\}$ households, so two allocations in $\mathcal{A}_\mathcal{C}$ differ in at most $2s_B$ coordinates by at most c_J each, and $\|\tau^\star - \hat{\tau}\| \le c_J\sqrt{2s_B}$. Averaging over the data and combining this cap with the bound on the estimated posterior mean gaps behind Proposition 4.2 gives the regret of the feasible rule.

Proposition C.3 (Regret of the Feasible Menu Rule): *Let Assumptions 2.1, 4.1, 4.2, and 4.3 hold, condition on σ, and let Δ_n, $\bar{\sigma}$, and p be as in Proposition 4.2. Write $\hat{\tau}_\mathcal{C}^{\mathrm{EB}}$ for the feasible menu rule and $\tau_{G,\mathcal{C}}^\star$ for an oracle Bayes rule over $\mathcal{A}_\mathcal{C}$, with Bayes regret defined as in Section 4.5. For all n large enough that $\log n > 1/p$,*

$$\mathrm{BR}\big(\hat{\tau}_\mathcal{C}^{\mathrm{EB}}, \tau_{G,\mathcal{C}}^\star \mid \sigma\big) \le 2c_J\sqrt{\frac{2s_B}{n}}\,\sqrt{2\bar{\sigma}\Delta_n + \Delta_n^2}.$$

The lemma bounds regret by the estimation error itself, while the statistical ingredient of the main text controls the average squared error, by $2\bar{\sigma}\Delta_n + \Delta_n^2$. Taking the square root of that bound turns it into a bound on the average error, and this is the square root in the proposition. The bound rises with c_J, which prices each changed assignment, and with s_B, which counts the assignments an error can change. With the menu fixed and a budget that grows with the population, so that the program can serve a nonvanishing share of households, the factor $2c_J\sqrt{2s_B/n}$ is bounded, and the regret bound vanishes at rate $n^{-p/(4(p+1))}$ up to logarithmic factors, against $n^{-p/(2(p+1))}$ in Proposition 4.2. The regret bound still vanishes, but more slowly. For $p = 2$, the two rates are $n^{-1/6}$ and $n^{-1/3}$, and as p grows, they approach $n^{-1/4}$ and $n^{-1/2}$.[26]

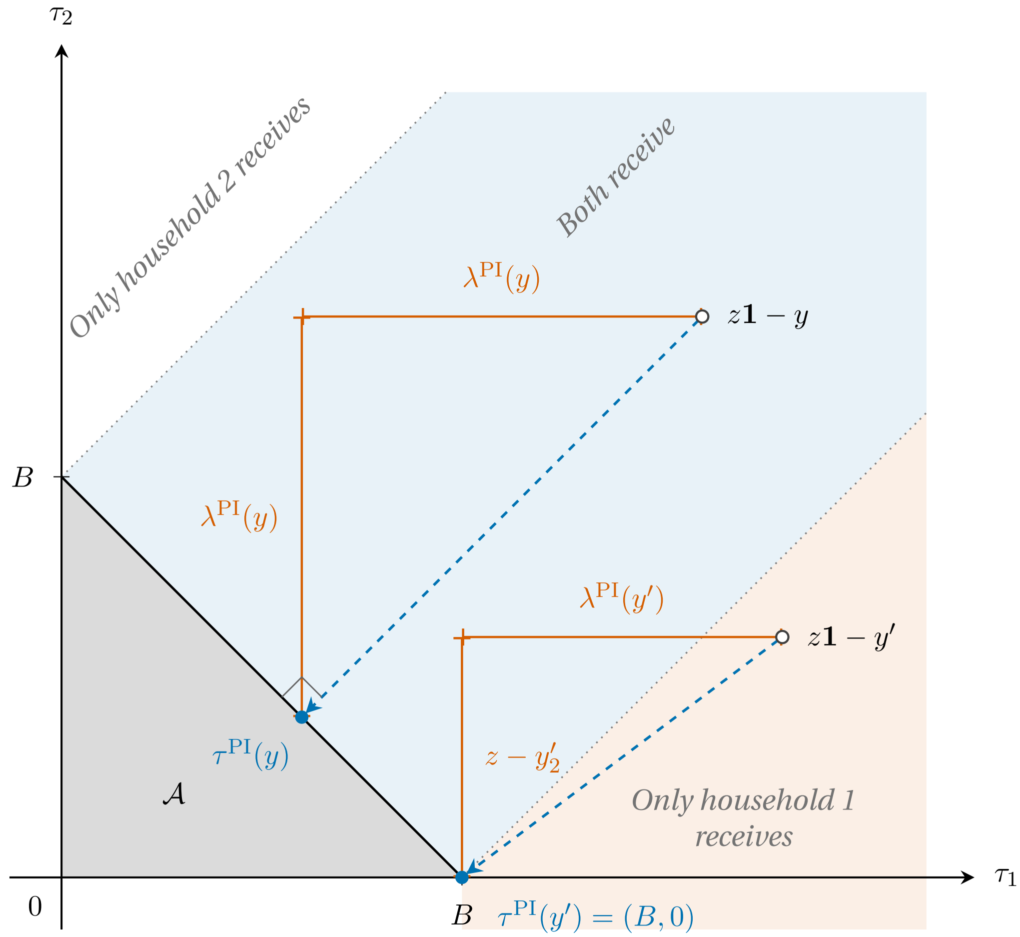
The menu costs accuracy near assignment boundaries but protects it away from them. Under a menu, an error changes the allocation only if it moves the estimated vector far enough to change which allocation is nearest. If the menu allocation nearest to the true posterior mean gaps is unique, every other allocation is strictly farther, and because there are finitely many, the same allocation stays nearest for every vector close enough to the true one. Estimation errors within that margin leave the assignment unchanged and cost nothing.

[26] With B fixed instead, s_B is eventually constant, and the bound is of order $\sqrt{\Delta_n/n}$. The faster rate reflects a bounded program averaged over a growing population rather than an advantage of the menu, since the lemma holds for nearest points of any set and gives a bound of the same order when applied to the continuous budget set, whose diameter is $\sqrt{2}B$.

D. Additional Figures and Tables

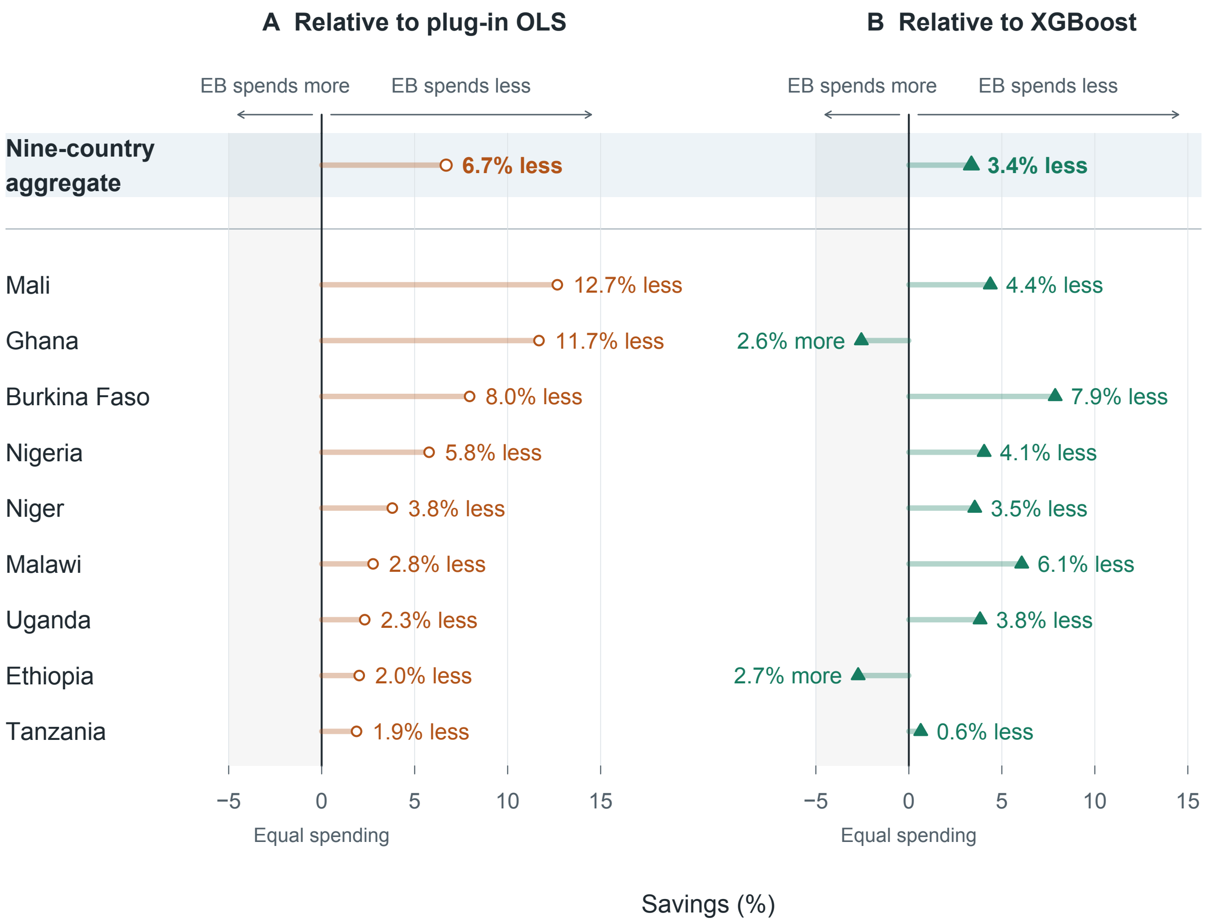
D.1 Figures

Figure D.1: Perfect-information allocation as a projection



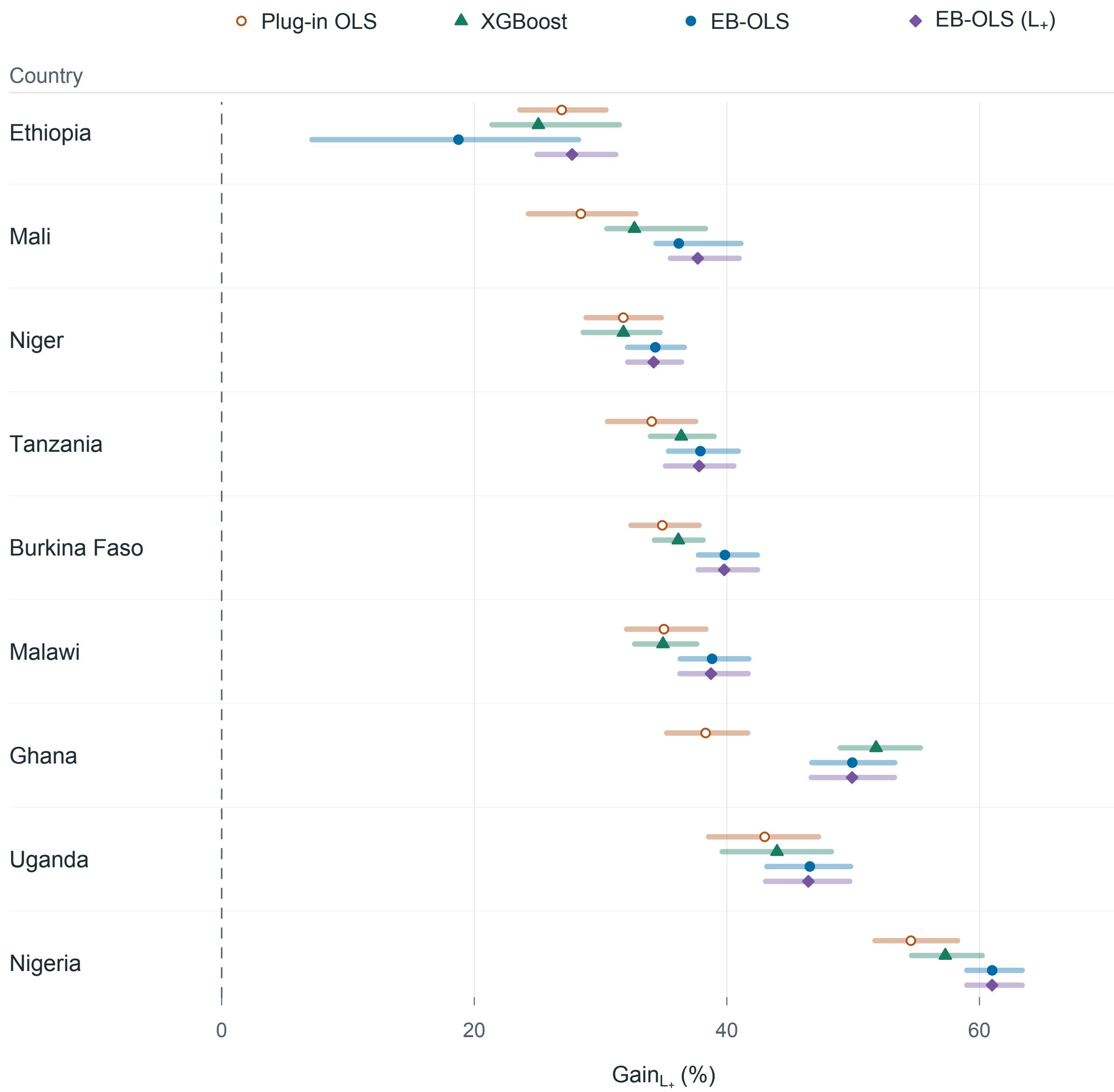
Notes. Under quadratic loss, the gray triangle is the feasible set $\mathcal{A}$. Hollow points are poverty gap vectors; blue points are their projections. Orange segments show the remaining poverty gaps. For y, both households receive transfers, leaving the same gap $\lambda^{\mathrm{PI}}(y)$. For y', $0 < z - y_2' < \lambda^{\mathrm{PI}}(y')$: household 2 is poor but receives nothing.

Figure D.2: Actual transfer spending saved by EB-OLS at equal poverty gap reduction



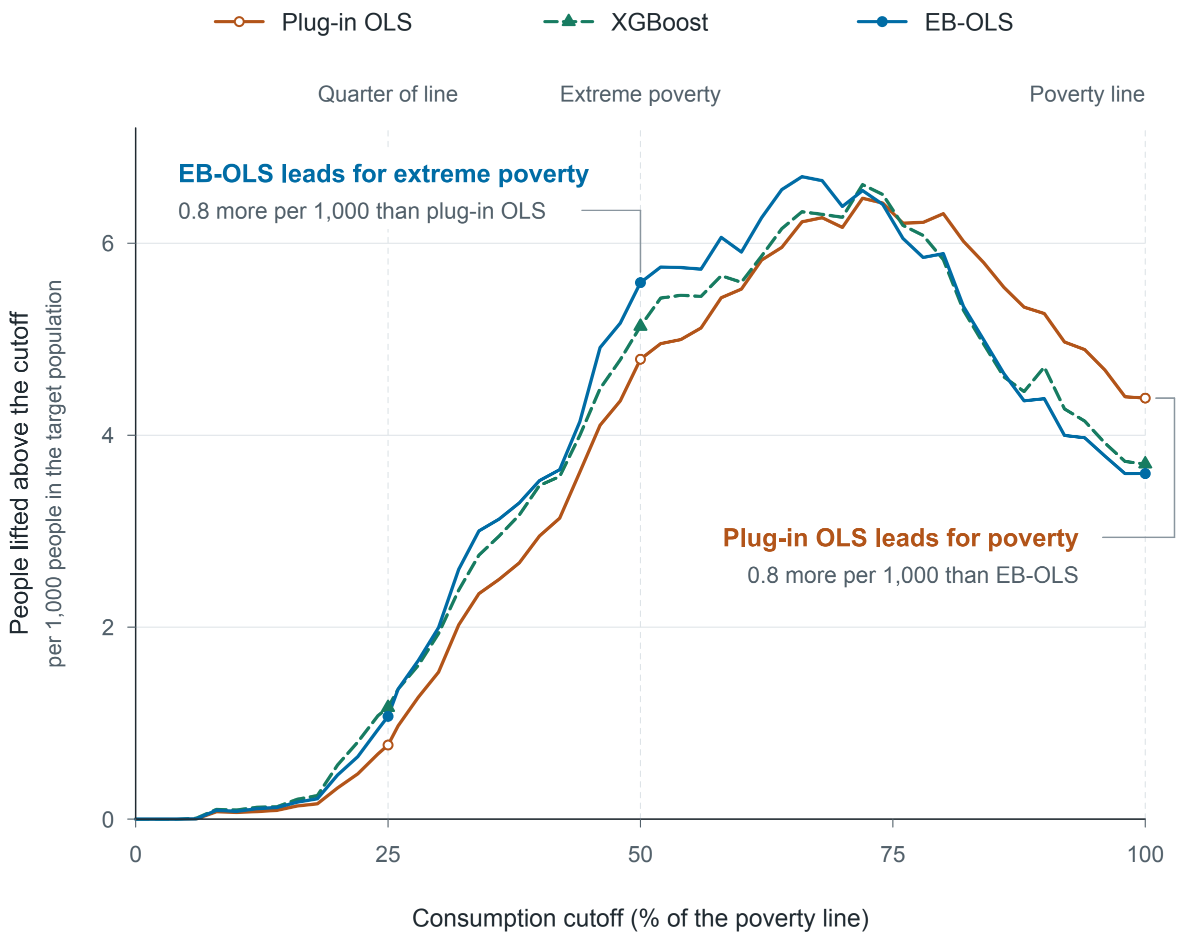
Notes. Points report savings in mean actual transfer spending when EB-OLS matches each comparator's mean realized poverty gap reduction. Comparators face the budget ceiling $b = 0.005$. Savings are measured relative to the comparator's mean actual spending. Positions show signed savings; labels state how much less or more EB-OLS spends than the comparator. Means are computed across simulation replications. The nine-country aggregate uses a common EB-OLS budget ceiling to match the equal-weighted mean poverty gap reduction across countries, with spending also averaged using equal country weights. Countries are ordered by savings relative to plug-in OLS, from highest to lowest.

Figure D.3: Targeting performance under one-sided loss



Notes. Symbols report country-specific mean gains under one-sided loss over 500 simulation replications at the common budget $b = 0.005$. Gains are reported as percentages of the perfect-information improvement over no transfers. A gain of 0% matches the no-transfer benchmark, and 100% matches the perfect-information benchmark, which is outside the displayed range. Pale intervals span the 25th to 75th percentiles across replications. Plug-in OLS, XGBoost, and EB-OLS use the allocations constructed under quadratic loss and are evaluated here under L_+. EB-OLS (L_+) is reoptimized for the one-sided loss. Countries are ordered by mean plug-in OLS gain under L_+, from lowest to highest.

Figure D.4: Rankings by headcount reduction depend on the consumption cutoff



Notes. Each curve reports the number of people whose consumption rises from at or below a given cutoff to above it, per 1,000 people in the entire target population. This equals the reduction in the person-weighted headcount relative to no transfers. Cutoffs are expressed relative to each country's poverty line, z; extreme poverty is defined as consumption at or below $0.5z$. All rules retain their quadratic-loss allocations at the common budget ceiling $b = 0.005$ as the measurement cutoff varies. Results average 500 replications within each of nine countries and give countries equal weight. Lines connect empirical results at intervals of $0.02z$ and at $0.25z$; symbols mark $0.25z$, $0.5z$, and z. Callouts report differences in mean headcount reductions.

D.2 Tables

Table D.1: Country environments and prediction quality

Panel A: Target populations and poverty

Country	Target N	Poverty rate (%)	Mean gap (share of z)	Extreme poor (%)	Near line (%)	Budget / poverty gap (%)
Burkina Faso	9,897	39.9	0.096	2.6	16.7	5.2
Ethiopia	4,693	40.2	0.127	7.9	13.3	3.9
Ghana	4,214	39.9	0.152	12.6	9.4	3.3
Malawi	3,467	40.1	0.118	6.2	12.9	4.2
Mali	2,740	39.9	0.130	8.4	10.0	3.8
Niger	3,359	40.0	0.098	3.1	16.2	5.1
Nigeria	3,814	40.0	0.132	9.1	11.0	3.8
Tanzania	4,282	40.2	0.130	8.4	12.5	3.9
Uganda	2,178	41.1	0.135	9.3	13.0	3.7

Panel B: Overall out-of-sample prediction quality

	R^2		RMSE		Weighted rank correlation	
Country	OLS	XGBoost	OLS	XGBoost	OLS	XGBoost
Burkina Faso	0.496	0.503	0.587	0.583	0.609	0.581
Ethiopia	0.119	0.136	0.818	0.810	0.368	0.362
Ghana	0.389	0.392	1.315	1.312	0.674	0.665
Malawi	0.401	0.373	0.673	0.689	0.566	0.535
Mali	0.365	0.352	1.001	1.012	0.563	0.555
Niger	0.414	0.410	0.488	0.489	0.551	0.544
Nigeria	0.481	0.477	0.643	0.645	0.713	0.704
Tanzania	0.451	0.469	0.785	0.772	0.612	0.607
Uganda	0.414	0.380	0.752	0.774	0.590	0.559

Panel C: Prediction quality in the lower tail

	RMSE among poor people		Predicted poor		Poor correctly classified	
Country	OLS	XGBoost	OLS (%)	XGBoost (%)	OLS (%)	XGBoost (%)
Burkina Faso	0.474	0.477	30.6	23.1	51.3	38.8
Ethiopia	0.677	0.716	18.6	8.8	27.4	13.6
Ghana	0.988	0.991	26.7	21.7	49.9	42.0
Malawi	0.592	0.617	23.7	20.5	40.3	33.7
Mali	0.702	0.693	19.5	11.4	32.7	20.1
Niger	0.403	0.437	32.4	23.8	50.5	38.0
Nigeria	0.505	0.488	31.7	30.6	60.7	58.5
Tanzania	0.598	0.586	31.3	26.0	51.0	43.0
Uganda	0.569	0.592	32.0	27.4	52.9	44.5

Notes: Country-specific simulation means are reported over 500 draws. All quantities use person weights in the held-out target population. In Panel A, mean gap is the average pre-transfer poverty gap expressed as a share of z. Extreme poor is the share of people with consumption below $0.5z$, and near line is the share with consumption between $0.9z$ and $1.1z$. Budget / poverty gap is the common budget $b = 0.005$ as a share of the aggregate pre-transfer poverty gap. In Panel B, the RMSE is expressed relative to z, and weighted rank correlation is the person-weighted Spearman correlation between predicted and realized consumption. In Panel C, predicted poor is the share of people whose predicted consumption is below z. Poor correctly classified is the share of poor people whose predicted consumption is below z.

Table D.2: Robustness of targeting gains

Panel A: Common budget levels

Budget per person (% of z)	Gain (L_2, %) Plug-in OLS	Gain (L_2, %) XGBoost	Gain (L_2, %) EB-OLS	EB advantage (p.p.) over OLS	EB advantage (p.p.) over XGBoost
0.125	18.9	24.0	28.6	9.7	4.6
0.250	17.0	22.6	26.8	9.8	4.2
0.500	14.6	20.7	24.5	9.9	3.8
1.000	11.0	17.7	21.2	10.2	3.5
2.000	5.1	13.4	16.5	11.4	3.1

Panel B: Alternative design choices at the headline budget

Design	Gain (L_2, %) Plug-in OLS	Gain (L_2, %) XGBoost	Gain (L_2, %) Shrinkage rule	Shrinkage-rule advantage (p.p.) over OLS	Shrinkage-rule advantage (p.p.) over XGBoost
Headline specification	14.6	20.7	24.5	9.9	3.8
Standard GLmix support	14.6	20.7	23.4	8.8	2.8
Budget spent in full	14.6	20.6	24.3	9.7	3.7
OLS covariance clustered by enumeration area	14.5	20.9	24.5	10.0	3.6
Person-weighted OLS fit	9.1	20.4	21.8	12.7	1.4
250 training households	−0.3	14.8	17.9	18.2	3.1
500 training households (common target)	14.6	20.6	24.4	9.8	3.8
1,000 training households	22.2	25.6	28.5	6.3	2.9
Training households retained in target	16.2	25.5	26.6	10.4	1.1
Fixed XGBoost hyperparameters	14.5	13.4	24.5	10.0	11.1
Positive-part James–Stein	14.5	20.9	16.1	1.6	−4.8
CLOSE-NPMLE	14.5	20.9	21.3	6.8	0.4

Panel C: Documented program budgets

Country	Program scenario	Budget per person (% of z)	Gain (L_2, %) OLS	Gain (L_2, %) XGBoost	Gain (L_2, %) EB-OLS	EB advantage (p.p.) over best plug-in
Ghana	LEAP 2008 (lower payment)	0.0048	44.1	41.6	45.1	1.0
	LEAP 2008 (upper payment)	0.0090	42.0	41.1	44.5	2.5
	LEAP 2010 plan (lower)	0.0148	40.0	41.2	44.1	2.9
	LEAP 2010 plan (upper)	0.0277	37.2	41.9	43.7	1.8
Malawi	SCTP 2013 (lower bound)	0.0677	29.9	21.9	28.1	−1.9
	SCTP 2015 plan	0.4512	19.8	15.9	22.8	3.0
Mali	Jigisemejiri design	0.2702	−12.2	7.7	11.3	3.6
Niger	PPFS pilot	0.0097	8.6	23.6	33.9	10.4
	PPFS 2017 plan	0.2566	13.1	14.2	19.4	5.2
Tanzania	PSSN phase costing	0.2142	18.2	17.3	19.5	1.2
	PSSN realized-scale costing	0.9425	8.5	10.7	13.7	3.0

Notes: Panels A and C and the first three rows of Panel B use 500 draws per country; the remaining rows of Panel B use 100. Gains are reported as percentages of the perfect-information improvement over no transfers. A gain of 0% matches the no-transfer benchmark, and 100% matches the perfect-information benchmark. Advantages are differences in gains, measured in percentage points (p.p.). Panel A reports equal-weighted averages of country-specific simulation means at different common budget levels. Budget per person is expressed as a percentage of the poverty line and equals $100b$. Panel B reports the corresponding averages under alternative design choices at $b = 0.005$, changing one element of the headline design at a time. "Shrinkage rule" refers to EB-OLS except in the James–Stein and CLOSE-NPMLE rows. James–Stein applies the heteroskedastic positive-part estimator of Xie, Kou, and Brown (2012) to the OLS-estimated poverty gaps before the same weighted projection. The budget-spent-in-full specification requires every rule to exhaust the available budget. Panel C reports country-specific simulation means at documented program budget levels. Its final column gives the difference between EB-OLS and the better-performing plug-in rule. Table D.3 documents the program budget conversions.

Table D.3: Program budgets and their normalized equivalents

		Program evidence			Normalization inputs			
Country	Program scenario	Year	Caseload (households)	Annualized outlay	Population (millions)	Annual poverty line per person	Budget per person (% of z)	Source
Ghana	LEAP 2008 (lower payment)	2008	8,200	GHS 787K	24.25	677	0.0048	[1]
	LEAP 2008 (upper payment)	2008	8,200	GHS 1.48M	24.25	677	0.0090	[1]
	LEAP 2010 plan (lower)	2010	35,000	GHS 3.36M	25.47	894	0.0148	[1]
	LEAP 2010 plan (upper)	2010	35,000	GHS 6.30M	25.47	894	0.0277	[1]
Malawi	SCTP 2013 (lower bound)	2013	30,000	MWK 972M	16.16	88,872	0.0677	[2]
	SCTP 2015 plan	2015	319,000	MWK 10.3B	17.09	134,074	0.4512	[2]
Mali	Jigisemejiri design	2013	62,000	XOF 7.44B	17.46	157,699	0.2702	[3]
Niger	PPFS pilot	2011	2,281	XOF 274M	17.18	165,015	0.0097	[4]
	PPFS 2017 plan	2017	80,000	XOF 9.60B	21.44	174,540	0.2566	[4]
Tanzania	PSSN phase costing	2015	250,000	TZS 63.0B	52.02	565,388	0.2142	[5]
	PSSN realized-scale costing	2015	1,100,000	TZS 277B	52.02	565,388	0.9425	[6]

Notes: Budget per person is $100b = 100T/(Pz)$, where T is the annualized transfer outlay, P is the program-year population, and z is the annual per-person poverty line in program-year prices. The headline simulation budget $b = 0.005$ corresponds to 0.5% of the poverty line. It exceeds the budget in 10 of the 11 documented scenarios and lies below only the Tanzania PSSN realized-scale costing. The PSSN realized-scale costing applies a monthly transfer of TZS 21,000 to the 1.1 million households reached by December 2015 for a full year, yielding an annualized outlay of TZS 277.2B. This is a costing estimate rather than audited annual expenditure. At the World Development Indicators 2015 period-average exchange rate of 1,991 TZS per US dollar, this is about 139 million dollars a year. The conclusion's dollar illustration applies Tanzania's simulated saving at the headline budget, $b = 0.005$, to this outlay; the PSSN costing corresponds to the larger budget $b = 0.009425$. Annualized outlays and poverty lines are expressed in the local currency shown in the outlay column. Poverty lines are converted to program-year prices using World Development Indicators consumer price indices, and population figures come from the same source. K, M, and B denote thousands, millions, and billions. The scenarios include historical bounds, pilots, plans, program designs, and constructed costings, so they should not all be interpreted as audited expenditure. Source numbers link to the underlying documents. The replication files record the source locations and conversion steps.

Table D.4: Targeting outcomes by country

		Gain (%)		Transfer coverage		Use of available budget (%)			
Country	Rule	Gain_{L_2}	Gain_{L_+}	Population treated (%)	Poor people reached (%)	Poverty gaps closed	Overshoot	Leakage	Unspent
Burkina Faso	Plug-in OLS	13.4	34.9	3.7	7.4	68.3	13.5	18.2	0.0
	XGBoost	14.7	36.2	7.4	13.4	68.2	6.9	24.7	0.2
	EB-OLS	24.4	39.8	6.9	13.4	73.8	7.6	18.6	0.0
Ethiopia	Plug-in OLS	−21.3	26.9	3.5	5.6	55.4	10.9	33.7	0.0
	XGBoost	−7.0	25.1	4.9	7.9	49.2	5.7	29.0	16.2
	EB-OLS	−6.8	18.7	4.2	6.8	39.2	6.1	22.2	32.5
Ghana	Plug-in OLS	22.9	38.3	2.0	4.2	72.2	12.4	15.4	0.0
	XGBoost	42.3	51.8	4.2	9.1	83.5	3.6	12.9	0.0
	EB-OLS	39.0	49.9	4.3	8.9	81.5	3.1	15.4	0.0
Malawi	Plug-in OLS	19.2	35.0	3.7	7.0	70.2	9.9	19.9	0.0
	XGBoost	15.5	34.9	5.2	9.2	67.9	4.7	27.4	0.0
	EB-OLS	22.5	38.8	6.5	12.1	72.1	5.0	22.9	0.0
Mali	Plug-in OLS	−12.2	28.4	3.6	6.6	60.6	14.3	25.1	0.0
	XGBoost	6.3	32.7	5.8	10.6	59.7	4.8	25.3	10.1
	EB-OLS	9.5	36.2	6.8	12.5	65.5	4.8	24.3	5.4
Niger	Plug-in OLS	10.2	31.8	5.9	10.4	64.4	9.2	26.4	0.0
	XGBoost	11.6	31.8	8.4	14.3	64.5	6.1	29.3	0.1
	EB-OLS	15.8	34.3	11.8	19.8	66.8	4.7	28.5	0.0
Nigeria	Plug-in OLS	50.0	54.6	3.6	8.3	88.5	5.9	5.5	0.0
	XGBoost	52.9	57.3	5.8	13.1	90.1	2.4	7.5	0.0
	EB-OLS	58.7	61.0	5.1	11.9	93.7	2.4	3.9	0.0
Tanzania	Plug-in OLS	13.2	34.0	3.3	6.1	68.4	9.4	22.2	0.0
	XGBoost	14.0	36.4	6.0	10.7	69.2	5.2	25.6	0.0
	EB-OLS	16.6	37.9	7.9	14.0	69.3	3.7	26.5	0.6
Uganda	Plug-in OLS	36.0	43.0	3.3	7.3	82.8	10.0	7.2	0.0
	XGBoost	35.8	44.0	5.5	10.8	81.6	4.3	14.1	0.0
	EB-OLS	40.7	46.6	5.7	11.7	84.7	4.8	10.5	0.0

Notes: Country-specific simulation means are reported over 500 draws. Outcomes are evaluated at realized consumption in the held-out target population using person weights. The common budget $b = 0.005$ equals 0.5% of the poverty line per represented person. Gains are reported as percentages of the perfect-information improvement over no transfers. A gain of 0% matches the no-transfer benchmark, and 100% matches the perfect-information benchmark. Gain_{L_+} evaluates the same allocations under L_+. Population treated is the share of persons receiving a positive transfer. Poor people reached is the corresponding share among those initially poor. The final four columns divide 100 budget units among poverty gaps closed, overshoot (spending on poor recipients beyond their initial gaps), leakage to initially nonpoor recipients, and unspent resources.

Table D.5: Recipient composition, transfer amounts, and extreme-poverty outcomes by country

		Recipient composition	Transfers among recipients		Extreme poverty	
Country	Rule	Nonpoor recipients (%)	Mean (share of z)	90th percentile (share of z)	People reached (%)	Gap closed (%)
Cross-country mean	Plug-in OLS	21.4	0.152	0.331	12.1	3.0
	XGBoost	24.5	0.087	0.185	17.2	2.6
	EB-OLS	22.7	0.088	0.174	19.6	2.8
Burkina Faso	Plug-in OLS	19.6	0.140	0.309	12.0	3.0
	XGBoost	27.2	0.070	0.155	17.1	2.4
	EB-OLS	21.4	0.083	0.169	19.1	2.8
Ethiopia	Plug-in OLS	35.7	0.148	0.323	7.7	2.1
	XGBoost	35.2	0.084	0.178	10.4	1.5
	EB-OLS	34.3	0.089	0.178	7.7	1.0
Ghana	Plug-in OLS	16.4	0.257	0.557	7.6	3.1
	XGBoost	14.1	0.121	0.255	15.8	3.1
	EB-OLS	16.3	0.127	0.244	15.0	3.0
Malawi	Plug-in OLS	23.1	0.139	0.307	13.7	3.5
	XGBoost	28.8	0.099	0.212	16.3	2.7
	EB-OLS	25.0	0.084	0.172	22.4	3.2
Mali	Plug-in OLS	26.6	0.150	0.320	10.6	2.2
	XGBoost	27.9	0.078	0.163	16.2	2.1
	EB-OLS	27.1	0.076	0.137	20.1	2.5
Niger	Plug-in OLS	29.9	0.086	0.184	16.0	3.3
	XGBoost	31.8	0.061	0.129	18.3	2.4
	EB-OLS	32.1	0.050	0.098	25.6	2.8
Nigeria	Plug-in OLS	7.0	0.140	0.306	17.5	4.4
	XGBoost	9.3	0.088	0.188	25.0	4.1
	EB-OLS	6.0	0.103	0.210	24.4	4.5
Tanzania	Plug-in OLS	25.0	0.156	0.341	9.4	2.4
	XGBoost	27.8	0.086	0.183	15.4	2.2
	EB-OLS	28.2	0.077	0.154	20.5	2.3
Uganda	Plug-in OLS	9.6	0.154	0.332	14.3	3.6
	XGBoost	18.5	0.095	0.204	19.7	3.2
	EB-OLS	13.8	0.104	0.206	21.6	3.4

Notes: Country-specific simulation means are reported over 500 draws. The cross-country mean gives equal weight to each country. Outcomes are evaluated at realized consumption in the held-out target population using person weights. The common budget is $b = 0.005$. Nonpoor recipients is the share of recipients who were initially nonpoor. Mean and 90th-percentile transfers are calculated among recipients and expressed as shares of z. Replications in which a rule makes no transfer are excluded from these conditional transfer-amount statistics. Extreme poor denotes people with pre-transfer consumption below $0.5z$. The final two columns report the share of extremely poor people receiving a positive transfer and the share of their aggregate pre-transfer poverty gap closed.

Table D.6: Empirical Bayes allocations optimized for alternative losses

Panel A: Gain by country and evaluation loss

Country	Gain_{L_2} (%)		Gain_{L_+} (%)	
	$\hat{\tau}^{\text{EB}}$	$\hat{\tau}_+^{\text{EB}}$	$\hat{\tau}^{\text{EB}}$	$\hat{\tau}_+^{\text{EB}}$
Cross-country mean	24.5	23.6	40.4	41.5
Burkina Faso	24.4	24.1	39.8	39.8
Ethiopia	-6.8	-12.0	18.7	27.7
Ghana	39.0	39.0	49.9	49.9
Malawi	22.5	22.4	38.8	38.7
Mali	9.5	8.2	36.2	37.7
Niger	15.8	15.4	34.3	34.2
Nigeria	58.7	58.7	61.0	61.0
Tanzania	16.6	16.1	37.9	37.8
Uganda	40.7	40.5	46.6	46.5

Panel B: Cross-country transfer coverage and use of available budget

Allocation rule	Transfer coverage		Use of available budget (%)			
	Population treated (%)	Poor people reached (%)	Poverty gaps closed	Overshoot	Leakage	Unspent
$\hat{\tau}^{\text{EB}}$	6.6	12.3	71.8	4.7	19.2	4.3
$\hat{\tau}_+^{\text{EB}}$	7.6	13.9	74.1	5.0	20.8	0.0

Notes: Country-specific simulation means are reported over 500 draws. Outcomes are evaluated at realized consumption in the held-out target population using person weights. Both allocation rules use the same OLS signals and fitted restricted GLmix prior at $b = 0.005$. The rule $\hat{\tau}^{\text{EB}}$ minimizes the posterior expected quadratic loss. The rule $\hat{\tau}_+^{\text{EB}}$ minimizes the posterior expected one-sided loss. Panel A evaluates each allocation under both losses. The cross-country mean gives equal weight to each country. Gains are reported as percentages of the perfect-information improvement over no transfers. A gain of 0% matches the no-transfer benchmark, and 100% matches the perfect-information benchmark. Panel B reports equal-weighted averages of country-specific simulation means. Population treated is the share of persons receiving a positive transfer. Poor people reached is the corresponding share among those initially poor. The final four columns divide 100 budget units among poverty gaps closed, overshoot, leakage, and unspent resources.

E. Technical Results and Proofs for the Online Appendix

This appendix collects the supporting results and proofs for Appendices B and C.

E.1 Supporting Lemmas for Appendix B

Lemma E.1 (Capped optimal allocations under L_+): *Under Assumptions 2.1, 4.1, and B.1, suppose* $0 < B < nz$. *Imposing* $\tau_i \le z$ *leaves the optimal values of the oracle and fitted posterior problems with* $\sum_i \tau_i = B$ *unchanged, and both problems admit optimal selections in* $[0, z]^n$ *that are measurable in the data. The oracle selection admits a measurable cutoff* $\lambda_{G,+} \ge 0$ *satisfying the conditions of Proposition B.2, and* $\max_i \tau^\star_{G,+,i} < z$ *whenever* $\lambda_{G,+} > 0$.

Proof of Lemma E.1. Fix $(\hat{y}, \sigma)$ and suppress the conditioning arguments. For $H \in \{G, \hat{G}\}$, let $Q_H(\tau) := n^{-1} \sum_{i=1}^n \mathbb{E}_H[(z - \mu_i - \tau_i)_+^2 \mid \hat{y}_i, \sigma_i]$, the average of household losses. This definition is used again in the proof of Lemma B.2. Because both priors are supported on nonnegative incomes, each household's posterior loss is nonincreasing in its transfer and equals zero at every transfer of at least z.

Take any nonnegative allocation τ with total spending B. Set $c_i = \min\{\tau_i, z\}$ and $e = B - \sum_i c_i \ge 0$. Capping leaves the objective unchanged, and $\sum_i (z - c_i) = nz - B + e > e$ because $B < nz$, so the amount e can be redistributed among coordinates below the cap. The resulting allocation spends B, lies in $[0, z]^n$, and has objective no greater than $Q_H(\tau)$. Hence imposing the cap does not change the optimal value.

On the fixed compact set $\{\tau \in [0, z]^n : \sum_i \tau_i = B\}$, the objective is continuous in τ and measurable in the data, since posterior expectations are ratios of measurable integrals with strictly positive Gaussian-mixture denominators. Both problems therefore admit measurable selections of their minimizers (Brown & Purves, 1973).

The selected oracle allocation also minimizes Q_G over the original set $\{\tau \ge 0 : \sum_i \tau_i = B\}$, a differentiable convex program with the strictly positive feasible point $(B/n)\mathbf{1}$. Apply the KKT conditions to the summed objective nQ_G, whose budget multiplier is ν, and set $\lambda_{G,+} := \nu/2$. By the derivative identity in Lemma E.2, the conditions are

$$g^+_{G,i}(\tau^\star_{G,+,i}) \le \lambda_{G,+}, \qquad g^+_{G,i}(\tau^\star_{G,+,i}) = \lambda_{G,+} \quad \text{if } \tau^\star_{G,+,i} > 0.$$

Since $B > 0$, at least one transfer is positive, so $\lambda_{G,+} = \max_i g^+_{G,i}(\tau^\star_{G,+,i}) \geq 0$, which is measurable in the data. Finally, $g^+_{G,i}(z) = 0$, so a transfer equal to z forces $\lambda_{G,+} = 0$, and a positive cutoff implies that every oracle transfer is strictly below the cap. QED.

Lemma E.2 (Geometry of posterior remaining-gap functions): *Let Assumption 2.1 hold, let H be supported on $[0, M]$, and fix a household and its signal. On $[0, z]$, the posterior remaining-gap function $g^+_{H,i}$ is convex, nonincreasing, and one-Lipschitz, with $g^+_{H,i}(z) = 0$. The household's posterior expected loss is convex and continuously differentiable with derivative $-2g^+_{H,i}(t)$. Moreover,*

$$g^+_{H,i}(u) \leq \frac{z-u}{z-t}\, g^+_{H,i}(t), \qquad 0 \leq t \leq u \leq z, \quad t < z.$$

Proof of Lemma E.2. Under the fixed household posterior, write $Q = z - \mu_i \leq z$ and $g(t) = \mathbb{E}[(Q-t)_+]$. Each hinge $t \mapsto (Q-t)_+$ is convex, nonincreasing, and one-Lipschitz, and vanishes at $t = z$. Taking expectations gives these properties for g. By convexity, $g(u)$ lies below the line joining $(t, g(t))$ and $(z, 0)$, which is the displayed inequality.

The squared hinge is continuously differentiable with derivative $-2(Q-t)_+$, bounded in absolute value by $2z$ on $[0, z]$. Differentiation under the expectation therefore gives the loss derivative $-2g(t)$, which is continuous because g is one-Lipschitz and nondecreasing because g is nonincreasing, so the loss is convex. QED.

Lemma E.3 (Posterior remaining-gap stability in the prior): *Under Assumption 2.1, there is a deterministic function $A : \mathbb{R} \to (0, \infty)$, depending only on z, M, and $\sigma_{\min}$, such that*

$$\sup_{t\in[0,z]} \left| g^+_H(t \mid y, \sigma) - g^+_K(t \mid y, \sigma) \right| \leq A(y)\, W_2(H, K)$$

for all probability distributions H and K supported on $[0, M]$, all $y \in \mathbb{R}$, and all $\sigma \in [\sigma_{\min}, \sigma_{\max}]$. Moreover, for $Z \sim N(0,1)$, $\mathbb{E}[A(\mu + \sigma Z)^2] \leq K_A$ uniformly over $\mu \in [0, M]$ and $\sigma \in [\sigma_{\min}, \sigma_{\max}]$, where $K_A < \infty$ depends only on z, M, $\sigma_{\min}$, and $\sigma_{\max}$.

Proof of Lemma E.3. Fix y, σ, and t, suppress these arguments, and set $h_t(u) = (z - u - t)_+$ and $w(u) = \exp\{yu/\sigma^2 - u^2/(2\sigma^2)\}$. The Gaussian factor that does not depend on u cancels from Bayes' formula, so $g_J^+ = \int h_t w \, dJ / \int w \, dJ$ for $J \in \{H, K\}$. With $c = g_K^+ \in [0, z]$, the identity $\int (h_t - c) w \, dK = 0$ gives

$$g_H^+ - g_K^+ = \frac{\int (h_t - c) w \, d(H - K)}{\int w \, dH}.$$

On $[0, M]$, $|\frac{d}{du} \log w(u)| = |y - u|/\sigma^2 \le (|y| + M)/\sigma_{\min}^2 =: \kappa(y)$, so $\sup w / \inf w \le e^{M\kappa(y)}$ and $\mathrm{Lip}(w) \le \kappa(y) \sup w$, with the extrema taken over $[0, M]$. Since h_t is one-Lipschitz and $|h_t - c| \le z$, $\mathrm{Lip}((h_t - c)w) \le [1 + z\kappa(y)] \sup w$. The Kantorovich–Rubinstein inequality applied to the numerator and $\int w \, dH \ge \inf w$ in the denominator give $|g_H^+ - g_K^+| \le [1 + z\kappa(y)] e^{M\kappa(y)} W_1(H, K) =: A(y) W_1(H, K)$. The bound is uniform in t and σ, and $W_1 \le W_2$, which proves the first claim.

This choice satisfies $A(y) \le C_0(1 + |y|)e^{C_1|y|}$ for finite constants depending only on z, M, and $\sigma_{\min}$. For $\mu \in [0, M]$ and $\sigma \in [\sigma_{\min}, \sigma_{\max}]$, $|\mu + \sigma Z| \le M + \sigma_{\max}|Z|$, so $A(\mu + \sigma Z)^2$ is bounded by $C(1 + |Z|)^2 e^{q|Z|}$ with $C, q < \infty$ depending only on z, M, $\sigma_{\min}$, and $\sigma_{\max}$. This variable has finite expectation under the standard normal law, which gives K_A. QED.

Lemma E.4 (Grid resolution and likelihood tolerance): *Let Assumption 2.1 hold and let G be supported on $[0, M]$. Let $\Gamma_n \subset [0, M]$ be a finite grid containing both endpoints, with mesh δ_n, and let ℓ_n be the average log likelihood of Assumption B.1. Suppose $\hat{G}$ is supported on Γ_n and attains a value of ℓ_n within $\eta_n \ge 0$ of its maximum over distributions supported on Γ_n. Then, for every data realization,*

$$\ell_n(G) - \ell_n(\hat{G}) \le \frac{\delta_n^2}{8\sigma_{\min}^2} + \eta_n.$$

Consequently, choosing $\delta_n^2 + \eta_n \lesssim (\log n)^2/n$ ensures the approximate-likelihood condition in Assumption B.1 for all sufficiently large n.

Proof of Lemma E.4. Fix the observed data. On each grid interval $[a, b]$, round $u \in [a, b]$ to a random endpoint V with $\Pr(V = a) = (b - u)/(b - a)$ and $\Pr(V = b) = (u - a)/(b - a)$, so that $\mathbb{E}[V \mid u] = u$ and $\mathrm{Var}(V \mid u) = (u - a)(b - u) \le \delta_n^2/4$. Let G^Γ be the distribution of V when $u \sim G$. It is supported on Γ_n. For observation i, write $k_i(v) = \phi_{\sigma_i}(\hat{Y}_i - v)$.

Since $\log k_i$ is quadratic in v and the rounding preserves the mean, $\mathbb{E}[\log k_i(V) \mid u] = \log k_i(u) - \mathrm{Var}(V \mid u)/(2\sigma_i^2)$, and Jensen's inequality gives

$$\mathbb{E}[k_i(V) \mid u] \geq \exp\{\mathbb{E}[\log k_i(V) \mid u]\} \geq k_i(u) \exp\{-\delta_n^2/(8\sigma_i^2)\}.$$

Integrating over $u \sim G$, taking logarithms, and averaging over observations yields $\ell_n(G) - \ell_n(G^\Gamma) \leq (\delta_n^2/8n) \sum_{i=1}^n \sigma_i^{-2} \leq \delta_n^2/(8\sigma_{\min}^2)$. Since G^Γ is feasible for the grid problem and $\hat{G}$ is within η_n of its maximum, $\ell_n(\hat{G}) \geq \ell_n(G^\Gamma) - \eta_n$. Combining the two inequalities proves the result. QED.

Lemma E.5 (Support-constrained Wasserstein rate): *Let Assumptions 2.1, 4.1, and B.1 hold. There are constants $C_W < \infty$ and $n_W < \infty$, depending only on M, $\sigma_{\min}$, $\sigma_{\max}$, and C_ℓ, such that for every $n \geq n_W$,*

$$P_{G,\hat{Y}}\left(W_2^2(G, \hat{G}) \leq \frac{C_W}{\log n} \,\middle|\, \sigma\right) \geq 1 - \frac{2}{n^8}.$$

Proof of Lemma E.5. Write $f_{H,\sigma_i} = \phi_{\sigma_i} * H$ and let $\bar{h}_n^2 := n^{-1} \sum_{i=1}^n h^2(f_{G,\sigma_i}, f_{\hat{G},\sigma_i})$ be the average squared Hellinger distance between the true and fitted signal densities. Theorem 7 of Soloff et al. (2025), with $d = 1$, $S = [0, M]$, covariance eigenvalues in $[\sigma_{\min}^2, \sigma_{\max}^2]$, Hellinger localization parameter proportional to $\sqrt{\log n}$, and deviation parameter $\sqrt{8}$, applies to any approximate maximizer whose average log likelihood is within a constant times $(\log n)^2/n$ of $\ell_n(G)$, which is the likelihood condition in Assumption B.1. It gives an event $\mathcal{H}_n$ such that

$$P_{G,\hat{Y}}(\mathcal{H}_n^c \mid \sigma) \leq \frac{2}{n^8}, \qquad \bar{h}_n^2 \leq C_H \frac{(\log n)^2}{n} \quad \text{on } \mathcal{H}_n,$$

where C_H depends only on the constants listed in the lemma.

The proof of their Theorem 10 converts Hellinger accuracy into Wasserstein accuracy by smoothing both priors, applying Nguyen's transport inequality, and controlling the smoothed density difference by Fourier inversion and $\bar{h}_n$. Its only use of the fitted-support condition is the moment bound $\mathcal{R}_s \leq 2^s\{m_s(G) + m_s(\hat{G}) + 2\delta^s m_s(K)\}$ for $s > 2$, in their notation, where the support of $\hat{G}$ is placed in the observed-data bounding box. Here $G([0, M]) = \hat{G}([0, M]) = 1$ gives $m_s(G) \leq M^s$ and $m_s(\hat{G}) \leq M^s$ deterministically, so no additional exceptional event is needed, and equations (34)–(37) of their supplementary material yield, on $\mathcal{H}_n$ for all sufficiently large n and $\bar{h}_n > 0$,

$$W_2^2(G, \hat{G}) \le C \left\{ \frac{1}{-\log \bar{h}_n} + (\log n)^{1/2} \bar{h}_n^{1/12} \right\}.$$

On $\mathcal{H}_n$, $\bar{h}_n \le C_H^{1/2} (\log n)/\sqrt{n}$, so $-\log \bar{h}_n \ge c \log n$ for large n, while the second term is at most $C n^{-1/24} (\log n)^{7/12} = o(1/\log n)$. If $\bar{h}_n = 0$, injectivity of Gaussian convolution gives $\hat{G} = G$. Enlarging C_W and n_W therefore gives $W_2^2(G, \hat{G}) \le C_W / \log n$ on $\mathcal{H}_n$, and the probability bound follows. QED.

Lemma E.6 (Lower tail of the oracle cutoff): *Let Assumptions 2.1, 4.1, B.1, and B.2 hold, and let $\lambda_{G,+}$ be the cutoff selected in Lemma E.1. There are constants $\underline{\lambda} > 0$ and $c_\lambda > 0$, depending only on G, z, M, $\sigma_{\min}$, $\sigma_{\max}$, and $\bar{b}$, such that whenever $0 < B/n \le \bar{b}$,*

$$P_{G,\hat{Y}}(\lambda_{G,+} < \underline{\lambda} \mid \sigma) \le e^{-c_\lambda n}.$$

Proof of Lemma E.6. Write $\mu_0 = \inf \operatorname{supp}(G)$, choose t with $\bar{b} < t < z - \mu_0$, and choose $R > M$ large enough that $p := \Pr\{|Z| \le (R - M)/\sigma_{\max}\} > \bar{b}/t$ for $Z \sim N(0,1)$. The indicators $I_i := \mathbb{1}\{|\hat{Y}_i| \le R\}$ are independent conditional on σ, and since $|\hat{Y}_i| \le M + \sigma_{\max}|Z_i|$, each has $\mathbb{E}[I_i \mid \sigma] \ge p$.

Because $\mu_0 \in \operatorname{supp}(G)$ and $\mu_0 < z - t$, the prior puts positive mass on incomes whose remaining gap after transfer t is positive, and the Gaussian likelihood is strictly positive, so $g_G^+(t \mid y, s) > 0$ for every signal y and every $s \in [\sigma_{\min}, \sigma_{\max}]$. This function is continuous in (y, s), so its minimum m over $|y| \le R$ and $s \in [\sigma_{\min}, \sigma_{\max}]$ is attained and positive. Set $\underline{\lambda} = m$.

On the event $\{\lambda_{G,+} < \underline{\lambda}\}$, the conditions of Proposition B.2 give $g_{G,i}^+(\tau_{G,+,i}^\star \mid \hat{Y}_i, \sigma_i) \le \lambda_{G,+} < m$ for every household. If $I_i = 1$, the household's remaining-gap function at t is at least m, so monotonicity forces $\tau_{G,+,i}^\star > t$. Since the oracle spends B, the event implies $n^{-1} \sum_{i=1}^n I_i < B/(nt) \le \bar{b}/t$. Hoeffding's inequality for the independent $[0,1]$-valued I_i therefore yields

$$P_{G,\hat{Y}}(\lambda_{G,+} < \underline{\lambda} \mid \sigma) \le P_{G,\hat{Y}}\left(\frac{1}{n} \sum_{i=1}^n I_i \le \frac{\bar{b}}{t} \,\middle|\, \sigma \right) \le \exp\left\{ -2n \left(p - \frac{\bar{b}}{t} \right)^2 \right\},$$

and $c_\lambda = 2(p - \bar{b}/t)^2 > 0$ proves the result. QED.

E.2 Proofs of Appendix B Results

This subsection contains the proofs of the results stated in Appendix B.

Proof of Lemma B.1. Fix a household i and condition on $(\mu, \hat{Y})$. Write $a_i := z - \mu_i - \delta_i(\hat{Y})$, which is fixed given $(\mu, \hat{Y})$, and note that $z - y_i - \delta_i(\hat{Y}) = a_i - \varepsilon_i$. Since $x \mapsto x_+^2$ is convex, conditional Jensen and $\mathbb{E}[\varepsilon_i \mid \mu, \hat{Y}] = 0$ give $\mathbb{E}[(a_i - \varepsilon_i)_+^2 \mid \mu, \hat{Y}] \geq \big(a_i - \mathbb{E}[\varepsilon_i \mid \mu, \hat{Y}]\big)_+^2 = (a_i)_+^2$. For the upper bound, the map $x \mapsto x_+^2$ is differentiable with derivative $2x_+$, which is 2-Lipschitz. The descent inequality therefore gives, for all $a, e \in \mathbb{R}$, $(a - e)_+^2 \leq a_+^2 - 2a_+e + e^2$. Setting $a = a_i$ and $e = \varepsilon_i$ and taking conditional expectations eliminates the linear term, so $\mathbb{E}[(a_i - \varepsilon_i)_+^2 \mid \mu, \hat{Y}] \leq (a_i)_+^2 + \mathbb{E}[\varepsilon_i^2 \mid \mu, \hat{Y}]$. Averaging over i, taking expectations over $\hat{Y}$, and applying the law of iterated expectations yields (8). This proves the lemma. QED.

Proof of Corollary B.1. For brevity, let $q_i := (z - y_i)_+$. Fix any feasible $\tau \in \mathcal{A}$, and define $\tilde{\tau}_i := \min\{\tau_i, q_i\}$ for each i. Since $\tilde{\tau}_i \leq \tau_i$ for every i, the vector $\tilde{\tau}$ remains feasible. Moreover, replacing τ_i by $\tilde{\tau}_i$ does not change the term $(z - y_i - \tau_i)_+$: if $\tau_i \leq q_i$, nothing changes, while if $\tau_i > q_i$, both $(z - y_i - \tau_i)_+$ and $(z - y_i - \tilde{\tau}_i)_+$ are equal to zero. Hence there exists an optimal solution satisfying $0 \leq \tau_i \leq q_i$ for all i.

On this restricted region, minimizing L_+ is equivalent to minimizing $\frac{1}{n}\sum_{i=1}^n (q_i - \tau_i)^2$ over $\mathcal{A}$. Since $q_i = (z - y_i)_+$, this differs from the quadratic objective in Proposition 2.1 only by the additive constant $\frac{1}{n}\sum_{i: y_i \geq z}(z - y_i)^2$, which does not depend on τ. Therefore $\tau^{\mathrm{PI}}(y)$ remains optimal under L_+.

If $B < \sum_{i=1}^n q_i$, every optimum lies in the restricted region. Suppose instead that an optimal τ has $\tau_j > q_j$ for some j. Then $\sum_{i \neq j} \tau_i \leq B - \tau_j < \sum_{i \neq j} q_i$, so some household $k \neq j$ has $\tau_k < q_k$. Moving an amount $\delta \in (0, \min\{\tau_j - q_j,\, q_k - \tau_k\}]$ from j to k keeps the allocation feasible, leaves the term for j at zero, and strictly lowers the term for k, contradicting optimality. On the restricted region, L_+ is a strictly convex quadratic plus a constant and the region is convex, so the minimizer is unique.

If $B \geq \sum_{i=1}^n q_i$, choosing $\tau_i = q_i$ for all i is feasible and yields $L_+ = 0$. Hence the minimum value is zero. When $B = \sum_{i=1}^n q_i$, zero loss requires $\tau_i \geq q_i$ for all i, and feasibility then forces $\tau_i = q_i$ exactly, so the minimizer remains unique at the boundary. When $B > \sum_{i=1}^n q_i$, the set of minimizers is not a singleton, since additional transfers to households already at or above the poverty line do not change the loss. QED.

Proof of Proposition B.1. Let $e(x) := \sum_{i=1}^n \big(\tau_i^{\text{plug}}(x) - z\big)_+$. For feasibility, fix any $x \in \mathbb{R}^n$. By construction, each component of $\tau^{\text{cs}}(x)$ is nonnegative. Moreover, using $\min\{a, z\} + (a - z)_+ = a$ for any $a \geq 0$, we have

$$\sum_{i=1}^n \tau_i^{\text{cs}}(x) = \sum_{i=1}^n \min\{\tau_i^{\text{plug}}(x), z\} + e(x) = \sum_{i=1}^n \tau_i^{\text{plug}}(x) \leq B.$$

Hence $\tau^{\text{cs}}(x) \in \mathcal{A}$ for every x.

To prove part (i), fix $\theta \in (-\infty, z]^n$ and $x \in \mathbb{R}^n$. If $\tau_i^{\text{plug}}(x) > z$, then $\theta_i \leq z < \tau_i^{\text{plug}}(x)$ and also $\tau_i^{\text{cs}}(x) \geq z \geq \theta_i$, so both rules generate zero loss in coordinate i. If instead $\tau_i^{\text{plug}}(x) \leq z$, then $\tau_i^{\text{cs}}(x) = \tau_i^{\text{plug}}(x) + e(x)/n \geq \tau_i^{\text{plug}}(x)$, and since $t \mapsto (\theta_i - t)_+^2$ is nonincreasing, the loss in coordinate i weakly falls. Therefore, $L_+(\tau^{\text{cs}}(x), \theta) \leq L_+(\tau^{\text{plug}}(x), \theta) \qquad \text{for every } x.$ Taking expectations yields $R_+(\tau^{\text{cs}}, \theta) \leq R_+(\tau^{\text{plug}}, \theta)$.

For part (ii), fix $\theta \in (-\infty, z]^n$ such that $\sum_{i=1}^n (\theta_i)_+ > B$, and consider any realization x with $e(x) > 0$. Since $\sum_{i=1}^n \tau_i^{\text{plug}}(x) \leq B < \sum_{i=1}^n (\theta_i)_+$, there must exist some coordinate m such that $\tau_m^{\text{plug}}(x) < (\theta_m)_+$. Because $\tau_m^{\text{plug}}(x) \geq 0$, this implies $\theta_m > 0$, hence $(\theta_m)_+ = \theta_m \leq z$. Therefore coordinate m is not capped, and $\tau_m^{\text{cs}}(x) = \tau_m^{\text{plug}}(x) + \frac{e(x)}{n} > \tau_m^{\text{plug}}(x)$. Since $t \mapsto (\theta_m - t)_+^2$ is strictly decreasing for $t < \theta_m$, the loss in coordinate m falls strictly, while every other coordinate improves weakly by part (i). Hence $L_+(\tau^{\text{cs}}(x), \theta) < L_+(\tau^{\text{plug}}(x), \theta) \qquad \text{whenever } e(x) > 0.$

It remains to show that $e(X) > 0$ with positive probability. Since $B > z$, any x with $x_1 > B$ and $x_j < 0$ for all $j \geq 2$ satisfies $x_+ = (x_1, 0, \ldots, 0)$, so its projection onto $\mathcal{A}$ is $(B, 0, \ldots, 0)$. On this event, $e(x) = B - z > 0$. Under Assumption 2.1, the distribution of X has a density that is strictly positive on $\mathbb{R}^n$, so this event has positive probability for every θ. Therefore $\Pr_\theta(e(X) > 0) > 0$.

Combining this with the pointwise strict inequality on $\{e(X) > 0\}$ and weak inequality elsewhere gives $R_+(\tau^{\text{cs}}, \theta) < R_+(\tau^{\text{plug}}, \theta)$. This proves part (ii). The final inadmissibility claim follows immediately from the definition of admissibility. QED.

Proof of Proposition B.2. Fix $(\hat{y}, \sigma)$. Write $g^+_{G,i}(t) = g^+_{G,i}(t \mid \hat{y}_i, \sigma_i)$ and $\psi_i(t) = \mathbb{E}_G[(z - \mu_i - t)^2_+ \mid \hat{y}_i, \sigma_i]$ for $t \geq 0$.

For each realization of μ_i, the map $t \mapsto (z - \mu_i - t)^2_+$ is convex and continuously differentiable on $\mathbb{R}_+$, with derivative $-2(z - \mu_i - t)_+$. The maintained second-moment condition implies $\mathbb{E}_G[(z - \mu_i)_+ \mid \hat{y}_i, \sigma_i] < \infty$, and $2(z - \mu_i)_+$ is an integrable envelope for the relevant difference quotients. Dominated convergence therefore justifies differentiating under the conditional expectation, yielding $\psi'_i(t) = -2g^+_{G,i}(t)$.

The feasible set $\{\tau \in \mathbb{R}^n_+ : \sum_i \tau_i = B\}$ is nonempty and compact, and $\sum_i \psi_i$ is continuous and convex, so an optimum exists. Slater's condition holds at $\tau = (B/n)\mathbf{1}$, so the KKT conditions are necessary and sufficient. Since the factor $1/n$ does not affect the minimizer, work with $\sum_i \psi_i(\tau_i)$. There exist $\nu \in \mathbb{R}$ and $\xi_i \geq 0$ such that, for every i,

$$\psi'_i(\tau^\star_{G,+,i}) + \nu - \xi_i = 0, \qquad \xi_i \tau^\star_{G,+,i} = 0, \qquad \tau^\star_{G,+,i} \geq 0, \qquad \sum_i \tau^\star_{G,+,i} = B.$$

Substituting $\psi'_i(t) = -2g^+_{G,i}(t)$ gives $2g^+_{G,i}(\tau^\star_{G,+,i}) = \nu - \xi_i \leq \nu$. Write $\lambda_{G,+} := \nu/2$, so that $g^+_{G,i}(\tau^\star_{G,+,i}) \leq \lambda_{G,+}$. When $\tau^\star_{G,+,i} > 0$, complementary slackness gives $\xi_i = 0$, so $g^+_{G,i}(\tau^\star_{G,+,i}) = \lambda_{G,+}$. If $\tau^\star_{G,+,i} = 0$, the same display gives $g^+_{G,i}(0) \leq \lambda_{G,+}$. Since $B > 0$ and $\sum_i \tau^\star_{G,+,i} = B$, at least one coordinate is active. For any active coordinate, $\lambda_{G,+} = g^+_{G,i}(\tau^\star_{G,+,i}) \geq 0$. This proves the KKT characterization.

The cutoff $\lambda_{G,+}$ is determined by $\sum_i \tau^\star_{G,+,i} = B$. QED.

Proof of Lemma B.2. Fix $(\hat{y}, \sigma)$, hold the fitted prior $\hat{G} = \hat{G}(\hat{y}, \sigma)$ fixed, and let Q_H be as defined in the proof of Lemma E.1, for $H \in \{G, \hat{G}\}$. Under the true model, household independence makes $Q_G(\tau)$ the posterior expected loss of τ conditional on the full data. Suppress the signal arguments and abbreviate $\tau^\star = \tau^\star_{G,+}$, $\hat{\tau} = \hat{\tau}^{\mathrm{EB}}_+$. The lemma assumes $\lambda_{G,+} > 0$. By Lemma E.1, both allocations belong to $\{\tau \in [0, z]^n : \sum_i \tau_i = B\}$, every $\tau^\star_i < z$, and $g^+_{G,i}(\tau^\star_i) \leq \lambda_{G,+}$ with equality whenever $\tau^\star_i > 0$.

We first establish quadratic growth away from the oracle allocation. For any τ in the capped budget set, let $d_i = \tau_i - \tau^\star_i$. Since $\sum_i d_i = 0$, the derivative identity in Lemma E.2 gives

$$n\{Q_G(\tau) - Q_G(\tau^\star)\} = 2\sum_{i=1}^n \int_{\tau^\star_i}^{\tau_i} \{\lambda_{G,+} - g^+_{G,i}(u)\}\, du.$$

If $d_i \geq 0$, the chord inequality in Lemma E.2 gives, for $0 \leq v \leq d_i$, $g^+_{G,i}(\tau^\star_i + v) \leq g^+_{G,i}(\tau^\star_i)(z-$

$\tau_i^\star - v)/(z - \tau_i^\star) \le \lambda_{G,+}(1 - v/z)$. If $d_i < 0$, then $\tau_i^\star > 0$, so $g^+_{G,i}(\tau_i^\star) = \lambda_{G,+}$, and the chord inequality with $t = \tau_i^\star - v$ and $u = \tau_i^\star$ rearranges to $g^+_{G,i}(\tau_i^\star - v) \ge \lambda_{G,+}(z - \tau_i^\star + v)/(z - \tau_i^\star) \ge \lambda_{G,+}(1 + v/z)$ for $0 \le v \le -d_i$. Integrating, with reversed limits when $d_i < 0$, bounds each coordinate's contribution to the display from below by $\lambda_{G,+} d_i^2/z$. Summing yields

$$Q_G(\tau) - Q_G(\tau^\star) \ge \frac{\lambda_{G,+}}{z} \frac{\|\tau - \tau^\star\|^2}{n}.$$

We next compare the true and fitted objectives. Let $\varepsilon_i := \sup_{t \in [0,z]} |\hat{g}_i^+(t) - g^+_{G,i}(t)|$ and $q^2 := \|\hat\tau - \tau^\star\|^2/n$, so that $r_+^2 = n^{-1}\sum_i \varepsilon_i^2$. Convexity of Q_G gives $Q_G(\hat\tau) - Q_G(\tau^\star) \le \langle \nabla Q_G(\hat\tau), \hat\tau - \tau^\star\rangle$, and optimality of $\hat\tau$ for $Q_{\hat G}$ on the same convex set gives $\langle \nabla Q_{\hat G}(\hat\tau), \hat\tau - \tau^\star\rangle \le 0$. Subtracting, the derivative identity bounds the ith coordinate of $\nabla Q_G(\hat\tau) - \nabla Q_{\hat G}(\hat\tau)$ by $2\varepsilon_i/n$ in absolute value, and the Cauchy–Schwarz inequality gives

$$Q_G(\hat\tau) - Q_G(\tau^\star) \le \frac{2}{n}\sum_{i=1}^n \varepsilon_i |\hat\tau_i - \tau_i^\star| \le 2r_+ q.$$

Applying the growth bound at $\tau = \hat\tau$ gives $(\lambda_{G,+}/z)\, q^2 \le 2r_+ q$. If $q = 0$, the regret is zero. Otherwise $q \le 2zr_+/\lambda_{G,+}$, and substitution gives $Q_G(\hat\tau) - Q_G(\tau^\star) \le 4zr_+^2/\lambda_{G,+}$, as claimed. QED.

Proof of Lemma B.3. Fix σ. Since both priors are supported on $[0, M]$, Lemma E.3 applied household by household gives

$$r_+^2(\hat Y, \sigma) \le W_2^2(G, \hat G)\, \frac{1}{n}\sum_{i=1}^n A(\hat Y_i)^2$$

for every data realization, including when $\hat G$ is estimated from the same signals. Both remaining-gap functions take values in $[0, z]$, so also $r_+^2 \le z^2$. Averaging the moment bound of Lemma E.3 over $\mu_i \sim G$ gives $\mathbb{E}[A(\hat Y_i)^2 \mid \sigma] \le K_A$ for every household. For $n \ge \max\{n_W, 2\}$, let $\mathcal{W}_n := \{W_2^2(G, \hat G) \le C_W/\log n\}$, with C_W and n_W from Lemma E.5, so that $P_{G,\hat Y}(\mathcal{W}_n^c \mid \sigma) \le 2/n^8$. Splitting the expectation over $\mathcal{W}_n$ and its complement,

$$\mathbb{E}_{G,\hat Y}[r_+^2(\hat Y, \sigma) \mid \sigma] \le \frac{C_W}{\log n}\frac{1}{n}\sum_{i=1}^n \mathbb{E}[A(\hat Y_i)^2 \mathbb{1}_{\mathcal{W}_n} \mid \sigma] + z^2 P_{G,\hat Y}(\mathcal{W}_n^c \mid \sigma) \le \frac{C_W K_A}{\log n} + \frac{2z^2}{n^8},$$

where the second inequality drops the indicator and needs no independence between the event and the signals. Taking $C_r = \max\{C_W K_A, 2z^2\}$ and $n_r = \max\{n_W, 2\}$ proves the claim, and the dependence of the constants follows from Lemmas E.3 and E.5. QED.

Proof of Proposition B.3. Choose $n_0 \geq \max\{n_r, 2\}$ such that $0 < B/n \leq \bar{b}$ for every $n \geq n_0$. Since $\bar{b} < z - \inf \operatorname{supp}(G) \leq z$, also $B < nz$. Fix $n \geq n_0$ and σ, and let $\Delta(\hat{Y}) := \mathbb{E}_G[L_+(\hat{\tau}_+^{\mathrm{EB}}, \mu) - L_+(\tau_{G,+}^\star, \mu) \mid \hat{Y}, \sigma]$. Oracle optimality gives $\Delta(\hat{Y}) \geq 0$, nonnegative incomes and transfers give $0 \leq L_+ \leq z^2$ and hence $\Delta(\hat{Y}) \leq z^2$, and iterated expectations give $\mathrm{BR}_+(\hat{\tau}_+^{\mathrm{EB}}, \tau_{G,+}^\star \mid \sigma) = \mathbb{E}[\Delta(\hat{Y}) \mid \sigma]$. Let $\underline{\lambda}$ and c_λ be supplied by Lemma E.6 and set $\mathcal{S}_n = \{\lambda_{G,+} \geq \underline{\lambda}\}$. On $\mathcal{S}_n$, Lemma B.2 gives $\Delta(\hat{Y}) \leq (4z/\underline{\lambda})\, r_+^2(\hat{Y}, \sigma)$. Splitting the expectation over $\mathcal{S}_n$ and its complement, dropping the indicator from the nonnegative curve-error term, and applying Lemmas B.3 and E.6,

$$\mathbb{E}[\Delta(\hat{Y}) \mid \sigma] \leq \frac{4z}{\underline{\lambda}} \mathbb{E}[r_+^2(\hat{Y}, \sigma) \mid \sigma] + z^2 P_{G,\hat{Y}}(\mathcal{S}_n^c \mid \sigma) \leq \frac{4zC_r}{\underline{\lambda}} \left(\frac{1}{\log n} + \frac{1}{n^8} \right) + z^2 e^{-c_\lambda n}.$$

Taking $C = \max\{4zC_r/\underline{\lambda}, z^2\}$ and $c = c_\lambda$ gives the stated bound. For fixed G and budget sequence, these constants and n_0 are uniform over noise-level vectors satisfying Assumption 2.1. QED.

E.3 Proofs of Appendix C Results

Proof of Proposition C.1. The zero allocation is feasible and $\mathcal{A}_\mathcal{C}$ is finite, so an optimal allocation exists. For part (i), n times the loss of any allocation is at least $\sum_{i=1}^n \min_{c \in \mathcal{C}} (z - y_i - c)^2$, with equality exactly when every household receives a menu level nearest to its poverty gap. If some allocation of this form is feasible, the optimal allocations are therefore exactly the feasible allocations of this form. When no poverty gap lies at a midpoint between adjacent menu levels, each household has a unique nearest level, so exactly one such allocation exists. For part (ii), suppose an optimal allocation τ has $y_i < y_k$ and $\tau_i < \tau_k$. Exchanging the two transfers preserves the multiset of assigned levels and the total cost, hence feasibility, and changes n times the loss by

$$(z - y_i - \tau_k)^2 + (z - y_k - \tau_i)^2 - (z - y_i - \tau_i)^2 - (z - y_k - \tau_k)^2 = -2(y_k - y_i)(\tau_k - \tau_i) < 0,$$

contradicting optimality. Within a group of equal incomes, reordering the transfers in nonincreasing order changes neither the loss nor the cost, so an optimal allocation with

$\tau_1 \geq \cdots \geq \tau_n$ exists. QED.

Proof of Proposition C.2. By the posterior loss decomposition in the proof of Lemma 3.1, posterior expected loss equals $n^{-1}\|\tau - g_G(\hat{y}, \sigma)\|^2$ plus a term that does not depend on τ, so at each realization the allocations that minimize posterior expected loss are the menu allocations nearest to $g_G(\hat{y}, \sigma)$. Throughout, when several menu allocations are equally close to a vector, $P_{\mathcal{A}_\mathcal{C}}$ selects the first of them in a fixed enumeration of $\mathcal{A}_\mathcal{C}$. Since $g_G(\cdot, \sigma)$ is measurable, the set of realizations at which a given allocation is nearest is measurable, being defined by finitely many comparisons of measurable functions, so $P_{\mathcal{A}_\mathcal{C}}(g_G(\hat{y}, \sigma))$ is a measurable rule. Every rule with values in $\mathcal{A}_\mathcal{C}$ has finite Bayes risk, since transfers are bounded and G has a finite second moment. By iterated expectations, a rule that minimizes posterior expected loss at almost every realization is therefore Bayes (Berger, 1985), and a rule that fails to do so on a set of realizations of positive probability has strictly larger Bayes risk. QED.

Proof of Corollary C.1. By Proposition C.2, it suffices to minimize $\|\tau - g_G(\hat{y}, \sigma)\|^2$ over $\mathcal{A}_\mathcal{C}$ at each realization. Write $\tau_i = cd_i$ with $d_i \in \{0, 1\}$ and $g_i = g_{G,i}(\hat{y}_i, \sigma_i)$. Relative to no transfer, enrolling household i lowers n times posterior expected loss by $g_i^2 - (g_i - c)^2 = 2cg_i - c^2$, and the budget allows at most $\lfloor B/c \rfloor$ grants, so the problem is to maximize $\sum_i d_i(2cg_i - c^2)$ subject to $\sum_i d_i \leq \min\{n, \lfloor B/c \rfloor\}$. Since every grant has the same cost, taking the largest gains first and stopping at capacity or when the next gain is negative is optimal, which is the stated rule. A gain of zero, at $g_i = c/2$, leaves the objective unchanged, so enrolling such households while capacity remains is also optimal. Breaking ranking ties by household index makes the rule measurable, as in the proof of Proposition C.2. QED.

Proof of Lemma C.1. Write $p := P_{\mathcal{A}_\mathcal{C}}(u)$ and $q := P_{\mathcal{A}_\mathcal{C}}(v)$. Since q is nearest to v in $\mathcal{A}_\mathcal{C}$ and p belongs to the same set, $\|v - q\|^2 \leq \|v - p\|^2$. Expanding the squared distances gives

$$\|u - q\|^2 - \|u - p\|^2 = \|v - q\|^2 - \|v - p\|^2 + 2\langle u - v,\, p - q\rangle \quad \leq 2\|u - v\|\,\|p - q\|,$$

by the first inequality and Cauchy–Schwarz. The left side is nonnegative because p is nearest to u. QED.

Proof of Proposition C.3. Let $\tau^\star = P_{\mathcal{A}_{\mathcal{C}}}(g_G(\hat{Y}, \sigma))$ and $\hat{\tau} = P_{\mathcal{A}_{\mathcal{C}}}(\hat{g}(\hat{Y}, \sigma))$, both measurable by the argument of Proposition C.2. By the posterior loss decomposition in the proof of Lemma 3.1, the difference in posterior expected losses between $\hat{\tau}$ and $\tau^\star$ is $n^{-1}\{\|g_G - \hat{\tau}\|^2 - \|g_G - \tau^\star\|^2\}$ with $g_G = g_G(\hat{Y}, \sigma)$, and Bayes regret is its expectation over $\hat{Y}$. Lemma C.1 at $u = g_G$ and $v = \hat{g}(\hat{Y}, \sigma)$ bounds the difference at each realization by $(2/n)\,\|\hat{g} - g_G\|\,\|\tau^\star - \hat{\tau}\|$, and $\|\tau^\star - \hat{\tau}\| \le c_J\sqrt{2s_B}$ as shown in the text. Hence, with all expectations conditional on σ,

$$\mathrm{BR}\big(\hat{\tau}^{\mathrm{EB}}_{\mathcal{C}}, \tau^\star_{G,\mathcal{C}} \mid \sigma\big) \le \frac{2c_J\sqrt{2s_B}}{n}\,\mathbb{E}\|\hat{g} - g_G\| \le 2c_J\sqrt{\frac{2s_B}{n}}\,\sqrt{\frac{1}{n}\mathbb{E}\|\hat{g} - g_G\|^2} \le 2c_J\sqrt{\frac{2s_B}{n}}\,\sqrt{2\bar{\sigma}\Delta_n + \Delta_n^2},$$

where the second inequality is Jensen's and the third is the bound on the estimated posterior mean gaps established in the proof of Proposition 4.2, valid for all n with $\log n > 1/p$. QED.

References for the Online Appendix

Barrientos, A. (2013). *Social assistance in developing countries*. Cambridge University Press.

Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis* (2nd ed.). Springer. doi: 10.1007/978-1-4757-4286-2

Brown, L. D., & Purves, R. (1973). Measurable selections of extrema. *The Annals of Statistics*, *1*(5), 902–912. doi: 10.1214/aos/1176342510

Chen, J. (2026). Empirical Bayes when estimation precision predicts parameters. *Econometrica*, *94*(2), 305–340. doi: 10.3982/ECTA22935

Gugushvili, D., & Le Nestour, A. (2019). *A detailed analysis of targeted social assistance and child poverty and simulations of the poverty-reducing effects of social transfers* (Tech. Rep.). Tbilisi, Georgia: UNICEF Georgia.

Soloff, J. A., Guntuboyina, A., & Sen, B. (2025). Multivariate, heteroscedastic empirical Bayes via nonparametric maximum likelihood. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *87*(1), 1–32. doi: 10.1093/jrsssb/qkae040

Xie, X., Kou, S. C., & Brown, L. D. (2012). SURE estimates for a heteroscedastic hierarchical model. *Journal of the American Statistical Association*, *107*(500), 1465–1479. doi: 10.1080/01621459.2012.728154