Columbia Business School

Roshni Sahoo
Assistant Professor of Business
9787991509
r.sahoo@columbia.edu

Dear Members of the Review Committee for the NBER Development Economics and Networks,

I am writing to apply to attend the NBER workshop on Development Economics and Networks. I am an Assistant Professor in the Decision, Risk, and Operations Division at Columbia Business School, and I completed my PhD in Computer Science at Stanford University in June 2026.

My research focuses on data-driven decision making for social impact. I develop statistical methods for learning from social data that may be affected by nonresponse, nonstationarity, and strategic behavior, and use these data to inform policy design in resource-constrained settings.

I would be excited to attend this workshop to connect and learn from members of the development economics community and form new collaborations. I would also be very interested to get feedback on my current research directions and learn about the types of problems that members of this community are working on. While my training was primarily in computer science, I was fortunate to be exposed to the econometrics and development economics communities during my PhD. Please find a research statement, CV, and working paper attached in this packet.

At the workshop, I would be excited to present a working paper called “What Would it Cost to End Extreme Poverty?” (joint work with co-authors Joshua Blumenstock, Paul Niehaus, Leo Selker, and Stefan Wager). Please contact Paul Niehaus (pniehaus@ucsd.edu) for a reference letter.

Thank you for your consideration,
Roshni Sahoo

Roshni Sahoo

Kravis Hall, Columbia University
New York, NY 10027

Website: roshni714.github.io
Email: r.sahoo@columbia.edu

Academic Appointments

Assistant Professor, Decision, Risk, and Operations Division 7/2026 –
Columbia Business School
Affiliate of Columbia Data Science Institute

Education

Ph.D. Computer Science, Stanford University 9/2020 – 6/2026
Advisor: Stefan Wager
Coursework: Theory of Statistics I, Theory of Statistics II, Theory of Probability I, Convex Optimization, Reinforcement Learning, Programming Languages, Social Algorithms

B.S. Computer Science and Engineering, MIT 9/2016 – 6/2020
B.S. Mathematics, MIT

Awards

Rising Star in ISyE-MS&E-IOE, Georgia Tech, 2025.

Rising Star in Data Science, University of Chicago, 2023.

Future Leader in Data Science, University of Michigan, 2023.

Phi Beta Kappa Honor Society, 2020.

IEEE Eta Kappa Nu (HKN) Honor Society, 2019.

Fellowship and Grants

Stanford Diversifying Academia, Recruiting Excellence Fellowship, 2024.

Spectrum Population Health Sciences Pilot Grant Recipient, 2023.

Stanford Data Science Scholar Fellowship, 2022.

McCoy Family Center for Ethics in Society Graduate Fellowship, 2021.

NSF Graduate Research Fellowship, 2020.

Research

Preprints

$\alpha\beta$ indicates alphabetical author order, † denotes undergraduate mentee.

1. **Sahoo, Roshni**, Joshua Blumenstock, Paul Niehaus, Leo Selker, Stefan Wager. (2025) What Would it Cost to End Extreme Poverty? (No. w34583). National Bureau of Economic Research.
2. Guan, Amy†, **Roshni Sahoo**, Joshua Salomon, Stefan Wager, Marissa Reitsma. (2025) Data Fusion for High-Resolution Estimation. *ArXiv e-prints, abs/2508.14858.*
3. ($\alpha\beta$) Lei, Lihua, **Roshni Sahoo**, Stefan Wager. (2023) Policy Learning under Biased Sample

Selection. *ArXiv e-prints, abs/2304.11735.*

Publications

1. **Sahoo, Roshni**, Lihua Lei, Stefan Wager. (2026) Learning from a Biased Sample. *ArXiv e-prints, abs/2209.01754.* Forthcoming in *Management Science.*
2. **Sahoo, Roshni**, Stefan Wager. (2025) Policy Learning with Competing Agents. *Operations Research*, 73(5), 2515–2529.
3. **Sahoo, Roshni**, Shengjia Zhao, Alyssa Chen, Stefano Ermon. (2021) Reliable Decisions with Threshold Calibration. *Advances in Neural Information Processing Systems.*
4. Zhao, Shengjia, Michael P. Kim, **Roshni Sahoo**, Tengyu Ma, Stefano Ermon. (2021) Calibrating Predictions to Decisions: A Novel Approach to Multi-Class Calibration. *Advances in Neural Information Processing Systems.*
5. Gilitschenski, Igor, **Roshni Sahoo**, Wilko Schwarting, Alexander Amini, Sertac Karaman, Daniela Rus. (2020) Deep Orientation Uncertainty Learning based on a Bingham Loss. In *International Conference on Learned Representations.*

Selected Talks

Data Fusion for High-Resolution Estimation
Statistics Seminar, Columbia University, 9/2026.
American Causal Inference Conference, 5/2026.
MIT CODE Conference, 11/2025.
Joint Statistical Meetings, 8/2025.
International Conference on Statistics and Data Science, 6/2025.

Data Leakage in Recommendation System A/B Tests
Bridging Industry and Academia in Causal Data Science Workshop, 11/2024.
Omega-Causal, Google Research, NYC, 7/2024.

What Would it Cost to End Extreme Poverty?
Sustainable Development Seminar, Columbia University, 11/2026.
AI in the Public Interest Seminar, Columbia University, 10/2026.
Statistics and Data Science & ICSD, Carnegie Mellon University, 2/2026.
Industrial and Systems Engineering, USC, 1/2026.
Operations Management & Operations Research and Statistics Groups, MIT Sloan, 1/2026.
Decision, Risk, and Operations Group, Columbia Business School, 11/2025.
Guest Lecture Cornell Tech ORIE 5355 (Decision Making Beyond Prediction), 11/2025.
INFORMS Annual Meeting, 10/2025.
Cornell Young Researchers Workshop, 10/2025.
AI in Social Science Conference, University of Chicago, 9/2025.
Statistics Seminar, Stanford University, 7/2025.
Emma Brunskill Group Meeting, 6/2025.
Workshop on AI and Analytics for Social Good, University of Maryland, 5/2025.
Guest Lecture Stanford CS 106EA, 2/2025.
Machine Learning in Economics Summer Conference, University of Chicago, 8/2024.

Bravo/SNSF Workshop on Using Data to Make Decisions, Brown University, 7/2024.
Algorithmic Fairness Seminar, Stanford University, 4/2024.
Machine Learning Lunch, Stanford University, 4/2024.
Econometrics Lunch, Stanford University, 4/2024.

Policy Learning under Biased Sample Selection

INFORMS Annual Meeting, 10/2024.
Interactive Causal Learning Conference, 11/2023.
Causal Data Science Meeting, 11/2023.
Econometrics Lunch, Stanford University, 10/2023.
Cornell Causal Reading Group, Cornell Tech, 10/2023.
Emma Brunskill Group Meeting, Stanford University, 5/2023.
Machine Learning Lunch, Stanford University, 4/2023.
Causal Inference Seminar, Stanford University, 4/2023.

Learning from a Biased Sample

American Causal Inference Conference, 5/2025.
BIRS Workshop on Bridging Prediction and Intervention Problems in Social Systems, 6/2024.
FDA Statistical Assessment Methodology and Diagnostic Biomarkers Meeting (SAMDB), 3/2024.
Rising Stars in Data Science, University of Chicago, 11/2023.
ACM EAAMO, 10/2023.
INFORMS Annual Meeting, 10/2023.
Joint Statistical Meetings, 8/2023.
Data Science for Social Good Summer Program, Stanford University, 8/2023.
MIDAS Future Leaders Summit, University of Michigan, 4/2023.
Machine Learning Lunch, Stanford University, 11/2022.
Algorithmic Fairness Seminar, Stanford University, 10/2022.
Causal Inference Seminar, Stanford University, 9/2022.

Policy Learning with Competing Agents

Causal Science Conference, Stanford University, 11/2022.
ACM EAAMO, 10/2022.
Algorithmic Fairness Seminar, Stanford University, 4/2022.
Causal Inference Seminar, Stanford University, 4/2022.
Machine Learning Lunch, Stanford University, 3/2022.

Deep Orientation Uncertainty Learning based on a Bingham Loss

Schlumberger Robotics and Intelligent Automation Webinar, 6/2020.

Reviewing

Journals (20): (# of papers in parentheses) Annals of Operations Research (1), Biometrika (3), Communications in Statistics – Simulation and Computation (1), Journal of the American Statistical Association (3), Journal of Econometrics (1), Journal of the Royal Statistical Society: Series B (5), Management Science (2), Operations Research (3), Quantitative Economics (1).

Conferences: EC 2026, FORC 2023, ACM FAccT (2023, 2024, 2025, 2026), NeurIPS 2025 (Workshop Proposals), NeurIPS 2023 (Conference), NeurIPS 2023 (Ethics), WiML (2023, 2024), ICLR (2024, 2025, 2026).

Other Academic Service

Outreach

Stanford FAST (Future Advancers of Science and Technology), Mentor, 9/2021 – 9/2023.

Stanford Graduate Women in Computer Science, Chair, 9/2022 – 9/2023.

Stanford Computing and Society, Organizer, 12/2020 – 4/2023.

Stanford CS Mentorship Program, Mentor, 9/2020 – 9/2021.

Stanford First-Generation and/or Low-Income (FLI) Mentorship Program, Mentor, 9/2020 – 9/2021.

Committees

Marketing Committee Lead, Stanford Data Science Conference, 2023.

Other

Session Chair, International Conference of Statistics and Data Science, 2025.

Discussant of "Principal-Agent Hypothesis Testing," International Seminar of Selective Inference, 2022.

Panelist for Stanford AI4ALL, 2022.

Teaching

Exploring Artificial Intelligence (CS 106EA), Stanford University, Course Assistant, 2025.

Machine Learning (CS229), Stanford University, Course Assistant, 2024.

Introduction to Deep Learning (6.S191), MIT, Teaching Assistant, 2020.

Elements of Software Construction (6.031), MIT, Lab Assistant, 2018.

Industry Experience

Google Research, Algorithms and Optimization Team, Student Researcher, 6/2024 – 9/2024.
Supervised by Jennifer Brennan, Jean Pouget-Abadie.

Two Sigma, Software Engineering Intern, 6/2019 – 8/2019.

Cruise Automation, Software Engineering Intern, 6/2018 – 8/2018.

Northrop Grumman, Machine Learning Intern, 6/2017 – 8/2017.

Outside Activities

None.

My research focuses on data-driven decision making for social impact. I develop statistical methods for learning from social data that may be affected by nonresponse, nonstationarity, and strategic behavior, and use these data to inform policy design in resource-constrained settings.

(1) How can we monitor socioeconomic outcomes in an evolving target population?

Nationally representative in-person surveys, such as those conducted by the World Bank's Living Standards Measurement Study, are the gold standard approach for collecting data from a target population to monitor socioeconomic outcomes. However, their high deployment costs preclude data collection at high temporal frequency [Kilic 2017, Barca and Hebber, 2020], so these surveys may be representative of the target population at a lagged time period. In contrast, telephone surveys permit frequent data collection, but they may suffer from sampling bias, meaning that some groups (characterized by observable or unobservable attributes) may be under- or over- represented relative to the true population. We consider the problem of estimating real-time socioeconomic indicators given representative but lagged data from the target population and real-time but potentially biased data drawn from a survey population that differs from the target population. When the selection mechanism into the survey is stationary, we demonstrate that the real-time mean outcome in the target population is a time-invariant reweighting of the survey data. The weights are identified as the solution to an augmented convex optimization problem on the data pooled from the target and survey populations across time. This is an ongoing work that builds on a line of my research on learning under biased sample selection and data fusion [Guan et al, 2025, Lei et al, 2023, Sahoo et al, 2026].

(2) How should cash-transfer programs be designed given that we can only allocate transfers on the basis of imperfect information on individuals' living standards?

The recent success of anti-poverty programs has motivated calls to end extreme poverty by funding direct cash transfers to vulnerable populations [Kharas and McArthur, 2023]. However, a barrier to implementing such a program is that policymakers only observe living standards for units who are sampled in household consumption surveys, which cover only 0.001% to 0.01% of a country's population. An open question in this area is how to efficiently and equitably allocate transfers based on imperfect but verifiable proxies of living standards (e.g. housing material), and how much it would cost to obtain meaningful reductions in the poverty rate by doing so. We study how to learn policies that allocate cash on the basis of imperfect information and provide poverty rate reduction and equity guarantees. Using nationally representative household consumption surveys from 34 countries that together account for 76% of the world's poor, we estimate that reducing the poverty rate to 1% (from a baseline of 13%) would cost $211B nominal per year. This is 4.0 times the corresponding reduction in the aggregate poverty gap, but only 19% of the cost of universal basic income. Extrapolated globally, the results imply a cost of 0.28% of global GDP to (approximately) end extreme poverty. This is ongoing work with Joshua Blumenstock (UC Berkeley), Paul Niehaus (UCSD), Leo Selker (UC Berkeley), and Stefan Wager (Stanford) [Sahoo et al, 2025].

(3) How can we efficiently allocate medical supplies prior to disease surges? What is the impact of targeted pro-active allocation of resources on disease surges?

The leading cause of post-neonatal child death in Ethiopia is preventable disease, including pneumonia, diarrhea, and malaria. Policymakers aim to reduce child mortality from preventable disease by allocating oxygen, antibiotics, antimalarials, and oral rehydration solutions to local hospitals but face uncertainty in the timing and severity of the disease surges across geographic areas. First, we aim to design a dynamic policy for resource allocation that attains the optimal tradeoff between the underage cost of having insufficient resources prior to a disease surge and the overage cost of oversupplying resources. Our approach leverages static regional-level observable information (DHS), weekly woreda-level climate and environmental information (ERA5, MERRA2), and weekly disease surveillance data (DHIS2), where we aim to minimize the cumulative amount of resources allocated while guaranteeing that sufficient resources arrive in each woreda prior to the disease surge. Second, in collaboration with researchers from Addis Ababa University and the Ethiopian Ministry of Health, we aim to evaluate the effect of this resource allocation system on disease case counts. This is an ongoing work with Kushal Patel (Stanford University), Rishi Mediratta (Stanford Medicine), and Lulu Mohe (Addis Ababa University).

Bibliography:

Barca, Valentina, and Madhumitha Hebbar. 2020. “On-Demand and Up-To-Date? Dynamic Inclusion and Data Updating for Social Assistance.” Bonn and Eschborn, Germany: GIZ.

Guan, Amy , **Roshni Sahoo**, Joshua Salomon, Stefan Wager, Marissa Reitsma. (2025) Data Fusion for High-Resolution Estimation. ArXiv e-prints, abs/2508.14858

Kilic, Talip, Umar Serajuddin, Hiroki Uematsu, and Nobuo Yoshida. Costing household surveys for monitoring progress toward ending extreme poverty and boosting shared prosperity. World Bank Policy Research Working Paper, (7951), 2017.

Lei, Lihua, **Roshni Sahoo**, Stefan Wager. (2023) Policy Learning under Biased Sample Selection. ArXiv e-prints, abs/2304.11735.

Sahoo, Roshni, Joshua Blumenstock, Paul Niehaus, Leo Selker, Stefan Wager. (2025) What Would it Cost to End Extreme Poverty? (No. w34583). National Bureau of Economic Research.

Sahoo, Roshni, Lihua Lei, Stefan Wager. (2026) Learning from a Biased Sample. ArXiv e-prints, abs/2209.01754. Forthcoming in Management Science.

What Would it Cost to End Extreme Poverty?

Roshni Sahoo* Joshua Blumenstock† Paul Niehaus‡ Leo Selker†
Stefan Wager§

September 16, 2026

Abstract

We study poverty minimization via direct transfers, framing this as a statistical learning problem while retaining the information constraints faced by real-world programs. Using nationally representative household consumption surveys from 34 countries that together account for 76% of the world's poor, we estimate that reducing the poverty rate to 1% (from a baseline of 13%) would cost $211B nominal per year. This is 4.0 times the corresponding reduction in the aggregate poverty gap, but only 19% of the cost of universal basic income. Extrapolated globally, the results imply a cost of 0.28% of global GDP to (approximately) end extreme poverty.

1 Introduction

The share of the world's population living in extreme poverty—defined as living on less than $2.15 in 2017 PPP dollars per day[1]—has declined dramatically, from an estimated 41% in 1981 to 8% in 2024.[2] This progress has historically taken place in conjunction with economic

*Corresponding author (email r.sahoo@columbia.edu), Columbia Business School, †School of Information, UC Berkeley, ‡Department of Economics, UC San Diego, §Graduate School of Business, Stanford University. We are grateful for helpful comments from seminar participants at the BREAD/NBER Program Meeting, Brown University Bravo/SNSF Workshop, Cornell University Young Researchers Workshop, Duke University, UChicago AI in Social Science Conference, UChicago Machine Learning in Economics Summer Conference, Harvard Data Science Institute, University of Maryland Workshop on AI and Analytics for Social Good, INFORMS, Stanford Econometrics, Stanford Statistics, and Yale University. We thank Apollinaire Abi, Alia Aghajanian, Vincent Armentano, Anik Ashraf, Oscar Barriga Cabanillas, Lina Cardona-Sosa, S. Chandrasekhar, Grown Chirongwe, Oeindrilla Dube, Elizabeth Foster, Tanzed Hossain, Roy Katayama, Patricia Koka, Arsène Koyangbo, Daniel Mahler, Juan Manuel Monroy Barragan, Jose Mpesayabo, Abhiroop Mukhopadhyay, Fabrice Mukendi Mutombo, Nicolas Ndayishimiye, Noé Nduwabike, Iary Rakotondradany, Khaled Saad, Nozipho Shabalala, Dhiraj Sharma, David Stifel, Jonathan Weigel, and Christina Weiser for their generous assistance accessing and interpreting data; Puneet Bhaskar, Qiqi Chen, Nolan Chu, Michael Cleary, Karl Gorski, Neela Kaushik, Henry Lopez Rodriguez, Matthewos Mesfin, Francesca Vescia, and Victor Zhenyi Wang for help with data preparation; and Aditi Acharya, Erick Rosas Lopez, and Shruthi Ramesh for excellent research assistance. This work was partially supported by NSF grant SES-2242876, a Stanford BGS Research Fund grant, and by gifts from Luke Ding and from the Manglaben Chimanbhai Patel Family Foundation. RS was supported by a Stanford Data Science Fellowship and a Stanford DARE (Diversifying Academia, Recruiting Excellence) Fellowship.

[1]Throughout the paper we refer to extreme poverty as having a material standard of living below some given threshold, a definition which may differ from local communities' concepts of poverty [Alatas et al., 2012]. We focus primarily on the traditional threshold of $2.15 in 2017 PPP dollars, but also report results for the newer threshold of $3.00 in 2021 PPP dollars [Foster et al., 2025].

[2]World Bank Poverty & Inequality Platform, https://pip.worldbank.org/poverty-calculator, accessed 7 November 2025. See also Armentano et al. [2025] for a more general discussion about the global decline in poverty.

growth [Pritchett, 2024a]. Yet it may not continue: analyzing current growth trends, Roser [2025] concludes that "progress against extreme poverty will come to a halt." In either case, the fact that hundreds of millions of people still live below very low consumption thresholds raises an immediate policy question: could extreme poverty be eliminated more rapidly through redistribution?

Recent proposals explicitly appeal to this possibility [e.g., the wealth transfers envisioned by Kharas et al., 2026]. And back-of-the-envelope calculations raise hope that the sums involved might be manageable: If policymakers could identify every household below the poverty line and transfer exactly the income needed to reach it, the total annual cost would equal the global poverty gap, estimated as of 2023 to be just 0.08% of global GDP.[3]

However, such transfers are not feasible. Poverty gap estimates rely on . Thus, for almost all households in low- and middle-income countries, there are no direct estimates of their standard of living. Collecting that information through regular surveys would be prohibitively expensive [Kilic et al., 2017] and, if used to determine transfer eligibility, would create incentives for households to misreport hard-to-verify quantities like consumption [Martinelli and Parker, 2009, Banerjee et al., 2020]. These cost and incentive considerations help explain why many existing programs instead determine eligibility using proxies for living standards that are cheaper to measure and verify [Hanna and Olken, 2018]. For example, in the common Proxy Means Test (PMT) approach, administrators use a small amount of survey data to identify a set of proxies that are jointly predictive of consumption (such as housing materials) and then collect those proxies for every household to determine eligibility in the full population [Grosh and Baker, 1995].

Learning to Target Income Transfers In this paper, we maintain this same information environment and ask how to design feasible policies that would end (up to a given tolerance) extreme poverty at the lowest possible cost. We formulate poverty targeting as a statistical learning problem, building on recent work that proposes to learn decision rules via loss minimization [Athey and Wager, 2021, Bertsimas and Kallus, 2020, Kitagawa and Tetenov, 2018]. The policymaker learns a transfer policy that maps a household's observable characteristics to a nonnegative transfer amount, minimizing expected loss subject to a budget constraint (or, equivalently, minimizing cost subject to a constraint on loss). We consider both unrestricted transfer policies, which allow each household to receive any nonnegative transfer, and binary policies, in which each household receives either zero or some common positive amount. The domain of the loss function is the household's post-transfer standard of living.[4]

Choosing the right loss function is a central concern. While the headcount poverty rate is the most easily understood and most widely used metric in public discourse about poverty,[5] it is well known that, with full information, minimizing the poverty rate directly can generate inequitable allocations [Atkinson, 1987, Bourguignon and Fields, 1990, Sen, 1976]. Intuitively, the most efficient way to reduce the poverty rate would be to allocate transfers to households just below

[3]Authors' calculations using countries' 2023 poverty gap indices, total population in 2023, 2017 market exchange rate, and PPP conversion factor for all countries in the World Bank PIP, and an estimate of 2023 global GDP of $107T from `https://data.worldbank.org/indicator/NY.GDP.MKTP.CD`.

[4]To be precise, and as we discuss below, this is the standard of living the household *could* attain if it spent the entire transfer on consumption. We abstract from dynamics, but consider households that save or invest a portion of their transfer as revealing that they are at least weakly better-off by doing so. One can thus interpret the exercise as learning policies that raise welfare to at least the level that ending today's consumption poverty would achieve.

[5]For example, Target 1.1 of the Sustainable Development Goals is "By 2030, eradicate extreme poverty for all people everywhere, currently measured as people living on less than $1.25 a day" and the corresponding indicator 1.1.1 is the "Proportion of the population living below the international poverty line by sex, age, employment status and geographic location (urban/rural)" [UN Statistics, 2019].

the poverty line, rather than to the poorest households. We show that this undesirable property generalizes to the imperfect-information case. Minimizing the poverty gap, on the other hand, is equitable in the sense that the resulting policies direct (weakly) greater support to poorer households. Gap minimization also delivers formal guarantees on poverty rate reduction: it is equivalent to minimizing the worst-case (across population subgroups) conditional poverty rate. We therefore emphasize the performance of gap-minimizing policies, while also reporting rate-minimizing results for comparison.

The Cost of Targeted Transfers We first characterize solutions to the gap- and rate-minimization problems analytically in order to (a) restrict the class of transfer policies that must be considered, and (b) identify specific functionals of the population distribution that must be learned. Gap minimization requires learning conditional quantiles of the distribution of living standards, letting us use modern approaches based on neural networks and large predictor sets, while rate minimization requires learning conditional densities, which require a bespoke semi-parametric approach and smaller predictor sets. We use these results to construct computationally feasible algorithms for implementing both approaches.

We then quantify the aggregate cost of reducing poverty in a sample of 34 countries, including essentially all of those which (i) have an extreme poverty rate above 10% or account for more than 1% of the world's extreme poor, and for which (ii) there exists a recent, publicly-available, high-quality, nationally representative living standards survey that includes a consumption aggregate and is large enough to cover at least 1,000 poor households. The first condition focuses the analysis on the set of countries likely to influence aggregate costs the most; the second increases the reliability of our living standards measures and helps ensure that we observe enough poor households for effective statistical learning.[6] Each could be relaxed in the future. Collectively, this sample accounts for 76% of the world's extreme poor.[7]

We learn policies that target transfers to households based on observable characteristics that have been used by, or are analogous to others that have been used by, existing real-world PMTs. We omit characteristics that we consider too difficult to verify in practice (e.g., income) or likely to be politically controversial (e.g., ethnicity). That said, one should think of the exercise as characterizing what is technically feasible; if policies need to satisfy additional criteria in order to be feasible politically, this might affect their costs.

We estimate that reducing the $2.15 2017 PPP poverty rate to 1% in all countries in our sample, using a gap-minimizing (and hence weakly equitable) policy, would cost $211B nominal per year. This is substantially lower than the cost of blunter, uniform policies: it is 19% of the cost of providing a Universal Basic Income at the level of the poverty line, and 57% of the cost of achieving the same poverty rate using supplemental income at amounts that are universal within countries but vary across them ("Universal Supplemental Income"). These comparisons highlight the benefit of targeting with observable proxies of consumption over not targeting at all. On the other hand, the cost of the feasible gap-minimizing policy is 4.0 times the achieved reduction in the poverty gap, or equivalently 4.0 times the cost of achieving the same gap reduction if all households' living standards were perfectly observable (the "oracle"

[6]The resulting full list of countries covered is Burundi, Benin, Burkina Faso, Bangladesh, Central African Republic, Côte d'Ivoire, Congo - Kinshasa, Colombia, Ethiopia, Ghana, Guinea-Bissau, Indonesia, India, Kenya, Liberia, Madagascar, Mexico, Mali, Malawi, Namibia, Niger, Nigeria, Pakistan, Rwanda, Sudan, Senegal, Sierra Leone, Togo, Timor-Leste, Tanzania, Uganda, Yemen, South Africa, and Zimbabwe. See Section 3 for further discussion and Appendix D for details of the data sources.

[7]The estimate of the share of the world's poor that our sample accounts for is based on national poverty rates and populations for all countries from the World Bank PIP, accessed 9 April 2026. National poverty rates in the surveys are generally consistent with poverty rate estimates from the World Bank in the corresponding survey year.

scenario). This comparison quantifies the relative cost of targeting with proxies versus perfect information. Both ratios vary substantially across countries: in countries with low poverty rates the benefits of targeting are very high, for instance, while in some of the poorest countries universal transfers cost little more than the optimal targeted ones.

Real-world transfer policies are often simpler than the optimal gap-minimizing one. If we restrict attention to binary policies, the cost of reducing poverty rates to 1% increases by 39%. If we allocate a binary transfer using a PMT rather than loss minimization, but continue to learn the optimal size of that transfer from the data, we achieve similar performance. If we use a transfer sized like those used in practice, on the other hand, we cannot reduce poverty very far, as current transfer amounts are too small to raise more than a few households out of poverty. Overall, we conclude that reducing poverty using simpler transfer structures is feasible, but only if transfers are large enough, and at meaningful added cost.

A more fundamental concern is that minimizing poverty need not maximize welfare. Restricting to equitable, gap-minimizing policies may help here to some extent. But any poverty line is still inherently arbitrary, and gap-minimization does not value gains to households even just slightly above the line. In practice, however, we find that gap-minimizing policies are cost-effective both at reducing the poverty rate, and at increasing welfare. Specifically, policies learned to directly minimize the rate are weakly *more* expensive, for a given reduction in the rate, than policies learned to minimize the gap. And gap-minimizing policies increase mean log consumption by only 1% less than policies learned to directly maximize it. Gap-minimizing policies thus allow a planner to approximately maximize welfare, without sacrificing performance on easy-to-communicate goals like headcount poverty.

We examine comparative statics with respect to several additional parameters of the problem. Defining poverty using the new $3.00 2021 PPP poverty line, rather than the $2.15 2017 PPP line, raises costs by 64%, reflecting the real shift in objectives this definition implied [Foster et al., 2025]. Restricting to purely geographic targeting, which simplifies data requirements and reduces scope for gaming, increases costs by 45%. Alternatively, replacing household survey data with remote-sensed satellite imagery in one country (Togo) for which we can geo-match the two increases costs by 18%. And reducing the amount of training data available to the algorithm by 10% increases costs by 1%—a modest elasticity, but still large enough to suggest that collecting *more* data might have high returns, since surveys cost far less than the transfer policies themselves.

We then discuss financing and its implications. For our estimates to be valid, financing for the policies we learn would need to be incremental, not diverted from existing programming that separately affects living standards. A meaningful share of that incremental financing could plausibly come, in many countries, from domestic revenue. But most countries would likely also require international transfers. International transfers at such scale could in turn have substantial macroeconomic effects; we briefly review some pertinent recent evidence.[8] We discuss what scale of wealth transfers, as proposed by Kharas et al. [2026], might be required to achieve income gains equivalent to the income transfers we study here. We also discuss the potential administrative costs of implementing policies like those we learn; while non-trivial, these would likely be orders of magnitude smaller than the costs of the transfers themselves.

Finally, we ask what the results suggest about the cost of (approximately) ending poverty globally, using the results from the 34 countries where we have data to predict the cost of ending poverty in countries where we do not. This extrapolation implies that reducing the global poverty rate to 1% would cost $304B per year: 0.28% of global GDP, or 0.47% of OECD

[8]Recent work on the general equilibrium effects of cash transfers, for example, has documented substantial expansionary effects [Egger et al., 2022, Gerard et al., 2024], which we necessarily abstract from here.

GDP.[9]

Related Work Our work sits at the intersection of two long-standing traditions of thought about economic development. The first begins with an ambitious development goal and then seeks to calculate the resources required to reach it. From the 1960s onwards, for example, analysts at development banks used a "two-gap" framework based on the Harrod-Domar constant-returns growth model to calculate the capital investment required to achieve a target rate of economic growth [Domar, 1946, Easterly, 1999, Harrod, 1939]. In the early 2000s, as attention shifted to poverty reduction and a broader set of social objectives, the UN Millennium Project estimated that its proposed Millennium Development Goals (MDGs) could be achieved if rich countries increased foreign aid to 0.54% of their GNI in 2015, from 0.25% in 2003 [Sachs, 2005].

The analyses most closely related to ours have focused specifically on the cost of achieving the MDG for extreme poverty. These studies calculated the growth in GDP per capita required to achieve a given poverty rate (assuming that the relative distribution of income stayed fixed) and then inferred the aid needed to generate that growth, using either a "two-gap" model [e.g., Devarajan et al., 2002] or coefficients from cross-country aid-on-growth regressions [e.g., Anderson and Waddington, 2007]. As the authors took care to point out, the causal assumptions in these analyses were heroic, and indeed in some cases empirically rejected.[10] This was necessarily, in the words of Devarajan et al. [2002], "a highly speculative exercise." Our approach, in contrast, is built on the more mechanical idea that giving someone $1 increases their disposable income by $1. It is in this sense less speculative (though it may for that very reason yield conservative estimates, if other strategies that reduce poverty through less direct causal mechanisms are also cheaper).

The second literature concerns the optimal design of transfer programs. A natural point of departure is the body of older theoretical work on poverty measurement that assumed full knowledge of the income or consumption distribution [e.g., Foster et al., 1984, Greer and Thorbecke, 1986, Sen, 1976]. A logical next question would then be how to extend these ideas to environments where policymakers observe only limited information about households. Along these lines, Kanbur [1987] and Ravallion and Chao [1989] discussed theoretical properties of poverty-minimizing transfer policies under specific, stylized information structures,[11] and Glewwe [1992]—in perhaps the closest antecedent to our work—proposed and illustrated learning a transfer policy as a function of a (small) set of observable covariates via the plug-in principle, i.e., by minimizing a given poverty metric in-sample. In practice, however, most subsequent work has focused on the more straightforward problem of classifying households as poor or non-poor to determine their eligibility for a fixed transfer. Common methods for doing so include proxy means testing [Alatas et al., 2012, Brown et al., 2018, Grosh and Baker, 1995, Noriega-Campero et al., 2020], including recent approaches that use non-traditional data [Aiken et al., 2022, Smythe and Blumenstock, 2022]; community-based targeting [Alatas et al., 2012, Coady et al., 2004]; and geographic targeting [Baker and Grosh, 1994, Bigman and Fofack, 2000]. Our approach differs in that we seek to learn a policy that directly minimizes a poverty

[9] Note that achieving a 1% global rate requires reducing country-specific poverty rates to 1.6%, since some countries have essentially no extremely poor people.

[10] For example, Devarajan et al. [2002] used the "two-gap" approach despite the fact that it was soundly rejected by the data [Easterly, 1999] and had been widely criticized, including by one of the authors. Anderson and Waddington [2007] used coefficients from cross-country growth regressions despite the fact that many would have agreed with Mankiw et al. [1995] even a decade earlier that "using these regressions to decide how to foster growth is ... most likely a hopeless task."

[11] Specifically, they assumed that the population consists of pre-defined, mutually-exclusive groups with known income distributions, and that no further predictors are available.

criterion subject to resource constraints, rather than predicting a binary eligibility outcome. With modern computational tools and high-dimensional statistical learning methods, revisiting Glewwe's (1992) formulation is now feasible.

The methods we develop also contribute to a broader literature on data-driven decision making [Athey and Wager, 2021, Bertsimas and Kallus, 2020, Bhattacharya and Dupas, 2012, Kallus and Zhou, 2021, Kitagawa and Tetenov, 2018, Manski, 2004]. The problems we study are "prediction-policy problems" in the sense of Kleinberg et al. [2015], and do not require estimates of causal effects, as in recent work on learning treatment assignment policies [e.g., Bhattacharya and Dupas, 2012, Huang and Xu, 2020, Luedtke and van der Laan, 2016, Sun, 2021, Sun et al., 2021, Sverdrup et al., 2023, Wang et al., 2018], though one methodological parallel is that it can be useful to reformulate budget-constrained treatment assignment problems as (fractional) knapsack-style problems [Sun et al., 2021, Sverdrup et al., 2023] as we do here. Björkegren et al. [2022] and Haushofer et al. [2025] are particularly relevant, as they study welfare-maximizing allocations of fixed-size cash transfers in the presence of treatment effect heterogeneity.[12] Future work might combine methods for learning heterogeneous effects with approaches like ours to learn dynamic transfer policies that efficiently minimize poverty over time, taking advantage of the fact that some recipients have predictably higher investment returns (i.e., are more likely to "graduate") in response to a large initial transfer.

2 Targeting Transfers via Statistical Learning

Let $\mathcal{X} := \mathbb{R}^d$ denote the space of observable characteristics and let $\mathcal{Y} := \mathbb{R}_+$ denote the space of consumption values. A country has a population distribution F over these variables; while the results in this section will involve this population distribution, they anticipate the idea that we will be able to draw representative samples from it for learning.

Cash transfer policies are common in low- and middle-income countries and, due to the cost and incentive-compatibility considerations noted above, are typically allocated based on imperfect proxies $X_i \neq Y_i$ of living standards. Given an available (per-unit) budget $B \in \mathbb{R}_+$, define a transfer policy as a mapping from the space of observable characteristics $\mathcal{X}$ and possible budgets $\mathbb{R}_+$ to a nonnegative cash-transfer amount, i.e. $t : \mathcal{X} \times \mathbb{R}_+ \to \mathbb{R}_+$. This must satisfy the budget constraint $\mathbb{E}_F[t(X; B)] \leq B$.

We define a unit's post-transfer standard of living as $Y_i + t(X_i; B)$. In effect this means analyzing the consumption the unit *could* attain if they consumed their entire transfer and if Y_i were unaffected by transfers. This formulation is standard in the literature on optimal transfer policies [Bourguignon and Fields, 1990, Glewwe, 1992, Kanbur, 1987, Ravallion and Chao, 1989] and we view it as conservative in the sense that units who save or invest a portion of their transfer are presumed at least weakly better off by doing so. In practice, reviews by Banerjee et al. [2017] and Crosta et al. [2024] find that transfers tend to increase labor supply and earnings from other sources.[13] Note also that this formulation precludes general equilibrium effects; if transfers stimulate the economy, as Egger et al. [2022] and Gerard et al. [2024] find, then our approach is conservative in this regard.

Given this setting, there are many possible ways to learn transfer policies from data. As a point of reference we first outline a widely-used approach for allocating cash transfers based

[12] For an overview of empirical evidence on the effects of cash transfers, see Bastagli et al. [2019] and Crosta et al. [2024].

[13] There is also a distinct question whether the *rule* that determines eligibility for transfers might reduce Y_i by disincentivizing effort. Evidence on this point from low income countries is sparse, though Banerjee et al. [2020] find little evidence that eligibility rules distorted consumption choices in Indonesia.

on a proxy means test (PMT) [Alatas et al., 2012, Brown et al., 2018, Grosh and Baker, 1995, Hanna and Olken, 2018].

Definition 1 (Proxy Means Test)**.** In a Proxy Means Test (PMT), observable characteristics, or "proxies," X_i are used to define the conditional mean consumption $\mu(x) := \mathbb{E}_F[Y_i \mid X_i = x]$. The conditional mean consumption μ is then thresholded to determine the allocation of transfers to units in the population, i.e.

$$t_{\text{PMT}}(x; B) := \begin{cases} \bar{t} & \mu(x) \leq \eta(\bar{t}, B), \\ 0 & \mu(x) > \eta(\bar{t}, B), \end{cases} \tag{1}$$

for some transfer size $\bar{t} \in \mathbb{R}_+$ selected by the policymaker (but not necessarily learned from data) and eligibility threshold $\eta(\bar{t}, B)$ chosen so that $\mathbb{P}_F[\mu(X) \leq \eta(\bar{t}, B)] = B/\bar{t}$.

Note that the PMT policy (1) is a binary policy: each unit can receive zero or some common amount $\bar{t}$.

The PMT has a connection to *statistical* loss minimization in the sense that $\mu(\cdot)$ is the function that minimizes $\mathbb{E}_F\left[(Y_i - \hat{\mu}(X_i))^2\right]$, the mean-squared error between Y_i and a predictor $\hat{\mu}(X_i)$. But this does not imply that the transfer policy it yields minimizes any particular post-transfer loss. More generally, there is no obvious formal link between this procedure and minimization of any *welfare* loss. Given this, it may be possible to improve on a PMT's poverty-reducing performance by explicitly learning policies via minimization of a loss function, chosen to reflect the ultimate poverty reduction goal.

Definition 2 (Unrestricted Policies via Loss Minimization)**.** For any loss function $L : \mathbb{R} \to \mathbb{R}$, an optimal unrestricted transfer policy is one which minimizes the expected post-transfer loss $\mathbb{E}_F[L(Y_i + t(X_i))]$ subject to a budget constraint B, i.e. which solves

$$\min_{t(\cdot;B):\mathcal{X}\to\mathbb{R}_+} \{\mathbb{E}_F[L(Y_i + t(X_i; B))] : \mathbb{E}_F[t(X_i; B)] \leq B\}. \tag{2}$$

Given data $(X_i, Y_i) \sim F$ i.i.d. for units $i = 1, 2, \ldots n$ from a representative survey, a policymaker can estimate the solution of (2) via empirical minimization of the loss.

The transfer policies obtained in Definition 2 are unrestricted in that they can allocate transfers of any nonnegative amount. In practice many current programs allocate transfers in only a few different sizes, and sometimes only two: zero, and some fixed positive amount. To understand the impact on costs of restricting the support of transfers, we also study the problem of learning such "binary" policies.

Definition 3 (Binary Policies via Loss Minimization)**.** For any loss function $L : \mathbb{R} \to \mathbb{R}$, an optimal binary transfer policy is one which minimizes the expected loss $\mathbb{E}_F[L(Y_i + t(X_i))]$ subject to a budget constraint B and takes on only one non-zero value, i.e. which solves

$$\min_{\substack{t(\cdot;B):\mathcal{X}\to\mathbb{R}_+, \\ \bar{t}\in\mathbb{R}_+}} \{\mathbb{E}_F[L(Y_i + t(X_i; B))] : \mathbb{E}_F[t(X_i; B)] \leq B, \quad t(x) \in \{0, \bar{t}\} \quad \forall x \in \mathcal{X}\}. \tag{3}$$

Note that this approach requires learning both the optimal transfer size $\bar{t}$ and the transfer policy t, which determines who is eligible for the transfer.

Loss minimization as in Definition 3, and Proxy Means Testing as in Definition 1, both yield binary policies; the key difference is that the loss-minimization approach directly minimizes a post-transfer poverty measure. The cash-transfer policy that solves the loss minimization

problem (3) for a particular post-transfer poverty measure yields the greatest possible reduction in that measure, while the PMT policy does not provide formal poverty reduction guarantees.

The appropriate loss function depends on one's definition of poverty. This is a classic topic in economics, with seminal contributions from Sen [1976], Foster et al. [1984], and Greer and Thorbecke [1986], among others. One important idea in this literature was that a measure of poverty should be sensitive to changes in the living standards of people below a given poverty line. Nevertheless, the poverty measure that is most easily understood and widely used in public discourse is one that lacks this feature: the poverty rate.

Example 1 (Poverty Rate)**.** Given a poverty line $c \in \mathbb{R}_+$, the pre-transfer poverty rate is the fraction of units in the population whose per-capita consumption falls below the threshold, i.e. $\mathbb{P}_F[Y_i < c]$. The post-transfer poverty rate is $\mathbb{P}_F[Y_i + t(X_i) < c]$.

We can also consider minimizing other measures such as the poverty gap index, which is sensitive to changes below the poverty line.

Example 2 (Poverty Gap Index)**.** Given a poverty line $c \in \mathbb{R}_+$, the pre-transfer poverty gap index is the average shortfall from the poverty line relative to the poverty line, i.e. $\mathbb{E}_F[(c - Y_i)_+/c]$. The post-transfer poverty gap index is, analogously, $\mathbb{E}_F[(c - Y_i - t(X_i))_+/c]$.

Both the poverty rate and the poverty gap index are special cases of the class of FGT-α indices studied by Foster et al. [1984] and Greer and Thorbecke [1986], defined as follows.

Example 3 (FGT Index)**.** Given a poverty line $c \in \mathbb{R}_+$ and any $\alpha \geq 0$, the pre-transfer FGT-α index is $\mathbb{E}_F[\mathbb{I}(Y_i < c) \cdot (c - Y_i)^\alpha]$. The corresponding post-transfer index is

$$\mathbb{E}_F[\mathbb{I}(Y_i + t(X_i) < c) \cdot (c - Y_i - t(X_i))^\alpha].$$

The FGT-0 index corresponds to the poverty rate and the FGT-1 index corresponds to the poverty gap.

2.1 Weakly Equitable Policies

Due to the emphasis on the poverty rate in public discourse, it may be tempting to learn transfer policies via minimization of the empirical poverty rate as in Example 1, as this approach directly seeks the greatest reduction in the poverty rate under a limited budget. However, in the setting where the policymaker has perfect information on consumption, i.e., $X_i = Y_i$, it is well known that rate minimization yields a policy family that prioritizes transfers to the "richest poor" [Bourguignon and Fields, 1990].[14] This is because the post-transfer poverty rate can be minimized more efficiently by raising those very close to the poverty line over it than by raising those who are further away. To rule this out we introduce the following notion of equity, which prioritizes transfers to the worst-off units.

Definition 4 (Weakly Equitable Policy)**.** A family of policies $t(x; B)$ is weakly equitable with respect to a population distribution F if incremental transfers are monotone decreasing in post-transfer consumption, i.e. let $x, x' \in \mathcal{X}$ and budgets B, B' such that $0 \leq B' \leq B$, then

$$F_{Y+t(X;B')|X=x} \preceq_{\text{SD}} F_{Y+t(X;B')|X=x'} \implies t(x;B) - t(x;B') \geq t(x';B) - t(x';B'). \quad (5)$$

[14] In particular, and under the additional condition that the budget is not large enough to lift all units above the poverty line, i.e. $0 < B < \mathbb{E}_F[(c - Y_i)_+]$, and Y_i has positive support under F, the rate-minimizing policy is

$$t(y; B) = (c - y)_+ \cdot \mathbb{I}(\eta(B) < y < c), \quad (4)$$

where $\eta(B) \in (0, c)$ and $\eta(B)$ is decreasing in B. Clearly, transfers are not allocated to the worst-off units.

By Definition 4, a weakly equitable policy family allocates weakly larger funds to the poorer units compared to richer units in terms of post-transfer consumption. This property generalizes the requirement that transfers be monotone decreasing in (pre-transfer) consumption.[15] In the full information setting, this property ensures both that larger transfers are provided to poorer units and that transfers to poorer units are not so large that the post-transfer consumption of poorer units exceeds that of richer units who receive smaller transfers.[16]

Clearly, rate minimization is not per se weakly equitable in the perfect information regime, i.e. when $X_i = Y_i$. This raises the question of whether it is weakly equitable in the more realistic regime where policy is based on imperfect proxies, i.e. there is uncertainty in Y_i given X_i. The following assumption makes concrete this notion of imperfect proxies.

Assumption 1 (Imperfect Proxies)**.** For every $x \in \mathcal{X}$, the conditional distribution $F_{Y|X=x}$ is continuously differentiable and has positive density on $\mathcal{Y} \subseteq \mathbb{R}_+$ that contains $(0, c)$.

Theorem 1 then demonstrates the negative result that for any nonconvex loss there exists an environment that satisfies Assumption 1 where the optimal policy does not satisfy weak equity:

Theorem 1. *Consider Definition 2 with some loss function L that is decreasing, integrable on $\mathbb{R}_+$, and bounded on $\mathbb{R}_+$. If L is nonconvex, then there exists a population distribution F, satisfying Assumption 1, under which the optimal policy family $t_L(x; B)$ is not weakly equitable.*

This result applies in particular to the poverty rate loss function (Example 1), which is nonconvex. Fortunately, the opposite also holds for a broad class of relevant poverty measures.

Theorem 2. *Consider loss minimization in Definition 2. If (a) L is a loss function that either is decreasing, strictly convex, differentiable, and bounded below by $C > -\infty$ and the budget $B > 0$ or (b) L is proportional to the loss function of an FGT index with $\alpha \geq 1$ (Example 3) and the budget $0 < B < c$, then for any population distribution F that satisfies Assumption 1, there exists a unique optimal policy t_L induced by solving (2) and t_L is weakly equitable with respect to F.*

Theorem 2 provides the positive result that minimization of a convex poverty measure yields weakly equitable policy families. However, one might worry that this approach yields policies that are cost-ineffective at reducing the more familiar poverty rate metric. To help navigate this tradeoff we next show that minimizing one particular convex loss function—the poverty gap index (Example 2)—yields guarantees on both weak equity and poverty rate reduction. The first point is immediate, since the poverty gap index satisfies the conditions of Theorem 2. In terms of formal guarantees on poverty rate reduction, we first show that gap minimization solves a worst-case version of rate minimization: it is equivalent to minimization of the worst-case conditional poverty rate among covariate-defined subgroups of the population.[17]

[15] To see this, consider (5) with $B > 0$ and $B' = 0$; then

$$F_{Y|X=x} \preceq_{\text{SD}} F_{Y|X=x'} \implies t(x; B) \geq t(x'; B). \tag{6}$$

[16] In the full-information setting, we show that if a policy family $t(y; B)$ is continuous in budget and weakly equitable, then post-transfer consumption is monotone increasing in pre-transfer consumption. This result is stated formally in Appendix A.5 and proven in Appendix B.9.

[17] Lemma 3 is closely related to an observation of Kanbur [1987], who studies gap minimization in a population consisting of two groups and finds that relatively more should be spent on the group with the higher poverty rate.

Lemma 3. *Suppose that Assumption 1 holds. Consider gap minimization as specified in Example 2 with budget* $0 < B < c$*. The optimal policy that solves* (2) *also solves*

$$\min_{t:\mathcal{X}\to\mathbb{R}_+} \left\{ \max_{x\in\mathcal{X}} \mathbb{P}_F \left[Y + t(X) < c \mid X = x \right] : \mathbb{E}_F \left[t(X) \right] \leq B \right\}. \tag{7}$$

Given the desirable equity and poverty rate reduction guarantees of gap minimization, our empirical analysis will focus on gap-minimizing policies, while also presenting rate-minimizing policies to examine how much more the poverty rate could be reduced if this equity condition were ignored.

3 Data Sources and Uses

3.1 Data Sources and Preparation

Appendix D.1 describes in full detail the criteria and exceptions by which we select countries for inclusion in the analysis. In brief, we begin with those with poverty rates above 10% or that account for more than 1% of the world's poor, to focus attention on the countries that matter most for global poverty. Among these we limit attention to countries with data publicly available from a recent, high-quality, nationally representative LSMS-style survey [Grosh and Glewwe, 2000], including a consumption aggregate, and that we expect (given the sample size and published poverty rate estimates) to contain at least 1,000 poor households. These conditions are meant to ensure that we can reliably learn accurate national-level policy costs. We make a few further adjustments to the resulting list: including Bangladesh because of its size, and omitting Afghanistan, Somalia, and South Sudan because we were not able to obtain reliable currency conversion rates for the relevant years. In total this yields a sample of 34 countries that vary meaningfully in their headcount poverty rates, from 1% (in Mexico) to 84% (in Congo - Kinshasa), and that collectively account for 76% of the world's extreme poor.[18]

Table E.1 lists the specific household surveys, in the indicated countries and years, for which we learn transfer policies. These are in most cases the surveys from which recent estimates of the headcount poverty rate in the World Bank's Poverty and Inequality Platform (PIP, `pip.worldbank.org`) were calculated, and are part of the World Bank's flagship Living Standards Measurement Study (LSMS) program to collect high-quality, standardized household surveys on poverty and welfare [World Bank, 2014]. In a few cases, the data underlying PIP estimates are not publicly available, and we therefore use an alternative or earlier source. All surveys are designed to be nationally representative; that said, one caveat is that they are thought to under-represent refugees, internally displaced persons, and stateless populations, particularly when these groups are concentrated in institutional settings such as camps [EGRISS, 2024].[19]

Several years have passed since most surveys were conducted, and our estimates should thus be interpreted as the cost of reducing poverty at the time of the underlying survey. More generally, because the estimates draw on information about living standards and predictors collected at the same time, one should think of them as representing the costs of ending poverty

[18] As of 9 April 2026, as computed from World Bank poverty rate estimates.

[19] As an upper bound on the magnitude of this potential omission, the cost of providing a UBI of $2.15 per day in 2017 PPP to *all* refugees and IDPs identified by the UNHCR as of 2023 is 0.04% of global GDP, increasing the total cost of the policy we derive below from 0.28% to 0.32%. This calculation excludes refugees residing in Cuba, Hong Kong SAR China, Liechtenstein, Sint Maarten (Dutch part), Cayman Islands, and Monaco for which the market exchange rate in 2017 and PPP conversion factor are unavailable from the World Bank. Authors' calculations based on `https://www.unhcr.org/refugee-statistics/download`, accessed 7 December 2025.

if poverty measurement surveys were conducted regularly. One could also use data collected in one year to learn policies administered in future years, but because poverty tends to fluctuate substantially over time [Merfeld and Morduch, 2024, Armentano et al., 2025] this would reduce their accuracy, and one would want to learn policies that explicitly account for this, as for example Aiken et al. [2025] do for proxy means tests.[20]

All surveys collected detailed information on household expenditure and self-production, and we make use of consumption aggregates constructed from this information. We compute household per capita consumption in 2017 PPP USD by converting the household consumption aggregate into 2017 PPP USD per day and dividing by household size. As is standard, this approach treats all individuals in the household as having equal consumption.[21] We compute an analysis weight for each household by multiplying its household-level sampling weight by its size.[22] The estimated headcount poverty rates this procedure yields are generally close to the World Bank estimates for the corresponding country and year, and match exactly in most cases where both estimates derive from the same survey. See Appendix D.2 for details on outcome construction.

Selecting covariates to use as predictors of living standards requires some judgment. Conceptually, we wish to select only characteristics that are plausibly verifiable, to limit scope for strategic behavior by potential recipients [Grosh and Baker, 1995, Björkegren et al., 2021] and make it easier to hold program staff accountable for enforcing eligibility criteria [Niehaus et al., 2013]. We operationalize this idea as follows. First, we identify a collection of proxy means tests actually implemented in practice in a variety of low- and middle-income countries.[23] We take the union of the predictors used in these PMTs, which yields a list spanning six broad categories: household demographics, human capital, household assets, livelihood activities, geographic indicators, and community characteristics. Where appropriate we generalize from the specific variables used in past PMTs; if a PMT included a refrigerator as a predictor, for example, we would include all large household appliances on our list of eligible predictors. At the same time, we omit a few variables which, though they have been used in the past, would in our judgment typically be difficult to verify. Examples include the age of the household's primary dwelling, measures of food (in)security, indicators for use of fertilizer, or total income. We also omit indicators for the ethnicity, tribe, religion, or caste of the household or household head; while these have been used in past PMTs, they raise questions of ethical and political suitability in the given context. We prefer to err on the conservative side and omit them.[24] Appendix D.3 provides the rubric for covariate selection. We then select only variables from this list from each living standards survey for use as predictors in our analysis.

3.2 Empirical methods

We use data from each survey to learn transfer policies via empirical minimization of the poverty rate on the survey data; empirical minimization of the poverty gap; and proxy means testing. Explicit characterizations of the empirical rate-minimizing and gap-minimizing policies and algorithms for deriving them are in Appendix A. Here we provide a brief summary of the issues that arise and the methods we use to address them.

[20] See also Beuermann et al. [2025].

[21] The methods described in Section 3.2 would apply equally to individual-level measures of living standards, though these are uncommon.

[22] The same adjustment is used by the World Bank to construct poverty estimates for the population of individuals in a country rather than the population of households. See here for details.

[23] These were Bangladesh [Kidd and Wylde, 2011], Colombia [Camacho and Conover, 2011], India [Niehaus et al., 2013, Planning Commission, 2011, 2012], Indonesia [Kidd and Wylde, 2011, Alatas et al., 2012, Fernandez and Hadiwidjaja, 2018], Malawi [Handa et al., 2022], and Peru [Hanna and Olken, 2018].

[24] We thank Lant Pritchett for suggesting this.

Unrestricted gap minimization is a convex problem, and thus tractable. Optimal transfers are (we show) functions of the conditional (on covariates) quantile functions of the distribution of living standards. We learn these conditional quantile functions for a grid of quantiles using deep learning [Goodfellow et al., 2016], minimizing pinball loss [Koenker and Bassett, 1978]. When we impose the added constraint that policies must be binary, however, the problem becomes non-convex and requires more care. We express a discrete approximation of it as a nested optimization over (i) the size of the transfer, and then (ii) the eligibility rule used to determine which households received a transfer of that size. Learning the latter requires learning the expected change in loss associated with a transfer of the specified size to each household, effectively yielding a ranking of households; we do so using empirical risk minimization via deep learning. Nesting this algorithm within a grid search over values of the transfer yields the solution. The learning procedure for binary rate minimization is analogous to the binary gap minimization procedure.

Unrestricted rate minimization is also a non-convex problem. We solve a discrete approximation of it by first showing that rate-minimizing transfers must have a specific structure: conditional on a given value x of the covariates, they must either be at the boundaries (i.e. $t(x) = 0$ or $t(x) = c$), or satisfy $f_{Y|X=x}(c - t) = \alpha$ for some common value α. Intuitively, this condition states that the number of people that would be lifted out of poverty by marginally increasing the transfer amount must be the same for any values of x at which the transfer amount is in the interior. We then show that, for any given value of α, this reduces the problem to a fractional multiple-choice knapsack problem which can be solved using estimates of the covariate distribution f_X (for which we use the empirical covariate distribution) and of the conditional distribution $f_{Y|X=x}$ (which we learn using an extension of Lindsey's method [Efron and Tibshirani, 1996]). Nesting this algorithm within a grid search over values of α yields the solution.

All statistical learning methods described in Appendix A are tuned for out-of-sample prediction using a validation set [Hastie et al., 2009]. This implies that we do not need to limit the number of predictors we use to guard against overfitting up front and lets us use predictor sets that are often larger than those used in typical PMT exercises; our models use between 40 and 420 predictors (second column of Table E.2). That said, we will also find (below) that we can use substantially smaller predictor sets without meaningful performance degradation. Finally, note that we will not be able to take full advantage of large predictor sets when learning the conditional distributions needed for unrestricted rate minimization, as the sample complexity of density estimation scales exponentially in the data dimension; we will return to this point in discussing the results below.

Implementing these learning methods requires that we specify precise training, hyperparameter selection, and evaluation procedures. To prevent iterative analysis of the same data, which could lead to "human-in-the-loop" overfitting, we prescribed these procedures in advance in a Data Use Plan (DUP). A complete copy of this plan is available online;[25] its key features are as follows.

First, we randomly partition each dataset into a 60% training sample and a 40% test sample. We do so maintaining geographic stratification at the level at which the survey is designed to be representative. We evaluate subsample sizes for this purpose using simple observation counts; where possible, we ensure that this also splits the sampling weights in the same proportion.

Second, we use the training set to learn hyperparameters for the policy learning algorithms. These include, for example, the topology of the neural networks used to fit nuisance parameters for gap minimization or binary rate minimization; the number of quantiles to fit during quantile regression for unrestricted gap minimization; the dimensionality of the predictor set used for

[25] Our data use plan is available here.

unrestricted rate minimization; and various learning parameters, among others. We learn distinct hyperparameters for each policy type we wish to learn, learning them from a 67% sub-sample of the training set and evaluating them based on the performance they induce on the complementary 33% sub-sample (the validation set). We also considered two alternative approaches: hard-coding hyperparameters based on prior intuition, or learning them from synthetic data generated from the training set via a generative adversarial network [Athey et al., 2024]. When tested in our "sandbox" environment (see below), we found that all three approaches performed similarly. We therefore chose to learn hyperparameters directly from the training data as this approach is substantially simpler than using synthetic data, while providing some protection (relative to hard-coding) against "unexpected" data distributions we might subsequently encounter.

Third, with hyperparameters in place, we use the training set to estimate all nuisance parameters necessary for learning the transfer policies. After that, with access to the covariates from the test set, we generate the transfer schedule $\{t(X_i)\}$ for all units in the test set, and verify that our budget constraint is enforced on the actual set of test set units we consider deploying the transfers on. This paradigm, which is often referred to as the transductive setting [Vapnik, 1998], mirrors typical practice in real-world transfer program design. Finally, we use the held-out outcomes Y_i from the test set to evaluate the transfer schedules. This is the relevant paradigm when the policymaker expects to know the values of the covariates (though not the outcome) for the entire target population when deciding on a policy. This is the case when, for example, a government conducts a census gathering covariate information before finalizing the transfer rule for a program.

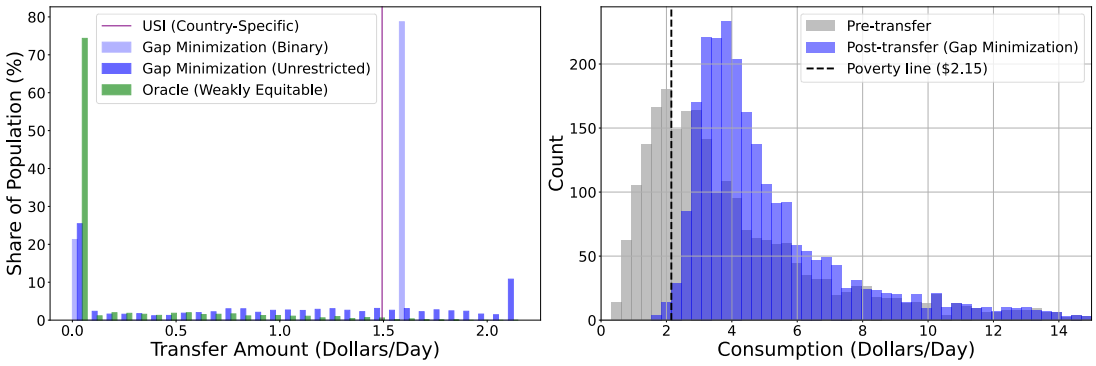
The main exception to our otherwise rigid implementation of this plan involves the 2018–2019 Integrated Household Survey in Malawi. We used this survey as a sandbox environment in which to iteratively experiment with various approaches and guide the choices to which we subsequently committed in our Data Use Plan. Results based on this survey should therefore be interpreted with this caveat in mind. That said, we will see below that the estimated benefits of targeting in Malawi (as inversely measured by the ratio of the cost of optimized transfers to the cost of universal ones; see Figure 6) are among the lowest of the countries in our sample, the opposite of what one might expect had our experimentation led to over-fitting.

Any other adjustments we made to the methodology over time are noted in a changelog within the DUP itself. To date the most salient of these have been modifying the analysis weights to reflect household size as well as survey sampling weights (which we had initially neglected to specify), and removing indicators related to ethnicity, race, caste, religion, or other such demographic groups (in response to audience feedback). For purposes of inference we also conduct an additional bootstrapping step: we hold fixed the nuisance parameters learned from the training data, draw bootstrap replicates from the test sample, and recalculate for each replicate the costs of achieving some given poverty goal. This quantifies cost uncertainty due to sampling variation in the test data.

4 Empirical Results

To build intuition we first describe learned policies and present cost results for a single country, Togo, in Section 4.1. We then apply our methods to the full sample and present the aggregate policy cost of achieving uniform poverty goals in Section 4.2.

Figure 1: Togo Transfer Distributions



This figure reports, for Togo, the distribution of transfer sizes under various learned policies all of which achieve a 1% post-transfer poverty rate (left), and the full pre- and post-transfer distribution of living standards under a gap-minimizing policy that achieves a 1% post-transfer poverty rate (right).

4.1 Togo Case Study

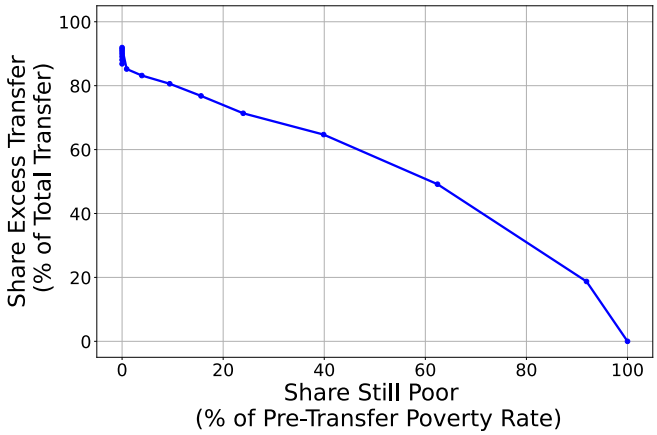
We begin with illustrative results from Togo. We choose Togo because (1) it is a representative country in the sense that our results place it near the middle of the sample in terms of the "difficulty" of targeting, as quantified below in Figure E.1, and (2) we uniquely have data for Togo letting us examine the performance of targeting based on satellite imagery, results that we will present in Section 4.3.4. Results here are based on Togo's 2018-2019 Enquête Harmonisée sur les Conditions de Vie des Ménages [WAEMU Commission, 2022c], a survey of $n = 6171$ households that contains $d = 160$ plausibly verifiable covariates.

The left panel of Figure 1 visualizes the distribution of transfer sizes under several policy alternatives, all of which achieve a post-transfer poverty rate of 1%. As expected, the (infeasible) oracle policy assigns transfer sizes of zero to many households—all those above the poverty line—and then a range of positive transfers to those below it. The binary and unrestricted gap-minimizing policies both give zero transfers to a smaller (but still meaningful) number of households; these are essentially the households that can reliably be assessed to have consumption above the poverty line based on proxies. The difference between the binary and unrestricted gap-minimization policies becomes salient for households that do receive positive transfers. The binary policy must give the same transfer amount (here the optimal choice is $1.59) to all households receiving any transfer. In contrast, the unrestricted gap-minimizing policy can tailor transfer sizes and utilize a wide range of positive transfer values to reduce costs while still meeting the poverty reduction target. Most gap-minimizing transfers are also smaller than the amount given to households by a Universal Supplemental Income.[26]

How well do learned transfers target the poor? The right panel of Figure 1 visualizes this for a gap-minimizing policy that achieves a 1% poverty rate. It plots the pre- and post-transfer distributions of per capita consumption. The share of households below the poverty line falls substantially, but the distribution of living standards above the line also shifts rightward as some funds accrue to non-poor households. For binary policies it is conventional to summarize such tradeoffs using rates of "inclusion error," i.e. of non-poor households being eligible, and of "exclusion error," i.e. of poor households being ineligible. Figure 2 presents a generalization

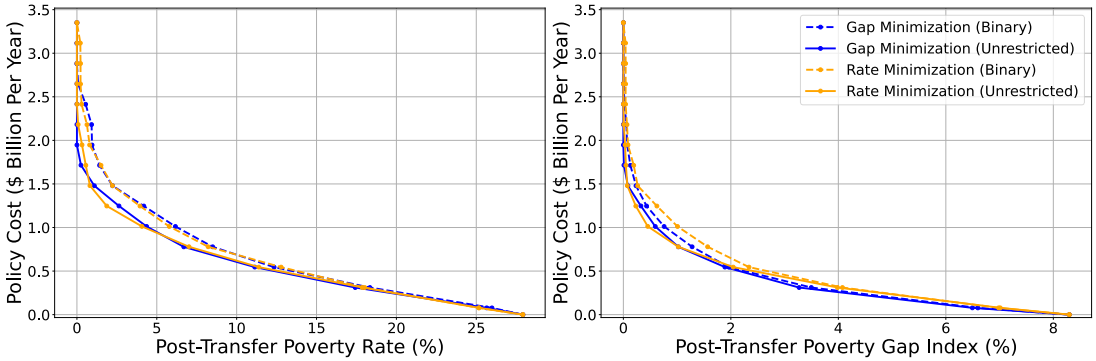
[26] A few are still quite large, however—over $2 in some cases—which is unlikely to be optimal since in the data essentially no one requires a transfer this large to exit poverty. This suggests there may be scope to further lower costs by discouraging the algorithm from learning very large transfer sizes.

Figure 2: "Inclusion error" vs. "exclusion error" for Togo



This figure reports, for Togo, the relationship between the share of those initially poor who were still poor after transfers (x-axis), which is analogous to exclusion error in conventional analyses, and the share of transfer dollars that do not contribute to closing the poverty gap (y-axis), which is analogous to inclusion error. All figures refer to gap-minimizing policies.

Figure 3: Gap- and Rate-minimizing Policy Costs, in Togo

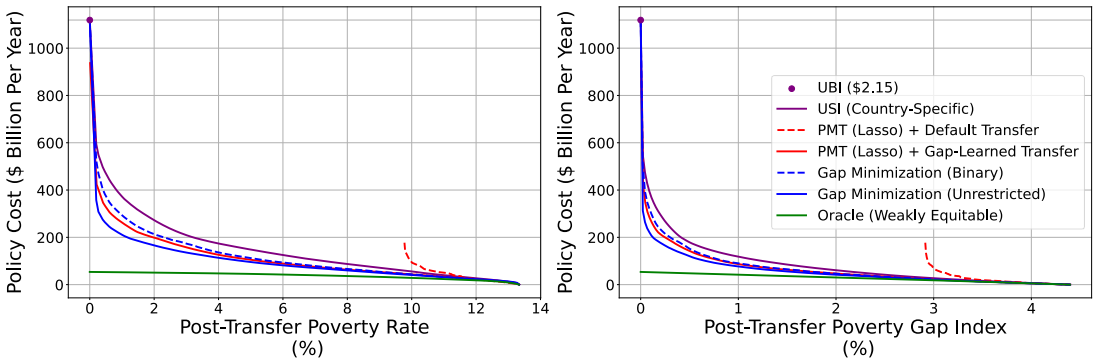


This figure reports the costs, in billions of nominal 2023 USD annually, of the indicated optimal transfer policies learned for Togo ($n = 6171$, $d = 160$). The left-hand panel plots these against the post-transfer poverty rate, while the right-hand panel plots these against the post-transfer poverty gap index.

of these concepts to the case of unrestricted transfers. It considers a range of gap-minimizing policies that vary in the share of the pre-transfer poor who are still poor after transfers, indicated on the x-axis, analogous to exclusion error. The fraction of transferred dollars that do not contribute to closing the poverty gap, on the y-axis, is analogous to the inclusion error rate. As one would expect, there is a tradeoff: the least ambitious policies, which reduce the poverty rate only slightly, are able to do so very efficiently in the sense that they route 20% or less of funds to non-poor households. More ambitious policies that lower the rate further do so at the cost of lower targeting efficiency.

Turning next to cost performance, Figure 3 compares the performance of gap-minimizing and rate-minimizing transfer policies by plotting the total annual policy cost (in nominal 2023

Figure 4: Gap minimization vs. Oracle, UBI, and PMT Benchmarks in the Full Sample



This figure reports the costs, in billions of nominal 2023 USD annually, of the indicated optimal transfer policies learned for the full sample of countries listed in Table E.2. The left-hand panel plots these against the resulting population-weighted post-transfer poverty rate, and the right-hand panel plots these against the resulting population-weighted post-transfer poverty gap index. See Section 4.2 for series definitions.

USD) of achieving a particular poverty goal (either the post-transfer poverty rate or post-transfer poverty gap index). Blue lines indicate methods that minimize the poverty gap; orange lines indicate methods that minimize the poverty rate. The solid and dashed lines correspond to unrestricted and binary policies, respectively.

Several patterns emerge from these cost series, which in general will also hold in the other countries we analyze. First, as one would expect, marginal costs increase as we approach zero poverty according to either metric. Literally ending extreme poverty requires a policy effectively amounting to UBI in order to reach every last poor household, but one can come close at far lower costs. Second, unrestricted policies are less costly than binary ones (i.e., the solid lines fall below the dashed lines). This is expected, since unrestricted policies can allocate smaller transfers to households closer to the poverty line. Finally, there is no systematic benefit to rate minimization relative to gap minimization in this setting.

4.2 Multi-Country Analysis

We now present results aggregated across the full sample of 34 countries, with the underlying country-specific results available in Appendix E. We aggregate by calculating the cost of achieving a given poverty rate in *every* country in our sample, and summing these. This necessarily costs (weakly) more than achieving the same poverty rate on *average*, potentially by reducing poverty more in countries where this is relatively inexpensive; we do it to ensure a degree of equity between the countries in our sample. We present cost curves in their entirety, but also discuss in more detail the costs of achieving a 1% poverty rate in each country using policies learned via empirical poverty-gap minimization. Note that achieving this target globally would imply that the global poverty rate is 0.7%.

Figure 4 presents the core results, plotting total annual policy cost (in nominal 2023 USD) against the resulting population-weighted headcount poverty rate (left-hand panel) or poverty gap index (right-hand panel). The primary policy of interest here is the unrestricted gap-minimizing policy, indicated in blue. The binary policy minimizes the same objective, but adds the restriction that within each country, transfers must either be zero or some common positive amount. Other series provide benchmarks. The "UBI ($2.15)" series indicates the

(constant) cost of giving every individual $2.15 2017 PPP per day, without any targeting. The "USI (Country-Specific)" series indicates the cost of giving every individual within a given country the same transfer amount less than or equal to $2.15, which we refer to as Universal Supplemental Income.

The two "PMT" series correspond to allocating transfers with a proxy means test, fitting conditional mean consumption via lasso regression [Hanna and Olken, 2018] and giving a fixed amount to individuals with mean predicted consumption below a threshold (determined by the transfer size and budget). Since the transfer size is not determined algorithmically in the PMT approach, we consider two benchmarks. In the "PMT + Default Transfer" series, we set the transfer at 20% of the mean pre-transfer per capita consumption among the poor, i.e. $0.2 \times \mathbb{E}[Y_i | Y_i \leq c]$. This roughly matches the sizing of two high-profile proxy means tested transfer programs (Progresa in Mexico, and the Benazir Income Support Programme in Pakistan), for example, and is more generally in the range we think of as typical for PMT-allocated transfers.[27] In the "PMT + Gap-Learned Transfer" series, we use the data to learn the transfer size that leads to the largest reduction in the poverty gap for a given budget when allocated via PMT. This does *not* correspond to common practice, but will help us distinguish the influence of optimal sizing of the transfer amount from that of end-to-end welfare-aware learning of the allocation rule when we compare the PMT to binary loss-minimizing policies.

Finally, the panels include an "Oracle" series in green indicating the (hypothetical) cost of achieving a given poverty gap index and poverty rate with a weakly equitable policy if full information on each household's consumption were available.[28] This is a useful benchmark given the salience of the global poverty gap in recent discourse about the cost of ending poverty [cf. Chandy et al., 2016, Kharas and McArthur, 2023, Sumner and Yusuf, 2024], but it is not a policy that could be feasibly implemented since program administrators do not directly observe the exact amount needed by every household in the population.

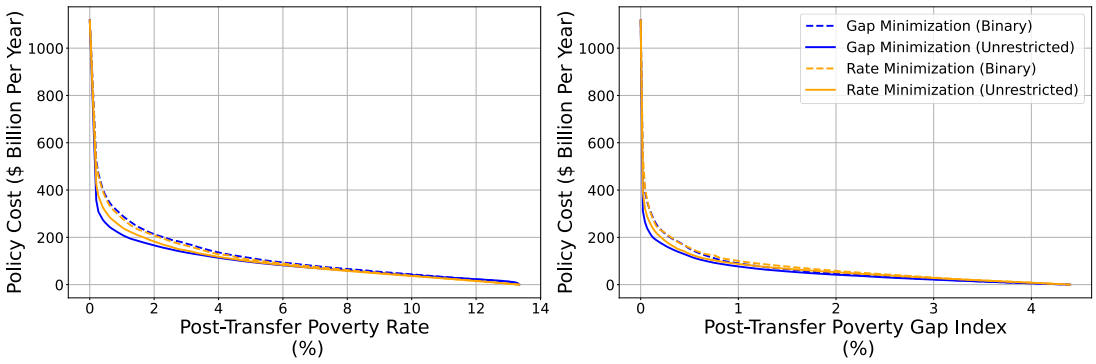
Overall the results show that the lowest-cost policies cost several times more than the aggregate poverty gap, but also substantially less than benchmark universal policies. To reduce the poverty rate in all countries to 1%, for example, would cost $211B per year, which is 4.0 times the oracle cost ($52B), but 19% of the cost of a $2.15 UBI, and 57% of the cost of a country-specific Universal Supplemental Income. This figure does not appear particularly sensitive to the fact that our test set is only a sample and not the full population: a bootstrapped 95% confidence interval for the cost is $207B to $215B. Similar ratios hold as we vary the desired post-transfer poverty rate. As the post-transfer poverty rate (or gap) approaches zero, however, the costs of all feasible policies increase at an increasing rate, as expected: requiring any of these algorithms to find the last needles in the population haystack becomes increasingly costly.

What if we restrict to binary policies? The binary gap-minimizing policy (dashed blue series) costs 39% more than the unrestricted gap-minimizing policy. This implies that, if politically or logistically expedient, the transfer policy can be reduced to a simple eligibility rule without massively increasing its cost. Allocating a binary transfer using a PMT, rather than via loss minimization, also yields respectable performance *provided* the size of that transfer is itself learned to minimize the poverty gap (solid red series). In that case, in fact, PMT targeting outperforms binary gap minimization for the most ambitious poverty rate targets. But if instead we set the transfer size to increase mean consumption among poor recipients by 20% (dashed red series) we cannot reduce either poverty metric very far with any budget. Such transfers are simply too small to raise more than a few poor households out of poverty.

[27] See Coady [2003, p. 14] and Barrientos [2013, p. 53], respectively.

[28] The oracle policy family that we visualize in both panels is the family of policies that minimizes the poverty rate while maintaining weak equity and continuity in budget. This policy is also an optimal gap-minimizing policy in the full-information setting

Figure 5: Gap- vs. Rate-minimizing Policy Costs in the Full Sample



This figure reports the costs, in billions of nominal 2023 USD annually, of the indicated optimal transfer policies learned for the full sample of countries listed in Table E.2. The left-hand panel plots these against the post-transfer poverty rate, while the right-hand panel plots these against the post-transfer poverty gap index.

While none of the policies examined in Figure 4 are optimized to reduce the poverty rate, the left-hand panel shows that gap minimization still substantially reduces the poverty rate. This suggests that it would be possible to implement gap-minimizing policies, which are weakly equitable, while still communicating with broad audiences about progress in terms of the poverty rate, a more familiar concept. That said, Figure 5 directly contrasts the empirical performance of policies optimized for the rate as opposed to the gap. Axes are as in Figure 4, but here the policies include two that minimize the rate (using either unrestricted or binary-valued transfers) as well as the corresponding policies minimizing the gap. The right-hand panel shows that, as expected, policies optimized for the gap consistently reduce the gap at a lower cost than the corresponding policies optimized for the rate. The left-hand panel shows, on the other hand, that a policy optimized to reduce the gap using unrestricted transfers actually performs somewhat *better* at reducing the rate than empirical rate minimization. In this sense there is little reason to prefer rate minimization in terms of performance.

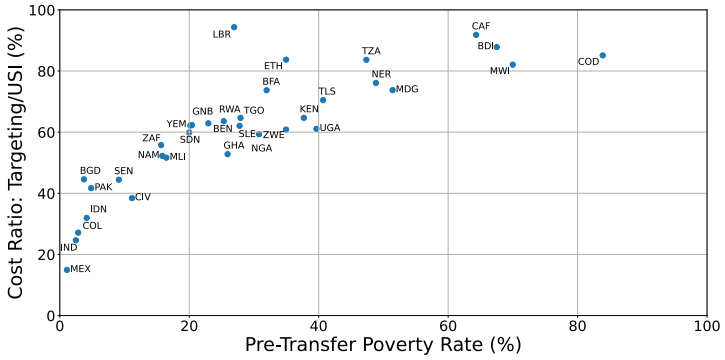
The driving factor behind this somewhat surprising result is that empirical rate minimization must solve a more challenging statistical problem than empirical gap minimization, requiring that we estimate the full conditional density functions as opposed to only their quantiles. A first, immediate consequence of this challenge is that empirical rate minimization has worse sample complexity, i.e., its errors decay more slowly with sample size. A second, more subtle consequence is that conditional density estimation requires use of a bespoke algorithm (unlike conditional quantile estimation which can be implemented using off-the-shelf neural networks)—and unsurprisingly we thus find our algorithm to be less robust at handling complex and/or high dimensional data. We seek to mitigate this issue by implementing data-driven feature selection for our rate minimization algorithm;[29] realistically, however, it is unlikely this procedure is able to make use of as many features as gap minimization.[30]

The full country-specific details underlying these results are presented in Figure E.2. Generally speaking, the comparisons above for the sample of countries as a whole hold within each

[29]No such feature selection is required for the quantile regression algorithm underlying our gap-minimization procedure, as off-the-shelf neural network implementations are already highly robust to receiving high-dimensional inputs.

[30]That said, we find that in practice gap minimization using the restricted set of features used for rate minimization performs almost identically (and, in particular, continues to reduce the rate more cost-effectively) indicating that algorithmic performance per se is the main driver (Figure E.3).

Figure 6: Cost Ratio of Gap Minimization to Universal Supplemental Income



This figure reports, for each country in Table E.2, the country's initial poverty rate (x-axis), and the ratio of the cost of a policy that achieves a 1% post-transfer poverty rate using transfers that vary person-to-person learned via gap minimization, and the cost of a policy that does so by giving a common transfer amount to all individuals (y-axis).

country as well.[31] Reducing the poverty rate to 1% with an optimal policy costs between 1.3 and 14.7 times the aggregate poverty gap (Figure E.1), but still substantially less than the cost of doing so using UBI at the level of the international poverty line. In some of the poorest countries, however, the cost of achieving this goal using a Universal Supplemental Income is not much greater than the cost of doing so using variably sized transfers (Figure 6), reflecting the fact that most households in these countries are below the poverty line.

4.3 Variations and extensions

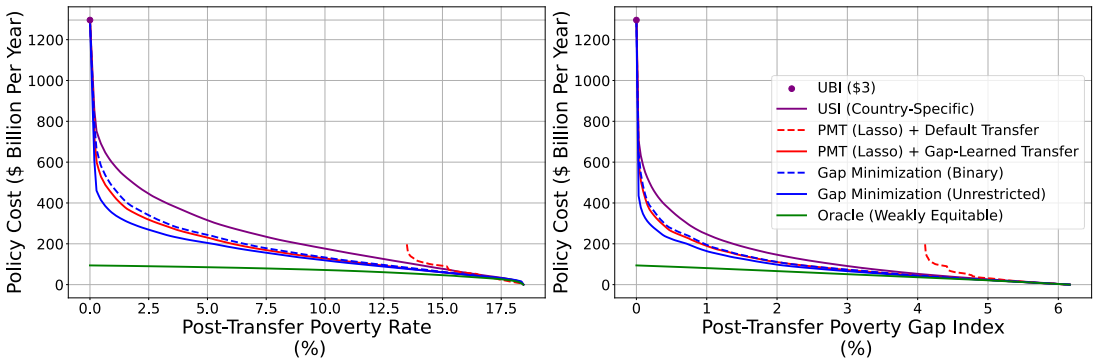
This section reports comparative statics for the minimized costs of achieving poverty goals. For simplicity of exposition we focus discussion primarily on the goal of achieving a national poverty rate of 1% in each country, as in our summary discussion above.

4.3.1 Alternative Poverty Line

The $2.15 2017 PPP line is a well-established benchmark in both public and policy discourse, derived (via periodic inflation-indexing) from the famous concept of "dollar-a-day" poverty introduced by the World Bank in 1990 and subsequently enshrined in the Millennium Development Goals and the Sustainable Development Goals. Yet it is also a very low bar. Lant Pritchett, among others, has argued forcefully that actors working to promote economic development should view "dollar-a-day" poverty as a milestone along the way, not as an end goal [Pritchett, 2024b, Pritchett and Viarengo, 2025], and the World Bank's 2025 revision to its global poverty line incorporated not only new information about prices, but also an upward shift in the *real* standard of living used to define the threshold. This was a meaningful change,

[31] For certain countries and targeting methods, we observe that the cost of learned policies can be non-monotone decreasing in the post-transfer poverty rate and gap. This behavior arises because our nuisance parameters and metrics are estimated with error in finite samples. Binary gap and rate minimization are especially prone to exhibiting non-monotonicities because the optimal solutions for these problems can rely on different nuisance parameters even for similar budgets. In contrast, unrestricted gap minimization is less fragile in this respect because similar budgets will yield solutions that rely on similar nuisance parameters with correlated errors.

Figure 7: Policy Costs for Poverty Reduction under $3.00 (2021 PPP) Poverty Line



This figure reports the costs, in billions of nominal 2023 USD annually, of the indicated optimal transfer policies learned for the full sample of countries to reduce extreme poverty, as defined as living on less than $3.00 (2021 PPP) per day. This figure is analogous to Figure 4, which reports costs to reduce extreme poverty under the $2.15 (2017 PPP) poverty line.

increasing by some 125 million the number of people defined as living in extreme poverty in 2022 [Foster et al., 2025].

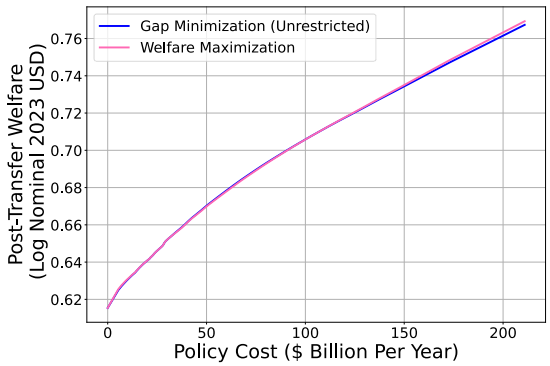
Our methods and data can be used flexibly to calculate the cost of achieving poverty goals with respect to any poverty line. Here we redo our analysis using the World Bank's new $3.00 2021 PPP line, which is of intrinsic interest as well as serving to illustrate the consequences of setting a higher bar more generally. Figure 7 replicates Figure 4, but using this higher line to define extreme poverty. Achieving a poverty rate of 1% in every country in our sample using policies learned via empirical poverty-gap minimization costs $347B per year in nominal 2023 dollars. This is a 64% increase in the cost (again, in nominal 2023 dollars) of achieving the same rate when poverty is defined using the $2.15 2017 PPP standard. This difference quantifies the extent to which ending extreme poverty under the new definition is indeed a substantially more ambitious goal than under the old.

4.3.2 Poverty Minimization vs. Welfare Maximization

Poverty minimization is a familiar and widely-accepted global goal with a clear end-point. But it implicitly places no value on gains for households *above* the extreme poverty line. This puts it in tension with broader ethical criteria for evaluating distributive justice which—while typically very sensitive to gains among the poor—also usually value gains among the non-poor. Economic welfare analysis, in particular, typically takes as its criterion the sum of individuals' utilities. To examine this tension, we next contrast policies learned to minimize the poverty gap with those learned to maximize welfare, defined as the mean of logarithmic consumption. This imposes a degree of curvature on utility functions broadly in line with available evidence [Layard et al., 2008]. We follow a process analogous to that above for poverty minimization, characterizing welfare-maximizing transfer policies analytically to the extent possible and then defining algorithms to learn solutions empirically; Appendix A.3 provides details.

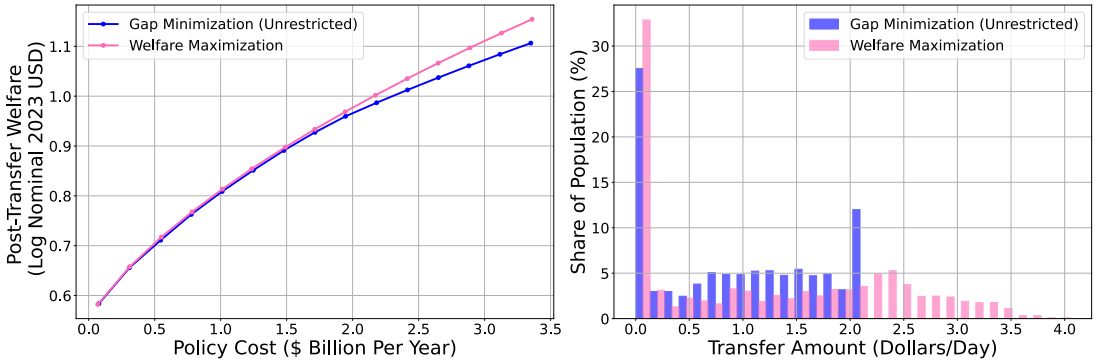
We find that welfare maximization yields higher welfare, as expected, but only slightly so. Gap-minimizing policies that achieve a 1% poverty rate in every sample country achieve a level of welfare that is only 1% lower than that which we obtain if we maximize welfare using the same country-specific budgets (see Figure 8). This is true despite the fact that the two approaches yield noticeably different transfer size distributions, as Figure 9 illustrates for Togo.

Figure 8: Aggregate welfare from poverty minimization vs. welfare maximization



This figure reports the mean realized welfare, evaluated using log utility, i.e. $\log(Y_i + t(X_i; B))$, from policies learned to minimize the poverty gap (blue series) and to maximize welfare (magenta series), for a range of budgets B.

Figure 9: Realized welfare and transfer size distributions in Togo



This figure reports, for Togo specifically, the mean realized welfare from policies as defined in Figure 8 (left-hand panel) and the corresponding transfer size distributions induced by those policies (right-hand panel).

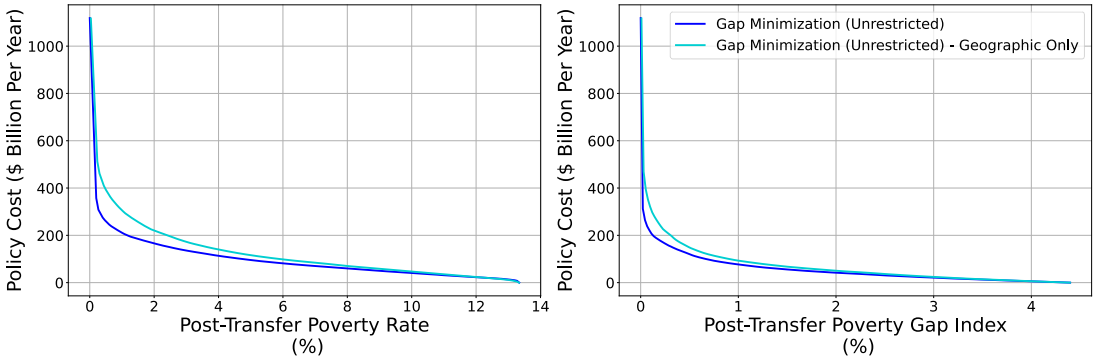
In particular, welfare maximization entails more large transfers, including some larger than $2.15, which can never be optimal for poverty minimization.

These results are not surprising, given that the households most likely to be poor are also likely to have the highest expected marginal utility. Welfare maximization would diverge strongly from poverty minimization if the available predictors allowed us to identify households unlikely to be below the poverty line, but likely to be just above it. In practice, that does not appear to be the case.

4.3.3 Geographic Predictors

One alternative to targeting based on the characteristics of individual households is to target based on regional indicators [Coady et al., 2004, Elbers et al., 2007]. One would expect such "geographic targeting" to be less accurate, as it uses fewer and less granular predictors, but

Figure 10: Geographic targeting



This figure reports the costs, in billions of nominal 2023 USD annually, of the indicated optimal transfer policies learned for the full sample of countries to reduce extreme poverty, analogous to Figure 4. It plots values separately for policies that use the full household survey data (blue series) and policies that use only regional indicators (light blue series).

simpler and cheaper to implement, and also less subject to gaming by potential recipients or to fraud by program personnel. We therefore examine next how loss-minimizing policies perform when we restrict the predictor sets to include *only* regional indicators. These are the same geographic indicators included in the larger predictor set: indicators for the finest administrative level at which the survey documentation indicates it was designed to be representative, and indicators for urban vs. rural status where available (see Appendix D.3 for details).

As expected, achieving poverty goals costs more when we limit to geographic predictors (Figure 10). But the differences are not prohibitive; achieving a 1% poverty rate using a gap-minimizing policy learned from only geographic predictors costs 45% more than when using the full household survey data. Poverty reduction using a purely geographic approach appears to be a plausible alternative.

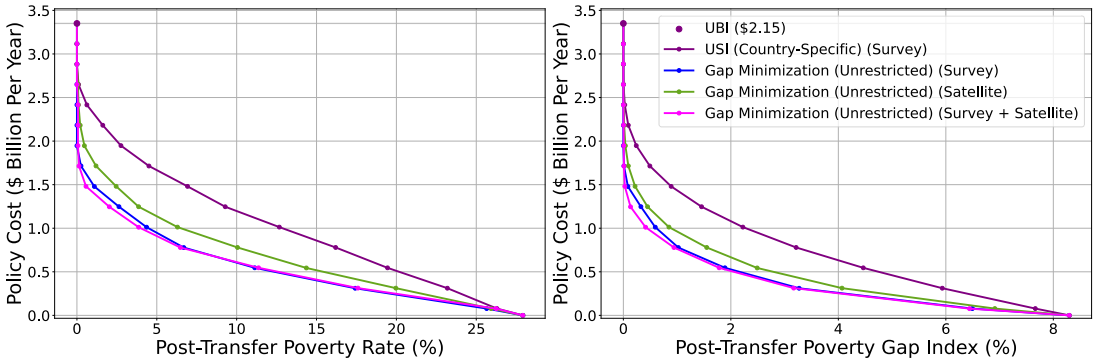
4.3.4 Satellite-Based Predictors

The global availability of remote-sensing data is creating new opportunities for targeting, as remotely-sensed predictors can be obtained at lower cost and higher frequency than survey-based ones [Smythe and Blumenstock, 2022, Yeh et al., 2020]. Here, we briefly explore the possibility of determining transfer amounts based on characteristics derived from satellite imagery, rather than collected using household surveys.

We conduct this analysis using data from Togo, where we can precisely link satellite imagery to the original (un-jittered) GPS coordinates of households from a nationally-representative consumption survey [WAEMU Commission, 2022c, Aiken et al., 2023]. For this more exploratory analysis of satellite imagery, we deviate from the protocol described in Section 3.2 by viewing policy performance for several (7) predictor sets before selecting one.[32] A full accounting of

[32]Although this deviation introduces some risk of human-in-the-loop overfitting, we believe this concern is limited by three factors. First, the number of predictor sets we considered (7) is small relative to the total number of hyperparameters tuned following the data use plan. Second, as discussed in Section 3.2, we did not see evidence of inflated performance in Malawi, which we used as a sandbox more so than we have done here in Togo. And third, selection among the satellite predictor sets we considered does not make a large difference: to reach a 1% post-transfer rate using the best-performing set costs 87% as much as using the worst-performing.

Figure 11: Policy Performance using Satellite Predictors (Togo)



This figure reports the costs, in billions of nominal 2023 USD annually, of the indicated optimal transfer policies learned for Togo. The left-hand panel plots these against the post-transfer poverty rate, while the right-hand panel plots these against the post-transfer poverty gap index.

this process can be found in our Data Use Plan.

We use the Google Satellite Embedding dataset as our predictor set [Brown et al., 2025].[33] The dataset is structured as a raster, with 64 features (embeddings) per raster point. We match each household in the survey with its nearest raster point which falls within the borders of Togo. We lack location information for 16 of the 6,171 respondent households; in such cases we impute each feature using the mean among households with location data. We then proceed as in Section 3.2.

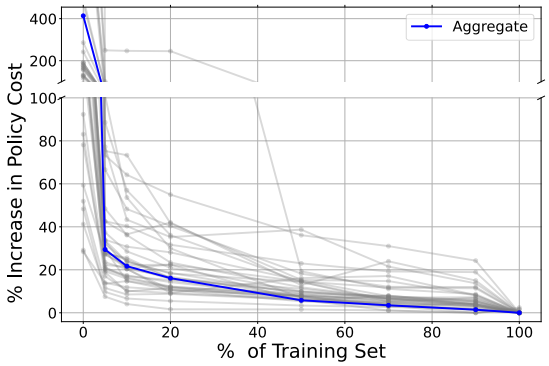
Figure 11 presents the performance of gap-minimizing policies using satellite features only, satellite features in conjunction with survey predictors, and survey predictors only. In isolation, satellite features are useful for targeting, as they reduce the cost of achieving a 1% national poverty rate by 24% relative to providing uniform USI. However, relative to targeting with survey data, the approach based on satellite data alone increases costs by 18%. Adding satellite features to survey predictors slightly improves performance, reducing costs by 6% relative to targeting with survey features alone. Overall, satellite imagery functions as an imperfect (but less expensive) substitute for survey predictors, and may be useful in conjunction with survey predictors, in this setting.

An offsetting advantage of targeting with satellite imagery alone is the avoided cost of an eligibility questionnaire. Table E.3 reports estimates of total program cost that take this into account, assuming costs of $416 per household surveyed to gather detailed consumption information for training and $7.93 per household to administer an eligibility questionnaire.[34] At these program scales, collecting survey predictors via a questionnaire pays off: it costs an additional $16M, but saves $272M in transfer costs. This reflects the more general point that, as we discuss further below, gathering eligibility data is inexpensive relative to the cost of transfers themselves.

[33] Accessed at 50-meter scale using the Google EarthEngine API on January 12, 2026.

[34] The former is the mean cost of six LSMS surveys from the Sahel region of Africa reported in Kilic et al. [2017]; the latter is the median of per-household costs in four African PMT surveys reported in Schnitzer and Stoeffler [2022].

Figure 12: Percent Increase in Policy Cost vs. Percent of Training Set



This figure reports the percentage increase in the cost (y-axis) of the optimal gap-minimizing policies as the fraction of the training set (x-axis) used for training is reduced, relative to the cost when using 100% of the training set.

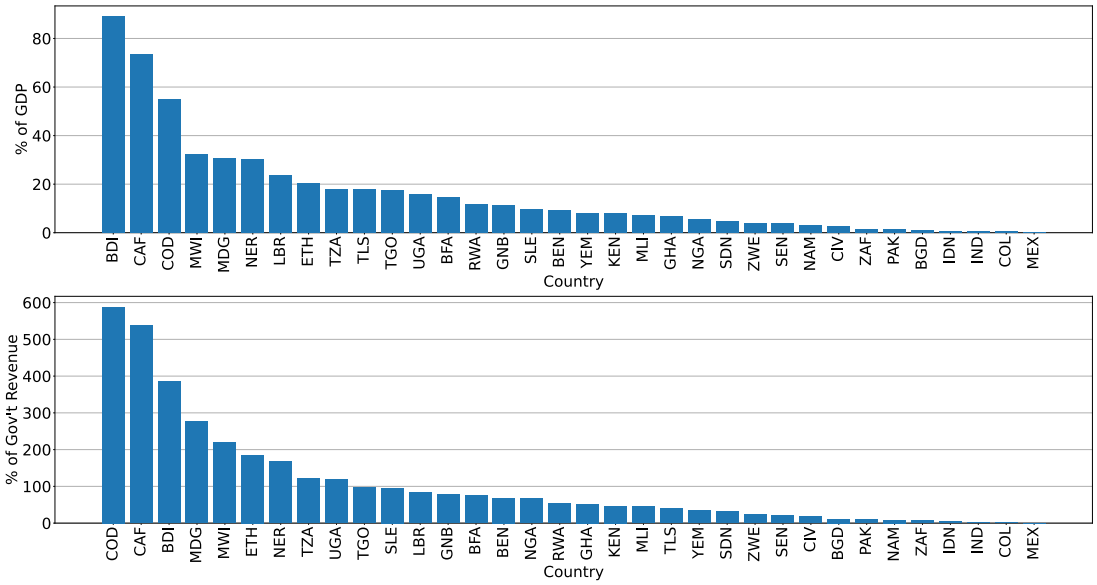
4.3.5 Training Data Size

In our final exercise we consider the potential benefits of varying the size of the survey dataset used to learn policies. The sample sizes for the underlying surveys on which we draw were set with objectives other than algorithmic policy learning in mind, raising the possibility that larger samples might yield better performance. Since the costs of additional surveys are third-order relative to the costs of the transfer policies we learn, even small gains in accuracy could yield very high returns on investment.

Of course, we cannot directly examine the accuracy benefits of surveys *larger* than those actually conducted. Instead we quantify here the performance penalties we incur if we use smaller subsets of the data actually available. Examining how these penalties vary with the size of the dataset may support some informed guesses about the benefits of larger samples. To that end, Figure 12 plots the relationship between the cost of a learned gap-minimizing policy that reduces the poverty rate to 1%, and the share of the available training data set from each country we utilize. Costs are expressed relative to the benchmark cost when we use all available data. Note that to smooth out sampling variability we repeat this exercise 15 times for 15 subsets of the training data sampled with replacement, and report averages across those draws.

The results suggest that the benefits of collecting additional data would be modest in relative terms. While costs spike when we use very small fractions of the data, they vary fairly little for values in the 80%–100% range. However, because the costs of survey data collection are typically so small relative to those of the transfers themselves, the *rate* of return could still be very high. For instance, costs decrease by 1% as we increase the share of the training set used from 90% to 100%. If they were to decrease by a further 1% were surveys 10% larger, the global savings (computed using the estimates in Section 5.1 below) would be on the order of $3B, while the added data collection cost would likely run in the $10Ms.

Figure 13: Policy Cost as a Share of GDP and of Government Revenue



This figure plots, for each indicated country, the ratio of the cost of a policy that achieves a 1% poverty rate via gap minimization to the country's GDP in the year the underlying survey was conducted (top panel) and to total government revenue in that year (bottom panel). We obtain country GDP from the World Bank (data source, accessed 13 August 2025) and government revenue percentages by country from the IMF (data source, accessed 14 July 2025).

5 Discussion & Implications

Table E.4 summarizes the estimated cost of reducing the national poverty rate to 1% in each country in our sample, along with reference information on the scale of national GDP and public revenue. Note that doing so would reduce the global poverty rate to 0.7%. We conclude in this section by discussing several implications of redistribution on the scale implied by those figures.

The first concerns sources of funds. Figure 13 plots the amount of money spent in each country, under the learned policy that reduces the national poverty rate to 1%, as a proportion of that country's GDP (top panel) and total government revenue (bottom panel). Taking a simple average over the countries in our sample, implementing such a scheme would cost 16% of GDP. To put this in context, this is equivalent to more than 90 years worth of the growth in tax revenue relative to GDP that Besley and Persson [2014] document in a sample of 18 countries during the 1900s.[35] It thus appears plausible that many countries could contribute substantially to the cost of policies like these, but implausible that many could finance them entirely on their own. This assessment is similar to that reached by Hanna and Olken [forthcoming]. They calculate the cost of transferring PPP $2.15 per day to everyone below the poverty line, taking this as a rough approximation to the true cost of eliminating poverty. They note that "in principle, one could bring everyone to the poverty line for less money than this if one could give larger transfers to those further away from the poverty line," but also that "this may understate

[35] Figure 3 in their paper shows tax revenue growing from 8% to 25% of GDP over the course of 100 years; 16% thus corresponds to 16% / 17% $\times$ 100 $=$ 94 years worth of growth.

the extent of the problem, since giving transfers only to those below the poverty line assumes that one can solve targeting challenges..." After taking both of these issues into account, the policies we learn here end up costing a similarly large share of national income.

Funding such policies using transfers from abroad, on the other hand, would raise additional questions about their macroeconomic effects. One can think of these in two parts: the effects of converting foreign currency to Local Currency Units (LCUs), and the effects of recipients spending or saving those LCUs. Both are large topics in their own right, and beyond the scope of our analysis, but we will remark briefly on related work.

With respect to currency conversion, the amounts in question are several multiples of status quo aid flows. In our sample, a country requires 16% of GDP on average to reduce the national poverty rate to 1%, while the 2023 ODA flow to each country is only 7% of its GDP on average. This calls to mind classic questions about whether exchange rate appreciation might undermine the effects of aid more generally [Rajan and Subramanian, 2011]. Any effects of aid inflows on exchange rates would be mitigated to the extent that transfer recipients subsequently purchased items with imported components, and credibly estimating the effects of net shocks to currency demand is an area of ongoing work.[36] That said, the broad point remains that effects on exchange rates could be large enough to matter.

Domestic spending might affect economic growth, through a demand channel (as in "big push" models such as that of Murphy et al. [1989]) or a credit channel [Stiglitz and Weiss, 1981], among others. Recent evidence from large-scale experiments or natural experiments broadly supports this view: Egger et al. [2022] estimate that the economies of villages in rural Kenya expanded by $2.5 for every $1 transferred into them, and Gerard et al. [2024] similarly find substantial effects of transfers on economic activity in Brazilian municipalities.[37] Any such knock-on benefits need not accrue proportionally, however—in fact, Haushofer et al. [2025] find that the transfers in Egger et al. [2022] had somewhat larger effects on the income and consumption of households that would otherwise have been somewhat less poor. As a result, it might be too optimistic to assume that one can simply divide the costs of the anti-poverty policies we learn here by an aggregate multiplier.

Another natural and related question is whether there is some upfront outlay that would be sufficient to achieve a given poverty goal on a lasting basis, as opposed to the flow-cost approach we have taken here. This of course inescapably requires assumptions about rates of return. One way to bound the up-front cost is to simply calculate the size of the endowment required to yield the annual flow expenditures we calculate, using any rate of return deemed plausible. But it may be possible to do better still by front-loading transfers to households themselves, as some of them may have access to much-higher-return investment opportunities than endowment managers do (see, for example, Haushofer et al. [2025] and Hussam et al. [2022], among many others). Here the volatility of poverty would pose a challenge. Historically, households have moved back and forth across the extreme poverty line quite frequently [Armentano et al., 2025, Baulch and Hoddinott, 2000], so that the amount one would ideally wish to transfer to

[36] One approach uses (instruments for) countries' open-market operations as a source of identifying variation; Adler et al. [2019], for example, estimate that a purchase of 1 percentage point of GDP causes a depreciation of the real exchange rate of 1.4–1.7%. Another approach uses shocks to currency demand caused by changes to the composition of global indices of emerging market bonds; these are arguably more clearly exogenous, but also harder to size, as passive funds' responses are mechanical and thus estimable but those of actively-managed funds are unknown. This approach yields larger elasticities. For example, Beltran and He [2025] estimate using this approach that in a sample of countries an inflow of 0.09% of domestic GDP led to a 1% appreciation of the exchange rate in the days immediately following, with effects dissipating only partially over the subsequent 12 months (see Table 5.1 and Figure 4 in their paper).

[37] While the transfers which Gerard et al. [2024] study were largely domestically financed, they study the effects of gross transfers (not net of taxation), which are the relevant ones for thinking about the impacts of externally financed transfers.

any given household could vary substantially from year to year [or even month to month; see Merfeld and Morduch, 2024]. A related issue is that, as noted above, policies learned from data at a given point in time may not perform as well in subsequent years if the relationship between consumption and PMT covariates changes over time [Aiken et al., 2025]. A natural extension of our framework would incorporate techniques for addressing such 'covariate shift' in the statistical learning framework of Section 2 [cf. Quinonero-Candela et al., 2008, Koh et al., 2021].

Finally, the gap-minimizing policies we estimate quantify the total cost of transfers themselves, but abstract from the administrative costs of delivering them at scale. In practice, real-world programs will incur costs to identify eligible households, enroll them, and disburse payments. Evidence from costing exercises for national social registries indicates that one-off, census-style registration campaigns generally cost on the order of US$1–3 per household registered [Lindert et al., 2020].[38] A second component of implementation cost arises from payment delivery: governments typically compensate payment providers either through small flat per-transaction fees (fractions of a US cent to a few US cents) or through ad valorem fees of roughly 1–3% of the amount transferred, as in Pakistan's BISP, India's Direct Benefit Transfer programs, and the Philippines' 4Ps [Lindert et al., 2020].[39] Relative to the total value of cash transfer policies we evaluate, these fixed and recurring costs thus appear likely to be modest. In countries without foundational identification or payment infrastructure, larger one-time investments may be required to establish those systems.

5.1 Costs at Global Scale

While the analysis above estimates costs for a sample of countries, as of the dates they were surveyed, it is natural to wonder what it implies for the cost of ending poverty at a global scale today. We consider in particular the cost of reducing the global poverty rate in 2023 to 1%. To do so it is sufficient to obtain a national poverty rate of 1.6% in all countries that currently have higher rates, as some other countries already have rates lower than 1% which bring down the (population-weighted) average.

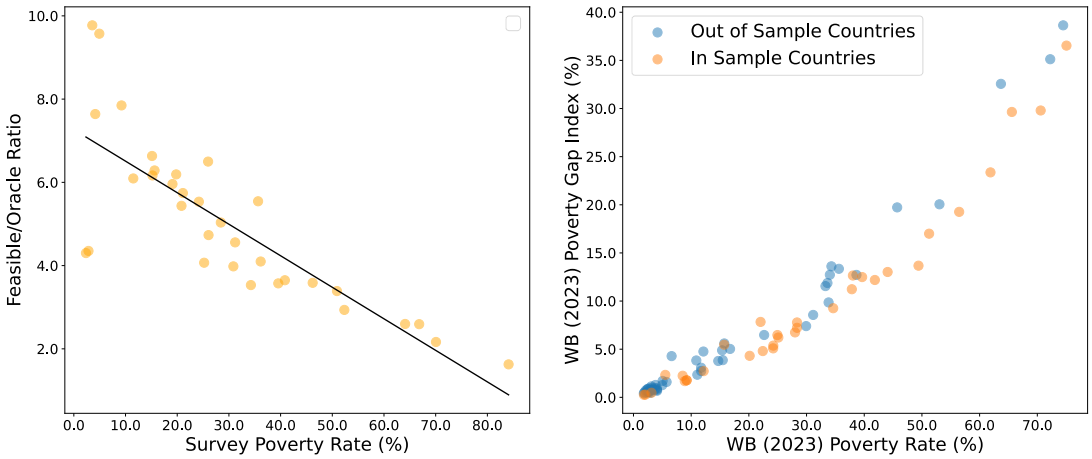
To estimate the cost of doing so, we first estimate the cost of achieving a 1.6% rate in each country in our sample using unrestricted gap-minimizing transfers. We then estimate, within that sample, the relationship between (i) the ratio of the feasible policy cost to the poverty gap—which one can think of as a measure of how much one must "overpay" due to imperfect information—and (ii) the poverty rate. Intuitively, we expect the ratio (i) to be higher in countries with lower poverty rates, since targeting there is a harder problem. The left-hand panel of Figure 14 shows that this is the case, and that we obtain an approximately linear fit, with an R^2 of 0.66.[40] We then use this estimated relationship to project the cost of achieving

[38] For example, Malawi's Unified Beneficiary Registry reports projected Phase 2 registration costs of approximately US$1.27 million for 785,000 households, or US$1.62 per household [Lindert et al., 2018]. For Turkey's Integrated Social Assistance System, system development costs of roughly US$13 million covering an estimated 10 million households imply a cost of about US$1.30 per household [Leite et al., 2017]. For Brazil's *Cadastro Único*, reported registration and re-certification costs range between US$2.03 and US$2.06 per household [Mostafa and Sátyro, 2014]. For Pakistan's National Socioeconomic Registry, per-household registration costs are estimated between US$2.30 and US$2.70 [Lindert et al., 2020].

[39] Transaction fee structures vary across programs but typically fall within a modest range. In Pakistan's BISP, the program pays a 3% transaction fee [Lindert et al., 2020, pp. 239–240]. In India's Direct Benefit Transfer system for MGNREGS payments, the Government pays a flat Re 0.50 per transaction (split across the sponsor bank, destination bank, and the National Payments Corporation of India), with an additional 1% service fee permitted for payment service providers [Lindert et al., 2020, p. 240]. In the Philippines' 4Ps program, the government negotiated a simplified service agreement that kept bank fees "under 1%" of the payment value [Lindert et al., 2020, p. 239].

[40] Note that we truncate fitted values so that (conservatively) they never fall below 1. We have also re-

Figure 14: In-sample fit and common support



The left-hand panel plots the in-sample relationship between the feasible-to-oracle policy cost ratio (given a 1.6% poverty rate target) and the poverty rate, calculated from the various surveys we use. The right-hand panel plots each country's poverty gap index and poverty rate, using estimates as of 2023 from the World Bank, differentiating between in-sample and out-of-sample countries.

a poverty rate of at most 1.6% in each country. We do so using estimates of poverty gaps and rates as of 2023 for all countries from the World Bank, which are based on the methodology in Mahler et al. [2026]. This exercise necessarily involves extrapolation across space and across time; that said, the supports of the distributions involved overlap well (see the right-hand panel of Figure 14).

The net result is an estimated cost of reducing the global poverty rate to 1% of $304B per year, equal to 0.28% of global GDP, or 0.47% of OECD GDP.[41] This is large in relation to some relevant benchmarks: OECD donors gave foreign aid equal to 0.21% of global GDP in 2023,[42] and aid levels have since fallen sharply, particularly in the United States. In terms of sheer fiscal feasibility, however, there is no question that wealthy countries could finance a policy at this scale. Nor is the amount large relative to other, arguably less critical, global expenditures. For instance, the world spends eight times as much—2.2% of global GDP—on alcoholic beverages.[43]

Over time the costs of such a scheme would fall if non-transfer incomes rose. This could be a result of the scheme itself if, as has typically been the case, recipients invested a share of their transfers. It could also result from economic growth unrelated to redistribution. Yet, as

estimated results using a quadratic fit, in which case the total estimated cost changes from 0.28% to 0.29% of global GDP. Channel Islands, Cuba, Isle Of Man, Liechtenstein, Monaco, Korea, Dem. People's Rep., and Taiwan are excluded from this calculation because the corresponding conversion factors are not available from the World Bank.

[41] Including China, the world's second-largest economy and now a major international aid donor, lowers this figure to 0.37%.

[42] In 2023, OECD foreign aid totaled $223B (https://www.oecd.org/en/events/2025/01/official-development-assistance-oda-2023-final-figures.html, accessed 6 December 2025) and global GDP totalled 107T (https://data.worldbank.org/indicator/NY.GDP.MKTP.CD, accessed 6 December 2025).

[43] Estimated 2024 global alcoholic beverages market size: $2,413B as per https://www.fortunebusinessinsights.com/alcoholic-beverages-market-107439, accessed 1 September 2025. Estimated 2024 GDP: $110,000B as per https://www.imf.org/external/datamapper/NGDPD@WEO/OEMDC/ADVEC/WEOWORLD, accessed 1 September 2025.

Table 1: Stated support for public & private contributions

	Support 0.5% of GDP policy		Would donate
	To governments	Direct to people	0.5% of income
OECD	73.5%	70.7%	59.5%
non-OECD	90.0%	90.5%	—
All	82.4%	81.3%	59.5%

Notes: $n = 14,150$ respondents from 9 OECD countries (Australia, Denmark, France, Germany, Italy, Spain, Sweden, UK, USA) & 9 non-OECD countries (Brazil, China, Ghana, India, Indonesia, Kenya, Nigeria, Pakistan, UAE). Responses collected between 15 January and 20 February 2026. Source: Banerjee et al. [2026].

we noted in opening the paper, current projections are that growth in most of the countries now home to extremely poor people will slow. To see what this might imply for the cost of ending poverty, we repeat the exercise above using estimated poverty gaps and rates from the World Poverty Clock [World Data Lab, 2022], which provides these both contemporaneously and also forecast to 2030. For 2023 these yield estimated costs of 0.27% of global GDP, or 0.45% of OECD GDP, reassuringly close to those we get using the World Bank estimates. For 2030, meanwhile, they yield estimated costs equal to 0.23% of (current) global GDP, or 0.37% of OECD GDP. This is lower (by 17%) than the 2023 WPC estimate, but also reinforces the point that on the current trajectory poverty will not vanish on its own.

Could redistribution at this scale garner widespread support? Recent political developments might seem to suggest otherwise. That said, Table 1 reports related polling numbers, reproduced with permission from Banerjee et al. [2026], collected in 18 countries during January and February of 2026. The first column indicates that a sizeable majority (73.5%) in OECD countries (and, less surprisingly, even more in non-OECD ones) state support for "a policy of rich countries contributing 0.5% of their GDP, to be given as foreign aid to governments in poor countries."[44] Voters might of course be more comfortable with transfers to foreign governments than to foreign individuals, but the next column shows that support remains essentially unchanged if the public aid were "to be given directly to people living in poor countries" rather than to their governments. The final column then considers the scope for voluntary redistribution, not coordinated and enforced by public authorities. Respondents were asked, "suppose that if everyone in rich countries voluntarily donated 0.5% of their income to people living in extreme poverty, this would be enough to eliminate extreme poverty around the world. Would you personally donate 0.5% of your income to such an effort?"— whether, in other words, they would voluntarily "do their part" if there were a convincing case that everyone doing so could end poverty. Even here, a sizeable majority (59.5%) agreed.

[44] Similarly, Fabre [2025] has recently estimated that there is majority stated support in eleven high-income countries for levels of global redistribution even higher than those contemplated here.

References

Gustavo Adler, Noëmie Lisack, and Rui C. Mano. Unveiling the effects of foreign exchange intervention: A panel approach. *Emerging Markets Review*, 40:100620, 2019. ISSN 1566-0141. doi: https://doi.org/10.1016/j.ememar.2019.100620. URL https://www.sciencedirect.com/science/article/pii/S1566014118303996.

Agence Nationale de la Statistique et de la Démographie (ANSD). Enquête harmonisée sur le conditions de vie des ménages (ehcvm) 2018-2019, 2022. Ref. SEN_2018_EHCVM_v02_M. Downloaded from https://microdata.worldbank.org/index.php/catalog/study/SEN_2018_EHCVM_v02_M on 2025-07-10.

Emily Aiken, Suzanne Bellue, Dean Karlan, Chris Udry, and Joshua E Blumenstock. Machine learning and phone data can improve targeting of humanitarian aid. *Nature*, 603(7903): 864–870, 2022.

Emily Aiken, Tim Ohlenburg, and Joshua Blumenstock. Moving targets: Can machine learning help proxy means tests better account for poverty dynamics. Technical report, UC Berkeley, January 2025.

Emily L. Aiken, Guadalupe Bedoya, Joshua E. Blumenstock, and Aidan Coville. Program targeting with machine learning and mobile phone data: Evidence from an anti-poverty intervention in Afghanistan. *Journal of Development Economics*, 161:103016, March 2023. ISSN 0304-3878. doi: 10.1016/j.jdeveco.2022.103016. URL https://www.sciencedirect.com/science/article/pii/S0304387822001584.

Vivi Alatas, Abhijit Banerjee, Rema Hanna, Benjamin A Olken, and Julia Tobias. Targeting the poor: evidence from a field experiment in indonesia. *American Economic Review*, 102 (4):1206–1240, 2012.

Nicola Amendola and Giovanni Vecchi. Durable goods and welfare measurement. *Journal of Economic Surveys*, 36(4):1179–1211, 2022.

Edward Anderson and Hugh Waddington. Aid and the millennium development goal poverty target: How much is required and how should it be allocated? Technical Report 1, 2007.

Vincent Armentano, Paul Niehaus, and Tom Vogl. How poverty fell. Technical report, UC San Diego, February 2025.

Susan Athey and Stefan Wager. Policy learning with observational data. *Econometrica*, 89(1): 133–161, 2021.

Susan Athey, Guido W. Imbens, Jonas Metzger, and Evan Munro. Using wasserstein generative adversarial networks for the design of monte carlo simulations. *Journal of Econometrics*, 240 (2):105076, 2024.

Anthony Barnes Atkinson. On the measurement of poverty. *Econometrica: Journal of the Econometric Society*, pages 749–764, 1987.

Badan Pusat Statistik (BPS). Survei sosial ekonomi nasional (susenas) 2018, 2018.

Judy L Baker and Margaret E Grosh. Poverty reduction through geographic targeting: How well does it work? *World development*, 22(7):983–995, 1994.

Abhijit Banerjee, Rema Hanna, Benjamin A. Olken, and Sudarno Sumarto. The (lack of) distortionary effects of proxy-means tests: Results from a nationwide experiment in indonesia. *Journal of Public Economics Plus*, 1:100001, 2020. ISSN 2666-5514. doi: https://doi.org/10.1016/j.pubecp.2020.100001. URL https://www.sciencedirect.com/science/article/pii/S2666551420300012.

Abhijit Banerjee, Esther Duflo, and Michael Greenstone. Support for climate compensation: Evidence from an international survey. Technical report, Working Paper, 2026. Unpublished manuscript.

Abhijit V. Banerjee, Rema Hanna, Gabriel E. Kreindler, and Benjamin A. Olken. Debunking the stereotype of the lazy welfare recipient: Evidence from cash transfer programs. *The World Bank Research Observer*, 32(2):155–184, 08 2017.

Bangladesh Bureau of Statistics (BBS). Household income and expenditure survey (hies) 2022, 2022.

Armando Barrientos. *Social Assistance in Developing Countries*. Cambridge University Press, Cambridge, 2013. ISBN 978-1-107-03902-5. doi: 10.1017/CBO9781139856058.

Francesca Bastagli, Jessica Hagen-Zanker, Luke Harman, Valentina Barca, Georgina Sturge, and Tanja Schmidt. The Impact of Cash Transfers: A Review of the Evidence from Low- and Middle-income Countries. *Journal of Social Policy*, 48(3):569–594, July 2019. ISSN 0047-2794, 1469-7823. doi: 10.1017/S0047279418000715. URL https://www.cambridge.org/core/journals/journal-of-social-policy/article/impact-of-cash-transfers-a-review-of-the-evidence-from-low-and-middleincome-countries/F8273371A30A504CBDCAFA32BF6F2EAD.

Bob Baulch and John Hoddinott. Economic mobility and poverty dynamics in developing countries. *Journal of Development Studies*, 36(6):1–24, 2000.

Paula Beltran and Chang He. Inelastic financial markets and foreign exchange interventions. Technical report, University of Virginia, March 2025.

Dimitris Bertsimas and Nathan Kallus. From predictive to prescriptive analytics. *Management Science*, 66(3):1025–1044, 2020.

Timothy Besley and Torsten Persson. Why do developing countries tax so little? *Journal of Economic Perspectives*, 28(4):99–120, November 2014.

Diether W. Beuermann, Bridget Hoffmann, Marco Stampini, David L. Vargas, and Diego Vera-Cossio. Shooting a moving target: Evaluating targeting tools for social programs when income fluctuates. *Journal of Development Economics*, 172:103395, 2025.

Debopam Bhattacharya and Pascaline Dupas. Inferring welfare maximizing treatment assignment under budget constraints. *Journal of Econometrics*, 167(1):168–196, 2012.

David Bigman and Hippolyte Fofack. Geographical targeting for poverty alleviation: An introduction to the special issue. *The World Bank Economic Review*, 14(1):129–145, 01 2000.

Daniel Björkegren, Joshua E. Blumenstock, and Samsun Knight. Manipulation-proof machine learning. *Working Paper*, 2021. URL https://www.jblumenstock.com/files/papers/manipulation.pdf. First version: November 30, 2018; this version: August 21, 2021.

Daniel Björkegren, Joshua E. Blumenstock, and Samsun Knight. (Machine) Learning What Policies Value, June 2022. URL http://arxiv.org/abs/2206.00727. arXiv:2206.00727 [cs, econ, q-fin].

Francois Bourguignon and Gary S Fields. Poverty measures and anti-poverty policy. *Recherches Économiques de Louvain/Louvain Economic Review*, 56(3-4):409–427, 1990.

C. F. Brown, M. R. Kazmierski, V. J. Pasquarella, W. J. Rucklidge, M. Samsikova, C. Zhang, E. Shelhamer, E. Lahera, O. Wiles, S. Ilyushchenko, N. Gorelick, L. L. Zhang, S. Alj, E. Schechter, S. Askay, O. Guinan, R. Moore, A. Boukouvalas, and P. Kohli. Alphaearth foundations: An embedding field model for accurate and efficient global mapping from sparse label data, 2025. URL https://arxiv.org/abs/2507.22291.

Caitlin Brown, Martin Ravallion, and Dominique Van de Walle. A poor means test? econometric targeting in africa. *Journal of Development Economics*, 134:109–124, 2018.

Adriana Camacho and Emily Conover. Manipulation of social program eligibility. *American Economic Journal: Economic Policy*, 3(2):41–65, 2011.

R.Andrés Castañeda and Damian Clarke. pip: Stata module to access World Bank's Global Poverty and Inequality data, 2026. URL https://worldbank.github.io/pip/.

Central Bureau of Statistics (CBS), Sudan. National baseline household budget survey 2009, 2009.

Central Statistical Organization (CSO), Yemen. Household budget survey 2014, 2014.

Laurence Chandy, Lorenz Noe, and Christine Zhang. The global poverty gap is falling. *Brookings Commentary*, January 2016. URL https://www.brookings.edu/articles/the-global-poverty-gap-is-falling-billionaires-could-help-close-it/.

David Coady. Alleviating structural poverty in developing countries: The approach of PROGRESA in Mexico. Background paper for the world development report 2004, World Bank, Washington, DC, 2003. URL https://documents.worldbank.org/en/publication/documents-reports/documentdetail/435991468757217796.

David Coady, Margaret Grosh, and John Hoddinott. Targeting outcomes redux. *The World Bank Research Observer*, 19(1):61–85, 2004.

Tommaso Crosta, Dean Karlan, Finley Ong, Julius Rüschenpöhler, and Christopher R. Udry. Unconditional Cash Transfers: A Bayesian Meta-Analysis of Randomized Evaluations in Low and Middle Income Countries, August 2024. URL https://www.nber.org/papers/w32779.

Carlo Del Ninno and Bradford Mills. *Safety nets in Africa: Effective mechanisms to reach the poor and most vulnerable*. The World Bank Group, 2015.

Departamento Administrativo Nacional de Estadística (DANE). Encuesta nacional de presupuestos de los hogares (enph) julio 2016 - julio 2017, 2020. Downloaded from https://microdatos.dane.gov.co/index.php/catalog/566/ on 2025-05-01.

Shantayanan Devarajan, Margaret Miller, and Eric Swanson. Goals for development: History, prospects, and costs. Technical Report 2819, World Bank, April 2002.

Evsey D. Domar. Capital expansion, rate of growth, and employment. *Econometrica*, 14(2): 137–147, 1946. URL http://www.jstor.org/stable/1905364.

William Easterly. The ghost of financing gap: testing the growth model used in the international financial institutions. *Journal of Development Economics*, 60(2):423–438, 1999. ISSN 0304-3878. doi: https://doi.org/10.1016/S0304-3878(99)00047-4. URL https://www.sciencedirect.com/science/article/pii/S0304387899000474.

Bradley Efron and Robert Tibshirani. Using specially designed exponential families for density estimation. *The Annals of Statistics*, 24(6):2431–2461, 1996.

Dennis Egger, Johannes Haushofer, Edward Miguel, Paul Niehaus, and Michael Walker. General equilibrium effects of cash transfers: Experimental evidence from kenya. *Econometrica*, 90 (6):2603–2643, 2022.

EGRISS. Capturing priority sdg indicators in refugee, internal displacement and statelessness contexts. Technical Report 3, Expert Group on Refugee, IDP and Statelessness Statistics, October 2024.

Chris Elbers, Tomoki Fujii, Peter Lanjouw, Berk Özler, and Wesley Yin. Poverty alleviation through geographic targeting: How much does disaggregation help? *Journal of Development Economics*, 83(1):198–213, May 2007. ISSN 0304-3878. doi: 10.1016/j.jdeveco.2006.02.001. URL https://www.sciencedirect.com/science/article/pii/S0304387806000150.

Ethiopian Statistical Service (ESS). Ethiopia socioeconomic panel survey, wave 5 (esps-5) 2021-2022, 2024. Public Use Dataset. Ref: ETH_2021_ESPS_W5_v02_M. Downloaded from https://microdata.worldbank.org/index.php/catalog/6161 on 2025-08-18.

Adrien Fabre. Public acceptance of international redistribution in high-income countries. Technical report, CIRED, November 2025.

Luisa Fernandez and Gracia Hadiwidjaja. Do household socioeconomic sstatus and characteristics change over a 3 year period in indonesia? evidence from susenas panel 2008-2010. Technical Report 3, Jakarta:TNP2K, 2018.

Elizabeth Mary Foster, Dean Mitchell Jolliffe, Gabriel Lara Ibarra, Christoph Lakner, and Samuel Kofi Tetteh Baah. Global poverty revisited using 2021 ppps and new data on consumption. Policy Research Working Paper Series 11137, The World Bank, Jun 2025.

James Foster, Joel Greer, and Erik Thorbecke. A class of decomposable poverty measures. *Econometrica: journal of the econometric society*, pages 761–766, 1984.

Francois Gerard, Joana Naritomi, and Joana Silva. Cash transfers and the local economy: Evidence from brazil. Technical report, London School of Economics, August 2024.

Ghana Statistical Service (GSS). Ghana living standards survey round 7 (glss 7) 2016-2017, 2018. Downloaded from https://microdata.statsghana.gov.gh/index.php/catalog/97 on 2025-09-30.

Paul Glewwe. Targeting assistance to the poor. efficient allocation of transfers when household income is not observed. *Journal of Development Economics*, 38(2):297–321, 1992.

Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

Joel Greer and Erik Thorbecke. A methodology for measuring food poverty applied to kenya. *Journal of Development Economics*, 24(1):59–74, 1986.

Margaret Grosh and Judy L Baker. Proxy means tests for targeting social programs. *Living standards measurement study working paper*, 118:1–49, 1995.

Margaret Ellen Grosh and Paul W. Glewwe. *Designing household survey questionnaires for developing countries: lessons from 15 years of the Living Standards Measurement Study*, volume 1. The World Bank, Washington, D.C., 2000. URL http://documents.worldbank.org/curated/en/452741468778781879.

Sudhanshu Handa, Maxton Tsoka, and Clare Barrington. Malawi social cash transfer programme (sctp) longitudinal impact evaluation: Baseline report. Technical report, University of North Carolina, December 2022. URL https://transfer.cpc.unc.edu/wp-content/uploads/2024/02/Malawi-SCTP-Baseline-Report-Final-Dec-2022.pdf.

Rema Hanna and Benjamin Olken. Introduction. In Rema Hanna and Benjamin Olken, editors, *The Handbook of Social Protection: Evidence to Inform Policy in Low- and Middle-Income Countries*. MIT Press, forthcoming.

Rema Hanna and Benjamin A Olken. Universal basic incomes versus targeted transfers: Anti-poverty programs in developing countries. *Journal of Economic Perspectives*, 32(4):201–226, 2018.

R. F. Harrod. An essay in dynamic theory. *The Economic Journal*, 49(193):14–33, 1939. URL http://www.jstor.org/stable/2225181.

Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York, 2nd edition, 2009. ISBN 978-0387848570. doi: 10.1007/978-0-387-84858-7.

Johannes Haushofer, Paul Niehaus, Carlos Paramo, Edward Miguel, and Michael Walker. Targeting Impact versus Deprivation. *American Economic Review*, 115(6):1936–1974, June 2025. ISSN 0002-8282. doi: 10.1257/aer.20221650. URL https://www.aeaweb.org/articles?id=10.1257/aer.20221650&from=f.

Xinyang Huang and Jin Xu. Estimating individualized treatment rules with risk constraint. *Biometrics*, 76(4):1310–1318, 2020.

Reshmaan Hussam, Natalia Rigol, and Benjamin N. Roth. Targeting high ability entrepreneurs using community information: Mechanism design in the field. *American Economic Review*, 112(3):861–98, March 2022.

IMF. Imf data portal. https://data.imf.org, 2025.

Institut Centrafricain des Statistiques et des Etudes Economiques et Sociales (ICASEES). Enquête harmonisée sur le conditions de vie des ménages (ehcvm) 2021, 2021.

Institut de Statistiques et d'Etudes Economiques du Burundi (ISTEEBU). Enquête sur les conditions de vie des ménages du burundi (eicvmb) 2020-2021, 2020.

Institut National de la Statistique et de la Démographie (INSD). Enquête harmonisée sur le conditions de vie des ménages (ehcvm) 2018-2019, 2022. Ref. BFA_EHCVM_2018_v03_M. Downloaded from https://microdata.worldbank.org/index.php/catalog/study/BFA_2018_EHCVM_v03_M on 2025-05-07.

Institut National de la Statistique et de l'Analyse Économique (INSAE). Enquête harmonisée sur le conditions de vie des ménages (ehcvm) 2018-2019, 2022. Ref. BEN_2018_EHCVM_v02_M. Downloaded from https://microdata.worldbank.org/index.php/catalog/study/BEN_2018_EHCVM_v02_M on 2025-06-03.

Institut National de la Statistique (INS). Enquête harmonisée sur le conditions de vie des ménages (ehcvm) 2018-2019, 2022. Ref. CIV_2018_EHCVM_v02_M. Downloaded from https://microdata.worldbank.org/index.php/catalog/study/CIV_2018_EHCVM_v02_M on 2025-05-08.

Institut National de la Statistique (INS), DRC. Enquête par grappes à indicateurs des odd (egi-odd) 2020, 2020.

Institut National de la Statistique (INSTAT). Enquête harmonisée sur le conditions de vie des ménages (ehcvm) 2018-2019, 2022. Ref. MLI_2018_EHCVM_v02_M. Downloaded from https://microdata.worldbank.org/index.php/catalog/study/MLI_2018_EHCVM_v02_M on 2025-05-08.

Institut National de la Statistique (INSTAT), Madagascar. Enquête permanente auprès des ménages (epm) 2021-2022, 2021.

Instituto Nacional de Estadística y Geografía (INEGI). Encuesta nacional de ingresos y gastos de los hogares (enigh) 2024, 2025. Downloaded from https://en.www.inegi.org.mx/programas/enigh/nc/2024/ on 2025-10-27.

Nathan Kallus and Angela Zhou. Minimax-optimal policy learning under unobserved confounding. *Management Science*, 67(5):2870–2890, 2021.

SM Ravi Kanbur. Measurement and alleviation of poverty: With an application to the effects of macroeconomic adjustment. *Staff Papers-International Monetary Fund*, pages 60–85, 1987.

Kenya National Bureau of Statistics (KNBS). Kenya continuous household survey programme (kchsp) 2021, 2023. Downloaded from https://statistics.knbs.or.ke/nada/index.php/catalog/123/ on 2025-04-14.

Homi Kharas and John McArthur. Updating institutional technologies: a purpose-driven fund to end extreme poverty. Technical Report 183, Brookings, 2023.

Homi Kharas, Michael Faye, Bertha Bangara-Chikadza, Joshua E. Blumenstock, Sembene Daouda, Dean Karlan, Bright Simons, and Zouera Youssoufou. End of extreme poverty: A case for action and a plan to start. *Brookings*, 2026. URL https://www.brookings.edu/articles/end-of-extreme-poverty-a-case-for-action-and-a-plan-to-start/.

Stephen Kidd and Emily Wylde. Targeting the poorest: An assessment of the proxy means test methodology. *AusAID Research Paper, Australian Agency for International Development, Canberra, Australia*, 2011.

Talip Kilic, Umar Serajuddin, Hiroki Uematsu, and Nobuo Yoshida. Costing household surveys for monitoring progress toward ending extreme poverty and boosting shared prosperity. *World Bank Policy Research Working Paper*, (7951), 2017.

Toru Kitagawa and Aleksey Tetenov. Who should be treated? empirical welfare maximization methods for treatment choice. *Econometrica*, 86(2):591–616, 2018.

Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan, and Ziad Obermeyer. Prediction policy problems. *American Economic Review*, 105(5):491–495, 2015.

Achim Klenke. *Probability theory: a comprehensive course*. Springer Science & Business Media, 2013.

Roger Koenker and Gilbert Bassett, Jr. Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50, 1978.

Pang Wei Koh, Shiori Sagawa, Henrik Marklund, Sang Michael Xie, Marvin Zhang, Akshay Balsubramani, Weihua Hu, Michihiro Yasunaga, Richard Lanas Phillips, Irena Gao, et al. Wilds: A benchmark of in-the-wild distribution shifts. In *International Conference on Machine Learning*, pages 5637–5664. PMLR, 2021.

R. Layard, G. Mayraz, and S. Nickell. The marginal utility of income. *Journal of Public Economics*, 92(8):1846–1857, August 2008. ISSN 0047-2727. doi: 10.1016/j.jpubeco.2008.01.007. URL https://www.sciencedirect.com/science/article/pii/S0047272708000248.

Phillippe Leite, Tina George, Changqing Sun, Theresa Jones, and Kathy Lindert. Social registries for social assistance and beyond: A guidance note & assessment tool. Social Protection and Jobs Discussion Paper 1704, World Bank, Washington, DC, 2017.

Liberia Institute for Statistics and Geo-Information Services (LISGIS). Liberia household income and expenditure survey (hies) 2014-2015, 2024. Downloaded from https://microdata.lisgislr.org/index.php/catalog/19 2025-09-19.

Kathy Lindert, Colin Andrews, Chipo Msowoya, Boban Varghese Paul, Elijah Chirwa, and Anita Mittal. Rapid social registry assessment: Malawi's unified beneficiary registry (ubr). Social Protection and Jobs Discussion Paper 1803, World Bank, Washington, DC, 2018.

Kathy Lindert, Tina George Karippacheril, Inés Rodríguez Caillava, and Kenichi Nishikawa Chávez, editors. *Sourcebook on the Foundations of Social Protection Delivery Systems*. World Bank, Washington, DC, 2020.

Alexander R Luedtke and Mark J van der Laan. Optimal individualized treatments in resource-limited settings. *The international journal of biostatistics*, 12(1):283–303, 2016.

Daniel Gerszon Mahler, Marta Schoch, Christoph Lakner, and Minh Cong Nguyen. A parsimonious approach to predicting income distributions. *Journal of Development Economics*, 180: 103695, 2026. ISSN 0304-3878. doi: https://doi.org/10.1016/j.jdeveco.2025.103695. URL https://www.sciencedirect.com/science/article/pii/S0304387825002469.

N. Gregory Mankiw, Edmund S. Phelps, and Paul M. Romer. The growth of nations. *Brookings Papers on Economic Activity*, 1995(1):275–326, 1995.

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

César Martinelli and Susan Wendy Parker. Deception and Misreporting in a Social Program. *Journal of the European Economic Association*, 7(4):886–908, 2009. ISSN 1542-4766. URL https://www.jstor.org/stable/40282792. Publisher: Oxford University Press.

Joshua Merfeld and Jonathan Morduch. Poverty at higher frequency. Technical report, NYU, August 2024.

Ministry of Statistics and Programme Implementation (MoSPI). Household consumption expenditure survey 2022-23, 2024. Downloaded from http://microdata.gov.in/NADA/index.php/catalog/224/ on 2025-01-12.

Joana Mostafa and Natália G. D. Sátyro. Cadastro único: A registry supported by a national public bank. Working Paper 126, International Policy Centre for Inclusive Growth, Brasilia, Brazil, 2014.

Kevin M. Murphy, Andrei Shleifer, and Robert W. Vishny. Industrialization and the big push. *Journal of Political Economy*, 97(5):1003–1026, 1989.

Namibia Statistics Agency (NSA). Namibia household income and expenditure survey (nhies) 2015/16, 2018. Downloaded from https://microdata.nsanamibia.com/index.php/catalog/12 on 2025-11-13.

National Bureau of Statistics (NBS). Living standards survey 2018-2019, 2021. Ref. NGA_2018_LSS_v01_M. Downloaded from https://microdata.worldbank.org/index.php/catalog/3827 on 2024-11-17.

National Bureau of Statistics (NBS), Ministry of Finance and Planning. Tanzania national panel survey 2020-21, wave 5, 2023. Ref: TZA_2020_NPS-R5_v02_M. Downloaded from https://microdata.worldbank.org/index.php/catalog/5639 on 2025-04-29.

National Institute of Statistics of Rwanda (NISR). Integrated household living conditions survey (eicv7) 2023-2024, 2025. Downloaded from https://microdata.statistics.gov.rw/index.php/catalog/119 on 2025-10-24.

National Statistical Office (NSO). Fifth integrated household survey 2019-2020, 2020. Ref. MWI_2019_IHS-V_v06_M. Downloaded from https://microdata.worldbank.org/index.php/catalog/3818 on 2024-02-26.

National Statistics Directorate (NSD), Timor-Leste. Timor-leste survey of living standards 2007 and extension 2008, 2011. Downloaded from https://microdata.worldbank.org/index.php/catalog/78 on 2025-07-10.

Paul Niehaus, Antonia Attanassova, Mariane Bertrand, and Sendhil Mullainathan. Targeting with agents. *American Economic Journal: Economic Policy*, 5(1):206–238, February 2013.

Alejandro Noriega-Campero, Bernardo Garcia-Bulle, Luis Fernando Cantu, Michiel A Bakker, Luis Tejerina, and Alex Pentland. Algorithmic targeting of social policies: fairness, accuracy, and distributed governance. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pages 241–251, 2020.

Pakistan Bureau of Statistics (PBS). Household income and expenditure survey (hies) 2018-19, 2020. Downloaded from https://www.pbs.gov.pk/wp-content/uploads/2020/07/data_in_stata.zip in October 2025.

Planning Commission. Bpl criteria. Technical report, Parliament of India, August 2011. URL https://rsdebate.nic.in/bitstream/123456789/594530/1/IQ_223_18082011_U1877_p78_p89.pdf.

Planning Commission. Report of the expert group to recommend the detailed methodology for identification of families living below poverty line in the urban areas. Technical report, Government of India, December 2012. URL https:

//des.mizoram.gov.in/uploads/attachments/981d16edfe868b9c691d76c48e8f1251/pages-53-bpl-criteria-in-urban.pdf.

Lant Pritchett. The case for economic growth as the path to better human wellbeing. Technical report, Harvard University, October 2024a.

Lant Pritchett. 'dollar a day' poverty was a development mileage marker, not the destination. Working paper, London School of Economics, February 2024b.

Lant Pritchett and Martina Viarengo. Raising the bar: A poverty line for global inclusion. Growth Lab Working Paper 250, Harvard University, 2025. URL https://growthlab.hks.harvard.edu/sites/projects.iq.harvard.edu/files/2025-05-glwp-250-poverty-line-global-inclusion.pdf.

Joaquin Quinonero-Candela, Masashi Sugiyama, Anton Schwaighofer, and Neil D Lawrence. *Dataset shift in machine learning*. Mit Press, 2008.

Raghuram G. Rajan and Arvind Subramanian. Aid, dutch disease, and manufacturing growth. *Journal of Development Economics*, 94(1):106–118, 2011. ISSN 0304-3878. doi: https://doi.org/10.1016/j.jdeveco.2009.12.004. URL https://www.sciencedirect.com/science/article/pii/S0304387809001278.

Martin Ravallion and Kalvin Chao. Targeted policies for poverty alleviation under imperfect information: algorithms and applications. *Journal of Policy Modeling*, 11(2):213–224, 1989.

Max Roser. The end of progress against extreme poverty? *Our World in Data*, 2025. https://archive.ourworldindata.org/20251125-173858/end-progress-extreme-poverty.html.

Jeffrey Sachs. *Investing in development : a practical plan to achieve the Millennium Development Goals*. Earthscan, London, UK, 2005.

Pascale Schnitzer and Quentin Stoeffler. Targeting for social safety nets: Evidence from nine programs in the sahel. *Available at SSRN 4017172*, 2022.

Amartya Sen. Poverty: an ordinal approach to measurement. *Econometrica: Journal of the Econometric Society*, pages 219–231, 1976.

Isabella S. Smythe and Joshua E. Blumenstock. Geographic microtargeting of social assistance with high-resolution poverty maps. *Proceedings of the National Academy of Sciences*, 119 (32):e2120025119, August 2022. doi: 10.1073/pnas.2120025119. URL https://www.pnas.org/doi/10.1073/pnas.2120025119. Publisher: Proceedings of the National Academy of Sciences.

Statistics Sierra Leone (Stats SL). Sierra leone integrated household survey (slihs) 2018, 2018. Downloaded from https://www.statistics.sl/index.php/sierra-leone-integrated-household-survey-slihs.html on 2025-11-13.

Statistics South Africa. Income and expenditure survey 2010-2011, 2013. Version 1. Pretoria: Statistics South Africa [producer], 2012. Cape Town: DataFirst [distributor], 2013. Ref. ZAF_2010_IES_v01_M. Downloaded from https://microdata.worldbank.org/index.php/catalog/1545 on 2025-05-16.

Joseph E Stiglitz and Andrew Weiss. Credit rationing in markets with imperfect information. *The American economic review*, 71(3):393–410, 1981.

Andy Sumner and Arief Anshory Yusuf. New estimates of the cost of ending poverty and its global distribution. *WIDER Working Paper Series*, 2024. URL https://ideas.repec.org//p/unu/wpaper/wp-2024-46.html. Number: wp-2024-46 Publisher: World Institute for Development Economic Research (UNU-WIDER).

Hao Sun, Evan Munro, Georgy Kalashnov, Shuyang Du, and Stefan Wager. Treatment allocation under uncertain costs. *arXiv preprint arXiv:2103.11066*, 2021.

Liyang Sun. Empirical welfare maximization with constraints. *arXiv preprint arXiv:2103.15298*, 2021.

Erik Sverdrup, Han Wu, Susan Athey, and Stefan Wager. Qini curves for multi-armed treatment rules. *arXiv preprint arXiv:2306.11979*, 2023.

Uganda Bureau of Statistics (UBOS). National panel survey 2019-2020, 2021. Ref. UGA_2019_UNPS_v03_M. Downloaded from https://microdata.worldbank.org/index.php/catalog/3902 on 2025-03-01.

UN Statistics. Global indicator framework for the sustainable development goals and targets of the 2030 agenda for sustainable development. *Developmental science and sustainable development goals for children and youth*, 439, 2019.

Vladimir N. Vapnik. *Statistical Learning Theory*. Wiley, New York, 1998. ISBN 978-0-471-03003-4.

WAEMU Commission. Enquête harmonisée sur le conditions de vie des ménages (ehcvm) 2018-2019, 2022a. Ref. GNB_2018_EHCVM_v02_M. Downloaded from https://microdata.worldbank.org/index.php/catalog/4293 on 2025-05-08.

WAEMU Commission. Enquête harmonisée sur le conditions de vie des ménages 2018-2019, 2022b. Ref. NER_2018_EHCVM_v02_M. Downloaded from https://microdata.worldbank.org/index.php/catalog/study/NER_2018_EHCVM_v02_M on 2025-05-20.

WAEMU Commission. Enquête harmonisée sur le conditions de vie des ménages 2018-2019, 2022c. Ref. TGO_2018_EHCVM_v02_M. Downloaded from https://microdata.worldbank.org/index.php/catalog/study/TGO_2018_EHCVM_v02_M on 2025-02-12.

Yuanjia Wang, Haoda Fu, and Donglin Zeng. Learning optimal personalized treatment rules in consideration of benefit and risk: with an application to treating type 2 diabetes patients with insulin therapies. *Journal of the American Statistical Association*, 113(521):1–13, 2018.

World Bank. Living standards measurement study, 2014. URL https://www.worldbank.org/en/programs/lsms.

World Bank. World bank open data. https://data.worldbank.org/, 2023.

World Bank. World bank poverty and inequality platform. https://pip.worldbank.org/, 2025a.

World Bank. Poverty and inequality platform methodology handbook. Technical report, World Bank, July 2025b. URL https://datanalytics.worldbank.org/PIP-Methodology/.

World Bank. Statistics online (SOL) portal. https://pip.worldbank.org/sol/app, 2025c. Accessed 2025–2026.

World Data Lab. World poverty clock. https://worldpoverty.io, 2022. URL https://worldpoverty.io. Accessed: October 2024.

Christopher Yeh, Anthony Perez, Anne Driscoll, George Azzari, Zhongyi Tang, David Lobell, Stefano Ermon, and Marshall Burke. Using publicly available satellite imagery and deep learning to understand economic well-being in Africa. *Nature Communications*, 11(1):2583, May 2020. ISSN 2041-1723. doi: 10.1038/s41467-020-16185-w. URL https://www.nature.com/articles/s41467-020-16185-w. Number: 1.

Eitan Zemel. The linear multiple choice knapsack problem. *Operations Research*, 28(6):1412–1423, 1980.

Zimbabwe National Statistics Agency (ZIMSTAT). Poverty income consumption and expenditure survey (pices) 2017, 2017.

A Algorithms

In this section, we outline algorithms for learning unrestricted policies that minimize the poverty gap index, minimize the poverty rate, and maximize welfare, as well as binary policies that minimize an arbitrary loss function. We also discuss the weakly equitable oracle policy that minimizes the poverty rate.

A.1 Poverty Gap Minimization

We aim to solve (2) where the objective corresponds to the poverty gap index. In this case, (2) has a convex objective and is straightforward to solve using convex duality. We provide a characterization of the optimal gap-minimizing policy, a high-level overview for learning the optimal gap-minimizing policy in the population regime where we have access to the population distribution F, and finite-sample implementation details.

Lemma 4. *Suppose that Assumption 1 holds. If $L(z) = (c - z)_+$ and $0 < B < c$, then the unique optimal policy that solves (2) satisfies*

$$t^*_{\text{gap}}(x) = (c - F^{-1}_{Y|X=x}(\lambda^*(B)))_+. \tag{8}$$

where $\lambda^*(B) \in [0, 1]$ and $\lambda^*(B)$ chosen so that the budget constraint is satisfied with equality.

Lemma 4 implies that the family of optimal gap-minimizing policies can be parametrized as

$$t_\lambda(x) = (c - F^{-1}_{Y|X=x}(\lambda))_+$$

for $\lambda \in [0, 1]$. Each policy in this family minimizes the poverty gap index for some budget B. Given this functional form, learning the optimal gap-minimizing policy for a particular budget B boils down to finding the appropriate choice of $\lambda^*(B)$. We can obtain the optimal gap-minimizing policy by discretizing the interval $[0, 1]$ into a grid of quantiles $\lambda_1, \lambda_2, \dots \lambda_J$ and selecting $j \in [J]$ that minimizes the difference $|\mathbb{E}_F \left[t_{\lambda_j}(X)\right] - B|$.

A.1.1 Finite-Data Regime

In realistic settings, we only have access to data sampled from F. In this section, we assume the true data distribution F is unknown but we have access to a training set $\mathcal{D}_{\text{train}} = \{(X_i, Y_i)\}_{i=1}^{n_{\text{train}}}$ where $(X_i, Y_i) \sim F$ i.i.d., and we aim to optimize allocations for unlabeled samples $\mathcal{D}_{\text{test}} = \{X_i\}_{i=1}^{n_{\text{test}}}$ where $X_i \sim F$ i.i.d. We describe the algorithm for estimating the optimal gap-minimizing policy for a budget B.

First, we discretize the interval $[0, 1]$ into a grid of quantiles $\lambda_1, \lambda_2, \dots \lambda_J$. Second, estimate each $q_{\lambda_j}(\cdot)$ for $j = 1, 2 \dots J$ with minimal distributional assumptions on F. Using the training set $\mathcal{D}_{\text{train}} = \{(X_i, Y_i)\}$, we estimate each $\hat{q}_{\lambda_j}(\cdot)$ by solving a conditional quantile regression problem for quantile λ_j via deep learning. In our implementation, for each j, we train a neural network to minimize the pinball loss parametrized by the quantile λ_j [Koenker and Bassett, 1978]. Recall that the pinball loss is given by

$$L_{\text{pinball}}(z; \lambda) = \lambda \cdot |z| \cdot \mathbb{I}(z \geq 0) + (1 - \lambda) \cdot |z| \cdot \mathbb{I}(z < 0),$$

and we have that

$$q_\lambda(\cdot) \in \operatorname*{argmin}_q \mathbb{E}_F \left[L_{\text{pinball}}(Y - q(X); \lambda)\right].$$

Our implementation represents q with a neural network and solves the following empirical risk minimization problem

$$\hat{q}_{\lambda_j}(\cdot) \in \operatorname*{argmin}_q \widehat{\mathbb{E}}_F \left[L_{\text{pinball}}(Y - q(X); \lambda_j) \right].$$

After estimating the conditional quantiles, we use covariates from the test set $\mathcal{D}_{\text{test}} = \{X_i\}_{i=1}^{n_{\text{test}}}$ to estimate the optimal transfer amount for each unit in the test set.

$$\hat{t}_{\lambda_j}(X_i) := (c - \hat{q}_{\lambda_j}(X_i))_+.$$

For each λ_j, we can estimate the empirical policy cost. Finally, we can select the choice of λ_j so that $\widehat{\mathbb{E}}_F \left[t_{\lambda_j}(X_i) \right] = \frac{1}{n_{\text{test}}} \sum_{i=1}^{n_{\text{test}}} \hat{t}_{\lambda_j}(X_i)$ is closest to B.

A.2 Poverty Rate Minimization

We aim to solve (2) where the objective corresponds to the poverty rate. In this case, (2) is not straightforward to solve because it has a non-convex objective. Nevertheless, we provide a characterization of the optimal deterministic rate-minimizing policy and outline an algorithm for learning a stochastic policy that is guaranteed to obtain lower or equal poverty rate than the optimal deterministic rate-minimizing policy. We also provide a high-level overview of the algorithm in the population regime where we have access to the population distribution F and finite-sample implementation details.

In this section, we make the following technical assumption that the covariate space $\mathcal{X}$ is finite and discrete to permit t to be finite-dimensional because the existence of minimizers of non-convex functionals over infinite-dimensional spaces is not guaranteed.

Assumption 2. The covariate space $\mathcal{X}$ is finite and discrete.

We show that when $L(z) := \mathbb{I}(z < c)$, (2) has an optimal deterministic policy, and any optimal deterministic policy has a highly-structured form; it satisfies a property called α-validity for some $\alpha > 0$.

Definition 5. For any $\alpha \geq 0$ and covariate $x \in \mathcal{X}$, the set of α-valid transfers at x is

$$\mathcal{T}_\alpha(x) := \{0, c\} \cup \{t \in (0, c) \mid f_{Y|X=x}(c - t) = \alpha\}. \tag{9}$$

Using the definition of α-valid transfers, we can now define an α-valid transfer policy.

Definition 6. A transfer policy $t(\cdot)$ is α-valid if $t(x) \in \mathcal{T}_\alpha(x)$ for every $x \in \mathcal{X}$. Note that $t(x)$ may make a potentially randomized choice from the set.

Theorem 5. *Suppose that Assumptions 1, 2 hold, and $L(z) := \mathbb{I}(z < c)$ and the maximum transfer size is c. There exists an optimal deterministic policy that solves (2) and any optimal deterministic policy must be α-valid for some $\alpha \geq 0$.*

Note that due to the non-convexity of the objective, the optimal deterministic policy is not necessarily unique. To find an optimal deterministic policy, we can recast poverty rate minimization as a two-level optimization problem, where the outer optimization sweeps over all possible α values and the inner optimization finds an optimal deterministic α-valid policy among the class of deterministic α-valid policies.

Corollary 6. *Suppose that Assumptions 1, 2 hold. If $L(z) := \mathbb{I}(z < c)$, (2) can be re-written as*

$$\min_{\alpha \geq 0} \min_{t: \mathcal{X} \to \mathbb{R}_+} \{\mathbb{P}_F \left[t(X) + Y < c \right] : \mathbb{E}_F \left[t(X) \right] \leq B \text{ and } t \text{ is } \alpha\text{-valid}\}. \tag{10}$$

Note that the inner optimization problem is always feasible because $t(x) = 0$ satisfies the budget constraint and is α-valid. The inner optimization problem of (10) finds an optimal deterministic α-valid policy among the class of deterministic α-valid policies. We demonstrate that this is equivalent to solving a multiple-choice knapsack problem.

For technical convenience, we assume that for any covariate $x \in \mathcal{X}$, the set of α-valid transfers $\mathcal{T}_\alpha(x)$ contains at most K points for some $K < \infty$.

Assumption 3. There exists $K < \infty$ such that $\sup_{x \in \mathcal{X}, \alpha \in \mathcal{A}} |\mathcal{T}_\alpha(x)| \leq K$.

This assumption requires the conditional distributions $f_{Y|X}$ to have a finite number of modes.

Let $\mathcal{A}$ be the space of plausible α values. We note that $\mathcal{A}$ is bounded because $\alpha \geq 0$ and $\alpha \leq \sup_{x \in \mathcal{X}, y \in \mathcal{Y}} f_{Y|X=x}(y)$. Define $z : \mathcal{X} \times \mathcal{A} \to \mathbb{R}^K$, where $z_k(x; \alpha)$ is the k-th smallest element of $\mathcal{T}_\alpha(x)$ and $z_k(x; \alpha) = c$ for $|\mathcal{T}_\alpha(x)| < k \leq K$. Let $p : \mathcal{X} \times \mathcal{A} \to \mathbb{R}^K$, where $p_k(x; \alpha) = \mathbb{P}_F \left[z_k(X; \alpha) + Y < c \mid X = x \right]$. Note that z is a vector that captures the α-valid transfer amounts, and p is a vector that captures the conditional post-transfer poverty rates at these amounts.

Corollary 7. *Fix $\alpha > 0$. Let $\pi : \mathcal{X} \to [0,1]^K$. Under Assumption 3, the solution to the inner optimization problem in (10) is given by*

$$t(x) = \langle \pi(x), z(x; \alpha) \rangle$$

where π is the solution to a multiple-choice knapsack problem:

$$\begin{aligned} \text{minimize } & \mathbb{E}_F \left[\langle \pi(X), p(X; \alpha) \rangle \right] \\ \text{subject to } & \mathbb{E}_F \left[\langle \pi(X), z(X; \alpha) \rangle \right] \leq B \\ & \langle \pi(x), \mathbf{1} \rangle = 1 \quad \forall x \in \mathcal{X} \\ & \pi_k(x) \in \{0, 1\} \quad \forall x \in \mathcal{X}, k \in [K]. \end{aligned} \tag{11}$$

The program in (11) is a multiple-choice knapsack problem. In the multiple-choice knapsack problem, the goal is to fill a knapsack up to a capacity B by selecting exactly one item out of each class of items, where each item has an associated "loss" and "weight." In our problem, each class corresponds to a unit with covariates x and the items in the class are the α-valid transfers $\mathcal{T}_\alpha(x)$. For clarity of exposition, if $\mathcal{T}_\alpha(x)$ contains less than K elements, we pad the class of items with transfers c in the formulation in (11) until each class consists of K transfers. For item k in the class corresponding to covariates x, the loss is given by the conditional post-transfer poverty rate $p_k(x; \alpha)$ and the weight is given by the transfer amount $z_k(x; \alpha)$.

Whereas the standard multiple-choice knapsack problem is NP-hard, its fractional relaxation can be solved using a computationally efficient algorithm. We leverage this connection to develop a practical algorithm for solving (11). In the fractional formulation, we permit the selection of fractional amounts of items, which permits fractional values $\pi_k(x) \in [0, 1]$. We emphasize that in the absence of additional assumptions, solving the fractional relaxation of (11) yields a stochastic policy that is guaranteed to obtain a poverty rate less than or equal to the optimal deterministic α-valid policy. We provide conditions under which the solution to the fractional relaxation is equal to the solution to the original problem (11) in Appendix A.2.1.

Thus, our overall algorithm for solving (2) where the objective corresponds to the poverty rate relies on the representation provided in (10). We define a grid $[\alpha_1, \alpha_2, \ldots, \alpha_J]$ over $\mathcal{A}$ and solve the fractional knapsack problem for every α_j in the grid to obtain a policy $t_{\alpha_j}(\cdot)$. Finally, we return the policy t^*_{rate} that yields the lowest poverty rate over all t_α for α in the grid.

A.2.1 Algorithm

We provide an algorithm for poverty rate minimization, which leverages a computationally efficient algorithm for solving the fractional multiple-choice knapsack problem.

Input : Conditional distributions $\{f_{Y|X=x}(\cdot)\}_{x\in\mathcal{X}}$,
Covariate distribution $f_X(\cdot)$,
Budget B,
Threshold c,
Grid size m
Output: Transfer policy t^*

rates $\leftarrow \emptyset$
policies $\leftarrow \emptyset$
Define $\alpha_{\max}$ using $\{f_{Y|X=x}(\cdot)\}_{x\in\mathcal{X}}$.
for $\alpha \in [\frac{\alpha_{\max}}{m}, \frac{2\alpha_{\max}}{m}, \dots \frac{(m-1)\cdot\alpha_{\max}}{m}, \alpha_{\max}]$ **do**
  for $x \in \mathcal{X}$ **do**
    $\mathcal{T}_\alpha(x) \leftarrow$ `ComputeAlphaValidTransfers(`$f_{Y|X=x}(\cdot)$, α, c`)`
    $\mathcal{C}_\alpha(x), \rho(x;\alpha) \leftarrow$ `ComputeLowerConvexHull(`$f_{Y|X=x}(\cdot)$, $\mathcal{T}_\alpha(x)$`)`
  end
  Apply Algorithm 2
  t_α, rate $\leftarrow$ `SolveFractionalMCKnapsack(`$f_X(\cdot)\,\{(\mathcal{C}_\alpha(x), \rho(x;\alpha))\}_{x\in\mathcal{X}}$, B`)`
  rates.append(rate)
  policies.append(t_α)
end
$t^* \leftarrow$ Minimum poverty rate policy in policies (based on rates)
return t^*

Algorithm 1: Two-level optimization procedure for poverty rate minimization.

The outer loop of our algorithm is a grid search over plausible α values. Let $C := \sup_{x\in\mathcal{X}, y\in\mathcal{Y}} f_{Y|X=x}(y)$. We note that for any $\alpha > C$ the set $\mathcal{T}_\alpha(x)$ will only consist of $\{0, c\}$ for all x, meaning that the inner optimization problem is identical for these values of α. As a result, it is sufficient to define $\alpha_{\max} = C + \delta$ for some small $\delta > 0$ and restrict α to $[0, \alpha_{\max}]$. In Algorithm 1, we grid this interval into m values.

The inner loop of our algorithm finds the optimal α-valid policy among the set of stochastic α-valid policies by solving a fractional multiple-choice knapsack problem. Nevertheless, the solution is guaranteed to have equal or lower policy cost than the optimal deterministic α-valid policy.

We solve the fractional multiple-choice knapsack problem using the computationally-efficient algorithm of Zemel [1980]. This procedure is also used by Sverdrup et al. [2023] to solve a cost-constrained treatment allocation problem. The algorithm relies on the key observation that in the optimal solution to the fractional relaxation of (11), the only transfer amounts that are active are the ones that lie on the lower convex hull of the loss-weight plane. For any $x \in \mathcal{X}$, define the lower convex hull of $(z_k(x;\alpha), p_k(x;\alpha))$ for $k = 1, \dots K$ to be a set $\mathcal{C}_\alpha(x)$ of points

with the ordering $k_1(x), \dots k_{|\mathcal{C}_\alpha(x)|}(x)$ such that

$$\begin{aligned}
z_{k_1(x)}(x;\alpha) &< z_{k_2(x)}(x;\alpha) < \cdots < z_{k_{|\mathcal{C}_\alpha(x)|}(x)}(x;\alpha), \\
p_{k_1(x)}(x;\alpha) &> p_{k_2(x)}(x;\alpha) > \cdots > p_{k_{|\mathcal{C}_\alpha(x)|}(x)}(x;\alpha), \\
\rho_{k_1(x)}(x;\alpha) &< \rho_{k_2(x)}(x;\alpha) < \cdots < \rho_{k_{|\mathcal{C}_\alpha(x)|}}(x;\alpha) < 0,
\end{aligned}$$

where ρ is the incremental cost-loss ratio for points on this convex hull as

$$\rho_{k_j(x)}(x;\alpha) := \frac{p_{k_{j+1}(x)}(x;\alpha) - p_{k_j(x)}(x;\alpha)}{z_{k_{j+1}(x)}(x;\alpha) - z_{k_j(x)}(x;\alpha)} \quad k_{j+1}(x), k_j(x) \in \mathcal{C}_\alpha(x) \tag{12}$$

and let $\rho_k(x) := \infty$ if $k \notin \{k_1(x), \dots k_{|\mathcal{C}_\alpha(x)|}(x)\}$.

We define a thresholding rule on the incremental cost-loss ratio (12).

Definition 7 (Thresholding rule, [Sverdrup et al., 2023])**.** Given cost-loss ratios ρ as in (12), a threshold $\lambda \leq 0$, and an interpolation value $\eta \in [0, 1)$, we define a thresholding rule as

$$T_{k_j(x)}(x;\rho,\lambda,\eta) = \begin{cases} 1 & \text{if } \rho_{k_j(x)}(x) < \lambda < \rho_{k_{j+1}(x)}(x) \\ \eta & \text{if } \rho_{k_j(x)}(x) = \lambda \\ 1-\eta & \text{if } \rho_{k_{j+1}(x)}(x) = \lambda \\ 0 & \text{o.w.} \end{cases}.$$

The fractional multiple-choice knapsack algorithm can be used to obtain the optimal (stochastic) policy. For a detailed description of the algorithm, we refer readers to Algorithm 2. At the start of the procedure, all units are assigned the transfer amount 0, which corresponds to the first point $k_1(x)$ on their convex hulls. Then, we initialize a priority queue. For each $x \in \mathcal{X}$, we add the pair of x and the second point $k_2(x)$ on its convex hull $\mathcal{C}(x)$ to the queue with priority $\rho_{k_1}(x)$. While the post-transfer poverty rate constraint has slack and the queue is nonempty, we pop points from the queue. When a unit with covariates x and point $k_j(x)$ on its convex hull is popped from the queue, the unit is assigned the corresponding transfer value $z_{k_j(x)}(x)$. If the unit has additional points on its convex hull, then the next point on its convex hull is added to the queue with priority equal to its cost-weight ratio $\rho_{k_{j+1}(x)}$. The sequence of updates is dictated by the priority queue ordered by ρ. When solving the problem in the population case, the time-complexity of the algorithm is log-linear in $|\mathcal{X}| \cdot K$. While this time-complexity may seem prohibitively large when the covariate space is continuous, in practice, this algorithm is only applied to units in the test set, so the time-complexity is at most $n_{\text{test}} \cdot K$.

The time complexity of this algorithm is $|\mathcal{X}| \cdot K \log |\mathcal{X}| \cdot K$. The worst-case run time arises when $B > \mathbb{E}_F \left[z_{k_{|\mathcal{C}_\alpha(X)|}}(X) \right]$, so all points on the convex hull for all units will be added and removed from the priority queue. Since there are at most $|\mathcal{X}| \cdot K$ points added and removed this yields a complexity of $|\mathcal{X}| \cdot K \log |\mathcal{X}| \cdot K$.

A.2.2 Finite-Data Regime

Thus far, we have given a characterization of the optimal deterministic policy that minimizes the poverty rate and a practical algorithm for poverty rate minimization. However, in realistic settings, we only have access to data sampled from F. In this section, we assume the true data distribution F is unknown but we have access to a training set $\mathcal{D}_{\text{train}} = \{(X_i, Y_i)\}_{i=1}^{n_{\text{train}}}$ where $(X_i, Y_i) \sim F$ i.i.d., and we aim to optimize allocations for unlabeled samples $\mathcal{D}_{\text{test}} = \{X_i\}_{i=1}^{n_{\text{test}}}$ where $X_i \sim F$ i.i.d.

Input : Convex hulls and incremental cost-loss ratios $\{(\mathcal{C}(x), \rho(x))\}_{x \in \mathcal{X}}$,
Covariate distribution $f_X(\cdot)$,
Budget B

Output: Transfer policy t, Poverty rate loss

Initialize priority queue.
cost $\leftarrow 0$, loss $\leftarrow 0$
queue $\leftarrow \emptyset$
for $x \in \mathcal{X}$ **do**
  Assign unit to first point on convex hull.
  $k_1(x) \leftarrow$ First point on convex hull $\mathcal{C}(x)$
  cost $+= f_X(x) \cdot z_{k_1(x)}(x)$, loss $+= f_X(x) \cdot p_{k_1(x)}(x)$
  Enqueue next point on hull.
  $k_2(x) \leftarrow$ Second point on convex hull $\mathcal{C}(x)$
  queue.add($(x, k_2(x))$ with priority $\rho_{k_1(x))}$)
end
while cost $< B$ *and* queue.*size()* > 0 **do**
  $(x, k_j(x)) \leftarrow$ queue.pop()
  Subtract current transfer to unit x from cost *and* loss.
  cost$- = f_X(x) \cdot z_{k_{j-1}(x)}(x)$, loss$- = f_X(x) \cdot p_{k_{j-1}(x)}(x)$
  Allocate transfer $k_j(x)$ to unit x, record new cost *and* loss.
  $t(x) \leftarrow z_{k_j(x)}(x)$
  cost$+ = f_X(x) \cdot z_{k_j(x)}(x)$, loss$+ = f_X(x) \cdot p_{k_j(x)}(x)$
  if cost $> B$ **then**
    Perform fractional adjustment for unit x– updating t, cost, loss.
    break
  end
  if *there remain points on convex hull for unit x* **then**
    $k_{j+1}(x) \leftarrow$Next point on $\mathcal{C}(x)$
    queue.add($(x, k_{j+1}(x))$ with priority $\rho_{k_j(x)}$)
  end
end
return t, loss.

Algorithm 2: Solve fractional multi-choice knapsack.

Algorithm 1 relies on the conditional density function $f_{Y|X}$ to compute the optimal transfer policy. In finite samples, we consider the plug-in estimator for the optimal transfer policy, which is obtained by running Algorithm 1 with an estimator of the conditional density $\hat{f}_{Y|X}$ and the empirical covariate distribution $\hat{f}_X(x) = \frac{1}{n_{\text{test}}} \cdot \mathbb{I}(X_i \in \mathcal{D}_{\text{test}})$ instead of the true conditional and covariate distributions.

To estimate $\hat{f}_{Y|X}$, we apply an extension of Lindsey's method [Efron and Tibshirani, 1996] to the training data $\mathcal{D}_{\text{train}}$. Lindsey's method is a popular technique for marginal density estimation, and our approach is a straightforward extension of this method for conditional density estimation. To model the conditional densities, we consider the following exponential family of densities on $\mathcal{Y}$

$$h(y \mid x) = h_0(y) \cdot \exp(s(y)^T \Theta x - \psi(\Theta x)). \tag{13}$$

Here, $h_0(y)$ represents a carrier density, $s(y)$ is a J-dimensional vector of sufficient statistics, Θ is a $J \times d$ matrix of parameters, and ψ is the log partition function, the normalizing function that ensures $h(y \mid x)$ integrates to 1 over $\mathcal{Y}$. Many choices of carrier density and sufficient statistics are possible. Following Lindsey's method, we set the carrier density to be a nonparametric estimate of f_Y, the marginal distribution of Y. In addition, we consider a basis of B-Spline functions to form the sufficient statistics.

We compute the conditional density estimator as follows.

1. We estimate f_Y, the marginal density of Y, by computing a kernel density estimate using $\{Y_i\}_{i=1}^{n_{\text{train}}}$. We denote this estimate as $\hat{f}_Y$.

2. Setting $\hat{f}_Y$ as the carrier density in (13), we estimate Θ via maximum likelihood. To estimate the log-likelihood of (13), we first approximate the log-partition function by discretizing $\mathcal{Y}$ into M bins with midpoints $[y_1, \dots y_M]$ and corresponding widths $[\Delta_1, \Delta_2, \dots \Delta_M]$. We note that the log partition function can be approximated by

$$\psi(z) = \log(\int \hat{f}_Y(y) \cdot \exp(s(y)^T z)) \approx \log(\sum_{j=1}^{M} \hat{f}_Y(y_j) \cdot \exp(s(y_j)^T z) \cdot \Delta_j).$$

   Thus, the log-likelihood of (13) can be written as

$$\log h(y_i \mid x_i; \Theta) = s(y_i)^T \Theta x_i - \log\Big(\sum_{j=1}^{M} \hat{f}_Y(y_j) \cdot \exp(s(y_j)^T \Theta x_i) \cdot \Delta_j\Big) + c, \tag{14}$$

   where c is a constant factor that does not depend on Θ. Notably, the log-likelihood is concave in Θ, so standard optimization tools can be applied to obtain an estimate of Θ. In our work, we optimize Θ via stochastic gradient descent.

Like Lindsey's method, this is a hybrid approach between a parametric and nonparametric method because it models the relationship between high-dimensional covariates and the density value with a generalized linear model but allows the density function to take on a flexible form by fitting the carrier density nonparametrically.

A.3 Welfare Maximization

We formally define the problem of welfare maximization. Let $w : \mathbb{R}_+ \to \mathbb{R}$ be an increasing, strictly concave, differentiable welfare function that is bounded above by a constant $C < \infty$.

Welfare maximization is defined as the following problem

$$\max_{t:\mathcal{X}\to\mathbb{R}_+} \{\mathbb{E}_F\left[w(Y+t(X))\right] : \mathbb{E}_F\left[t(X)\right] \leq B\}. \tag{15}$$

Note that welfare maximization is analogous to loss minimization, so we can apply the same techniques as in Theorem 2 to obtain the optimal policy for welfare maximization.

A.3.1 Algorithm

We compute the solution to a discrete approximation of (15). In our approximation, we consider policies that only allocate transfer amounts that lie in a discrete grid. In particular, let $\mathcal{T} := [t_1, t_2, \dots t_J]$ denote a grid of transfer amounts on $[0, \bar{t}]$, where $\bar{t}$ is the maximum transfer size.

First, we define the function $G_x(t) := \mathbb{E}_F\left[w(Y)+t) \mid X = x\right]$. We obtain a discrete approximation to the inverse function $(G'_x)^{-1} : \mathbb{R}_+ \to \mathbb{R}_+$. Let the function $H_x : \mathbb{R}_+ \to \mathcal{T}$ denote the discrete approximation. To construct H_x, obtain $G'_x(t) : \mathcal{X} \mapsto \mathbb{R}_+$ for $t \in \mathcal{T}$. Next, use the functions $\{G'_x(t)\}_{t\in\mathcal{T}}$ to define

$$H_x(\lambda) = t_{j^*}, \text{ where } j^* = \sup\{j \in [J] \mid G'_x(t_j) \geq \lambda\}.$$

Thus, $H_x(\cdot)$ is a discrete approximation of $(G'_x)^{-1}(\cdot)$.

Second, use $H_x(\cdot)$ to find the dual parameter λ corresponding to budget B. Define bounds $\lambda_{\min} := \inf_{x\in\mathcal{X}, j\in[J]} G'_x(t_j)$ and $\lambda_{\max} := \sup_{x\in\mathcal{X}, j\in[J]} G'_x(t_j)$. Consider a grid $\{\lambda_1, \lambda_2, \dots \lambda_J\}$ on $[\lambda_{\min}, \lambda_{\max}]$. We return the policy $t_j(x) = (H_x(\lambda_j))_+$ that minimizes $|\mathbb{E}_F\left[(H_X(\lambda_j))_+\right] - B|$ over the grid.

A.3.2 Finite-Data Regime

We consider the setting where we only have access to data sampled from F. In this section, we assume access to a training set $\mathcal{D}_{\text{train}} = \{(X_i, Y_i)\}_{i=1}^{n_{\text{train}}}$, where $(X_i, Y_i) \sim F$ i.i.d., and we aim to optimize allocations for unlabeled samples $\mathcal{D}_{\text{test}} = \{X_i\}_{i=1}^{n_{\text{test}}}$ where $X_i \sim F$ i.i.d. We use the following procedure to estimate the optimal policy for a budget B.

In finite samples, we must estimate the functional parameters that appear in the above algorithm. In particular, we must estimate $G'_x(t)$ for $t \in \mathcal{T}$ from the training data. To do so, we can apply any nonparametric regression technique to solve

$$\min_{h:\mathcal{X}\to\mathbb{R}_+} \mathbb{E}_F\left[(w'(t+Y) - h(X))^2\right].$$

A.4 Binary Policies

In this section, we solve (2) with the additional constraint that the policy must be binary-valued. In contrast to proxy means testing as in Definition 1, we minimize a particular poverty measure and optimize not only who receives the transfer but also the optimal transfer size.

Learning the optimal binary transfer policy that minimizes a particular loss function L among all possible binary transfer policy can be framed as the following optimization problem

$$\begin{aligned} \text{minimize } & \mathbb{E}_F\left[L(t(X)+Y)\right] \\ \text{subject to } & \mathbb{E}_F\left[t(X)\right] \leq B \\ & t(x) \in \{0, \bar{t}\} \quad \forall x \in \mathcal{X} \\ & \bar{t} \in [0, c]. \end{aligned} \tag{16}$$

However, the above optimization problem is non-convex because it involves jointly optimizing over the maximum transfer value $\bar{t}$ and the transfer policy t.

We can rewrite (16) as follows

$$\min_{\bar{t}\in[0,c]}\left\{\min_{\pi:\mathcal{X}\to\{0,1\}}\mathbb{E}_F\left[L(\bar{t}+Y)\cdot\pi(X)+L(Y)\cdot(1-\pi(X))\right]\text{ subject to }\mathbb{E}_F\left[\pi(X)\right]\leq B/\bar{t}\right\}. \tag{17}$$

Solving the inner minimization of (17) allows us to compute the optimal binary transfer policy among binary transfer policies where the maximum transfer value is a fixed value $\bar{t}$. The inner minimization of (17) can be viewed as a capacity-constrained classification problem. Let $\pi_{\bar{t}}(\cdot;B)$ be the solution to the inner minimization of (17). The corresponding binary transfer policy can be defined as $t_{\bar{t}}(\cdot;B) := \bar{t}\cdot\pi_{\bar{t}}(x;B)$.

The function $\pi_{\bar{t}}(\cdot;B)$ identifies the units who will benefit most (in terms of conditional loss reduction) from the $\bar{t}$-valued transfer. We can define the conditional loss reduction function for a $\bar{t}$-valued transfer

$$\rho_{\bar{t}}(x) := \mathbb{E}_F\left[L(\bar{t}+Y)-L(Y)\mid X=x\right]. \tag{18}$$

For poverty measures, L is decreasing so $\rho_{\bar{t}}(x)\leq 0$ for any $\bar{t}\in\mathbb{R}_+$.

Following from Theorem 1 from Sun et al. [2021], there are constants $\rho^*_{\bar{t},B}\in\mathbb{R}, a_{\bar{t},B}\in[0,1]$, such that the optimal solution to the inner minimization of (17) has the form

$$\pi_{\bar{t}}(x;B)=\begin{cases}1 & \rho_{\bar{t}}(x)<\rho^*_{\bar{t},B}\\ a_{\bar{t},B} & \rho_{\bar{t}}(x)=\rho^*_{\bar{t},B}\\ 0 & \rho_{\bar{t}}(x)>\rho^*_{\bar{t},B},\end{cases} \tag{19}$$

where either $\rho^*_{\bar{t},B}=a_{\bar{t},B}=0$ (i.e., we have sufficient budget to treat all units) or $\rho^*_{\bar{t},B}<0$ and the pair $(\rho^*_{\bar{t},B},a_{\bar{t},B})$ is the unique pair for which the policy cost is B. If X_i has continuous density, $\mathbb{P}_F\left[\rho_{\bar{t}}(X)=\rho^*_{\bar{t},B}\right]=0$ and the policy $\pi_{\bar{t}}(\cdot;B)$, and so the optimal policy will almost surely be integer-valued.

We operationalize this result to provide an algorithm for optimal binary policies. Our overall algorithm first discretizes the interval $[0,c]$ into a grid of possible maximum transfer amounts $[\bar{t}_1,\bar{t}_2,\ldots\bar{t}_J]$. For each $j=1,2,\ldots J$, we compute $\pi_{\bar{t}_j}(\cdot;B)$ by ranking units by $\rho_{\bar{t}_j}(\cdot)$ in increasing order and allocating transfers $\bar{t}_j$ sequentially until the budget B is exhausted. We set the optimal binary policy t^*_{binary} under loss function L to be the policy that yields the lowest expected loss over the grid, i.e.

$$t^*_{\text{binary}}(x;B)=\bar{t}_{j^*}\cdot\pi_{\bar{t}_{j^*}}(x;B),\quad\text{where } j^*=\operatorname*{argmin}_{j\in[J]}\mathbb{E}_F\left[L(t_{\bar{t}_j}(X_i;B)+Y_i)\right].$$

We note that this procedure will yield the optimal binary-valued policy under any decreasing poverty measure L and does not require the loss function L to be convex. As a result, we will use this approach to learn optimal binary-valued policies under both the poverty rate and the poverty gap index.

A.4.1 Finite-Data Regime

Again, we consider the setting where we only have access to data sampled from F. In this section, we assume the true data distribution F is unknown but we have access to a training set $\mathcal{D}_{\text{train}}=\{(X_i,Y_i)\}_{i=1}^{n_{\text{train}}}$ where $(X_i,Y_i)\sim F$ i.i.d., and we aim to optimize allocations for unlabeled samples $\mathcal{D}_{\text{test}}=\{X_i\}_{i=1}^{n_{\text{test}}}$ where $X_i\sim F$ i.i.d. We use the following procedure to

estimate the optimal binary policy for a budget B.

First, we discretize the interval $[0, c]$ into a grid of possible maximum transfer amounts $\bar{t}_1, \bar{t}_2, \ldots \bar{t}_J$. Second, we estimate nuisance parameters $\rho_{\bar{t}_j}(\cdot)$ for $j = 1, 2, \ldots J$. Using the training set $\mathcal{D}_{\text{train}} = \{(X_i, Y_i)\}_{i=1}^{n_{\text{train}}}$, we define pseudo-labels

$$Z_i = L(\bar{t} + Y_i) - L(Y_i)$$

for $i = 1, 2, \ldots, n_{\text{train}}$. Estimate $\hat{\rho}_{\bar{t}_j}(\cdot)$ via deep learning using $\mathcal{D}_{\text{train}}$ and pseudolabels. Observe that

$$\rho_{\bar{t}} \in \operatorname*{argmin}_{\rho} \mathbb{E}_F \left[(Z_i - \rho(X_i))^2 \right].$$

Our implementation represents $\rho_{\bar{t}}$ with a neural network and solves the following empirical risk minimization problem

$$\hat{\rho}_{\bar{t}_j}(\cdot) \in \operatorname*{argmin}_{\rho} \widehat{\mathbb{E}}_F \left[(Z_i - \rho(X_i))^2 \right].$$

Now, we must learn the optimal transfer size $\bar{t}_j$. For each $j = 1, 2, \ldots J$, create a ranked list of units by sorting $\hat{\rho}_{\bar{t}_j}(X_i)$ in increasing order for X_i in the training set. We can estimate the threshold $\rho^*_{\bar{t}_j, B}$ by allocating transfers to the $B/\bar{t}_j$-fraction of units at the top of the ranked list for each $j = 1, 2, \ldots J$. After that, we can form $\hat{t}_{\bar{t}_j}(\cdot; B)$ using the estimated conditional loss reduction function $\hat{\rho}_{\bar{t}_j}(\cdot)$ and estimated threshold $\hat{\rho}^*_{\bar{t}_j, B}$. Then, we can select

$$\hat{j}^* \in \operatorname*{argmin}_{j \in [J]} \widehat{\mathbb{E}}_F \left[L(\hat{t}_{\bar{t}_j}(X_i; B) + Y_i) \right].$$

We assign transfers to units in the test set $\mathcal{D}_{\text{test}}$ by ranking units in the test set by $\hat{\rho}_{\hat{j}^*}$ in increasing order and allocating $\bar{t}_{\hat{j}^*}$-valued transfers to units starting at the top of the ranked list until we hit the budget constraint B, which ensures that the budget constraint is exactly satisfied on the test set.

A.5 Weakly Equitable Oracle Policy

This subsection focuses on oracle policies, which can allocate transfers on the basis of individual-level consumption. We characterize the policy family that minimizes the poverty rate subject to the weak equity constraint and the constraint that the policy family is continuous in budget. To characterize this policy, we first introduce a definition of weak rank preservation and show that weak equity implies this property for policy families that are continuous in budget. Finally, we use the property of weak rank preservation to characterize the optimal policy.

Definition 8 (Weak Rank Preservation)**.** A full information policy family $t(y; B)$ weakly preserves ranks if for $y_1, y_2 \in \mathcal{Y}$

$$y_1 + t(y_1; B') \leq y_2 + t(y_2; B') \implies y_1 + t(y_1; B) \leq y_2 + t(y_2; B)$$

for all $0 \leq B' \leq B$.

Transfer policies that weakly preserve ranks do not permit a poorer unit's post-transfer consumption to exceed that of a unit who is initially richer. This rules out the scenario where a unit who is initially poorer than another receives a transfer $t(y_1; B)$ that allows them to "leapfrog" the other unit in terms of post-transfer consumption, i.e. $y_1 + t(y_1; B) > y_2 + t(y_2; B)$. We can consider such policies as yielding post-transfer consumption that is monotone increasing in pre-transfer consumption.

We observe that weak equity (Definition 4) and continuity in budget imply this property.

Lemma 8. *If $t(y;B)$ is a full-information weakly equitable policy family and $t(y;B)$ is continuous in B, then $t(y;B)$ weakly preserves ranks.*

Let $\mathcal{W}$ denote the class of policy families that satisfy weak equity and are continuous in budget. The following lemma characterizes the optimal rate-minimizing policy family among policy families that are weakly equitable and continuous in budget.

Lemma 9. *Suppose F is a population distribution over Y_i that has positive density on $(0,c)$. Consider minimizing the post-transfer poverty rate subject to a budget constraint and a weak equity constraint in the full-information setting:*

$$\min_{t(\cdot;\cdot):\mathcal{Y}\times\mathbb{R}_+\to\mathbb{R}_+} \{\mathbb{P}_F\left[Y + t(Y;B) < c\right] : \mathbb{E}_F\left[t(Y;B)\right] \leq B,\, t \in \mathcal{W}\}. \tag{20}$$

The optimal policy that solves this optimization problem is

$$t(y;B) = \begin{cases} c - \lambda(B) & 0 \leq y \leq \lambda(B) \\ c - y & \lambda(B) < y \leq c \\ 0 & y > c. \end{cases}. \tag{21}$$

B Proofs

B.1 Additional Lemmas

We define new notation for the following lemmas. Define

$$\begin{aligned} G_x(t) &= -\mathbb{E}_F\left[L(Y+t) \mid X = x\right] \\ G'_x(t) &= -\frac{d}{dt}\mathbb{E}_F\left[L(Y+t) \mid X = x\right]. \end{aligned} \tag{22}$$

Lemma 10. *Suppose Assumption 1 holds. Let L be decreasing, strictly convex, differentiable, bounded below by $C > -\infty$. Then, $G_x(t)$ is differentiable on $\mathbb{R}_+$ and strictly concave.*

Lemma 11. *Suppose Assumption 1 holds. Let L denote an FGT index with $\alpha \geq 1$. Then, $G_x(t)$ is differentiable on $\mathbb{R}_+$ and strictly concave on $[0,c]$.*

B.2 Proof of Theorem 1

We consider two cases. We first focus on the case where L is nonconvex in a region where it is continuously differentiable. Then, we will consider the case that L is nonconvex in a region where it is not continuously differentiable.

Our proof relies on the truncated Gaussian kernel to construct distributions under which weak equity violations occur. The truncated Gaussian kernel is given by

$$dK_h(y;m) = \frac{\phi((y-m)/h)}{h \cdot (1 - \Phi(-m/h))} \quad y \geq 0 \tag{23}$$

In addition, as $h \to 0$, $dK_h(y;m)$ converges weakly to δ_m, a point mass at m. We also note that if $m_1 < m_2$, then $K_h(\cdot;m_1) \preceq_{\text{SD}} K_h(\cdot;m_2)$.

The second half of our proof also uses the Gamma kernel to smooth loss functions that are not continuously differentiable. The Gamma kernel is given by

$$dQ_b(z) = \frac{z}{b^2} e^{-z/b} \quad z \geq 0. \tag{24}$$

B.2.1 Continuously Differentiable Case

In the case where L is nonconvex in a region where it is continuously differentiable, we show that it is possible to construct a population distribution F_0 with the following properties

1. F_0 is a mixture distribution over two types, i.e. $F_X = \frac{1}{2} \cdot \mathbb{I}(X = x_1) + \frac{1}{2} \cdot \mathbb{I}(X = x_2)$.
2. The conditional distribution $F_{0,Y|X=x_i}$ is a point mass, i.e. $F_{0,Y|X=x_i} = \mathbb{I}(y = y_i)$. In addition, $0 < y_1 < y_2$, so $F_{Y|X=x_1} \preceq_{\text{SD}} F_{Y|X=x_2}$.
3. There exists a strictly unique optimal transfer policy under F_0 and budget $B > 0$ called $t_0^*(\cdot; B)$. The optimal transfer policy $t_0^*(\cdot, B)$ is not weakly equitable under F_0. In particular, $t_0^*(x_1; B) - t_0^*(x_2; B) < 0$.

We will use F_0 to construct a population distribution that satisfies Assumption 1 under which the optimal policy that solves (2) is not weakly equitable. Given that F_0 has conditional distribution $F_{0,Y|X=x_i} = \mathbb{I}(y = y_i)$, we define F_h be a distribution over (X, Y) where $F_{h,X} = F_{0,X}$ and $F_{h,Y|X} = K_h(y; y_i)$, where we recall that K_h is the truncated Gaussian kernel defined in (23).

Construction of F_0 and Budget B. Since L is non-convex on a continuously differentiable region, there exists two points $y_1 < y_2$ with $L'(y_1) > L'(y_2)$. Since L' is continuous, there exists $\delta_1(\epsilon_1) > 0$, so that $|L'(y_1) - L'(z)| < \epsilon_1$ for $|y_1 - z| < \delta_1(\epsilon_1)$. Again, by the continuity of L', for any ϵ_2, there exists $\delta_2(\epsilon_2) > 0$, so that $|L'(y_2) - L'(z)| < \epsilon$ for $|y_2 - z| < \delta_2(\epsilon_2)$. We can choose ϵ_1, ϵ_2 sufficiently small so that $L'(y_2) + \epsilon_2 < L'(y_1) - \epsilon_1$. Let $\epsilon = \min_{i \in \{1,2\}} \epsilon_i$. It is straightforward to see that

$$L'(y_2) + \epsilon < L'(y_1) - \epsilon. \tag{25}$$

Let $\Delta = \min(\delta_1(\epsilon), \delta_2(\epsilon))/2$. For $z_i \in B_\Delta(y_i)$, we have that

$$L'(z_2) < L'(y_2) + \epsilon < L'(y_1) - \epsilon < L'(z_1).$$

This implies that for $t \in [0, \Delta]$, we have that

$$L'(y_1 + t) \geq L'(y_1) - \epsilon \tag{26}$$

and

$$L'(y_2 + t) \leq L'(y_2) + \epsilon. \tag{27}$$

Combining (26), (27) and (25), implies that for all $t_1, t_2 \in [0, \Delta]$

$$L'(y_2 + t_2) < L'(y_1 + t_1).$$

Set $F_{0,Y|X=x_i} = \mathbb{I}(y = y_i)$ for $i \in \{1, 2\}$. Also, note that we have selected Δ so that L is continuous on open neighborhoods O_1, O_2 that contain $[y_1, y_1 + \Delta]$ and $[y_2, y_2 + \Delta]$, respectively.

Optimal Policy under F_0 and Budget $B = \Delta/2$. For this choice of population distribution F_0 and $B = \Delta/2$, the optimal policy that solves (2) is $t^*(x) = 0 \cdot \mathbb{I}(X = x_1) + \Delta \cdot \mathbb{I}(X = x_2)$. Note that this policy is non-monotone increasing in consumption because $F_{0,Y|X=x_1} \preceq_{\text{SD}} F_{0,Y|X=x_2}$ but $t^*(x_2) > t^*(x_1)$ and is thus, not weakly equitable.

We consider the following optimization problem to characterize the optimal policy. Denote the objective as $J(t_1, t_2) := \frac{1}{2}(L(y_1 + t_1) + L(y_2 + t_2))$.

$$\min_{t_1, t_2 \in [0,\Delta]} \Big\{ J(t_1, t_2) : t_1 + t_2 \leq \Delta \Big\}. \tag{28}$$

First, we show that the budget constraint binds. For the sake of contradiction, suppose that (t_1^*, t_2^*) is the optimal policy but $t_1^* + t_2^* < \Delta$. Then, we can consider the policy $(t_1^*, t_2^* + \epsilon)$. Then we have that

$$\begin{aligned} J(t_1^*, t_2^* + \epsilon) - J(t_1^*, t_2^*) &= \frac{1}{2}(L(y_2 + t_2^* + \epsilon) - L(y_2 + t_2^*)) \\ &= \frac{1}{2} \cdot L'(y_2 + t) \cdot \epsilon \end{aligned}$$

for some $t \in [t_2^*, t_2^* + \epsilon]$ by the Mean Value Theorem. Since $L'(y_2 + t_2) < L'(y_1 + t_1) \leq 0$ where $t_1, t_2 \in [0, \Delta]$, we have that $J(t_1^*, t_2^* + \epsilon) - J(t_1^*, t_2^*) < 0$. This is a contradiction because we assumed that (t_1^*, t_2^*) was the optimal policy. Thus, the budget constraint must bind.

Since the budget constraint must bind, we can restrict to solutions that satisfy $t_1 + t_2 = \Delta$. Suppose the optimal policy has $t_1^* > 0$ (and satisfies the budget constraint). We consider the policy given by $(t_1^* - \epsilon, t_2^* + \epsilon)$, which still satisfies the budget constraint. Then we note that

$$\begin{aligned} J(t_1^* - \epsilon, t_2^* + \epsilon) - J(t_1^*, t_2^*) &= \frac{1}{2}\Big(L(y_1 + t_1^* - \epsilon) - L(y_1 + t_1^*) + L(y_2 + t_2^* + \epsilon) - L(y_2 + t_2^*)\Big) \\ &= \frac{1}{2} \cdot \epsilon \cdot (L'(y_2 + t_2) - L'(y_1 + t_1)), \end{aligned}$$

where $t_1 \in [t_1^* - \epsilon, t_1^*], t_2 \in [t_2^*, t_2^* + \epsilon]$. We note that we must have that $J(t_1^* - \epsilon, t_2^* + \epsilon) - J(t_1^*, t_2^*) < 0$ because $L'(y_2 + t_2) < L'(y_1 + t_1)$ and $L' \leq 0$ for $t_1, t_2 \in [0, \Delta]$. This is a contradiction, so the optimal policy must provide the entire budget to x_2. Thus, the optimal policy under F_0 and $B = \Delta/2$ is not weakly equitable.

Weak Equity Violation under F_h Recall that $F_{h,Y|X} = K_h(y; y_i)$ for $i = 1, 2$. Define $J(t_1, t_2; h) := \frac{1}{2} \sum_{i \in \{1,2\}} \mathbb{E}_{F_h} \left[L(Y + t_i) \mid X = x_i \right]$. Note that $J(t_1, t_2; h) = \mathbb{E}_{F_h} \left[L(Y + t(X)) \right]$, where $t(x_i) = t_i$ for $i = 1, 2$. We consider solving

$$\min_{t_1, t_2 \in [0,\Delta]} \{ J(t_1, t_2; h) : t_1 + t_2 \leq \Delta \}. \tag{29}$$

Let $\Theta^*(h)$ denote the set of minimizers (t_1, t_2) of (29) at kernel bandwidth h. We show that there exists $h > 0$ such that $\Theta^*(h)$ is nonempty and all $(t_1, t_2) \in \Theta^*(h)$ violate weak equity for (B, B') where $B = \Delta/2$ and $B' = 0$. In other words, there exists $h > 0$ such that all $(t_1, t_2) \in \Theta^*(h)$ have the property that $t_2 > t_1$.

We establish that $\Theta^*(h)$ is nonempty. Since L is monotone, it has countably many discontinuities. The set of discontinuities of $L(t + y)$ have measure zero under $dK_h(\cdot; y_i)$. Since L is bounded, the dominated convergence theorem yields that J is continuous in t_1, t_2. A continuous function has a minimizer on a compact set, so $\Theta^*(h)$ is nonempty for $h > 0$.

We first establish that uniform convergence on the feasible set: we show that

$$\sup_{t_1,t_2\in[0,\Delta]} |J(t_1,t_2;h) - J(t_1,t_2)| \to 0. \tag{30}$$

Recall that L is continuous on O_i that contains $[y_i, y_i+\Delta]$. Let C_i be a closed neighborhood that contains $[y_i, y_i+\Delta]$. Note that L is uniformly continuous on C_i. This means that for $\epsilon > 0$, there exists $\delta > 0$ so that $|y - y_i| < \delta$ implies $|L(t_i + y) - L(t_i + y_i)| < \epsilon$ (for all $t_i \in [0, \Delta]$). As a result, we have that

$$\begin{aligned}
&\sup_{t_i\in[0,\Delta]} \left|\mathbb{E}_{F_h}\left[L(t_i+Y) \mid X = x_i\right] - L(t_i+y_i)\right| \\
&\leq \sup_{t_i\in[0,\Delta]} \int |L(t_i+y) - L(t_i+y_i)| dK_h(y;y_i) \\
&\leq \sup_{t_i\in[0,\Delta]} \int_{y:|y-y_i|<\delta} |L(t_i+y) - L(t_i+y_i)| dK_h(y;y_i) \\
&\quad + \sup_{t_i\in[0,\Delta]} \int_{y:|y-y_i|>\delta} |L(t_i+y) - L(t_i+y_i)| dK_h(y;y_i) \\
&\leq \sup_{t_i\in[0,\Delta]} \int_{y:|y-y_i|<\delta} |L(t_i+y) - L(t_i+y_i)| dK_h(y;y_i) \\
&\quad + \sup_{t_i\in[0,\Delta]} \int_{y:|y-y_i|>\delta} (|L(t_i+y)| + |L(t_i+y_i)|) dK_h(y;y_i) \\
&= \epsilon + 2||L||_\infty \cdot \mathbb{P}_{F_h}\left[|Y - y_i| > \delta \mid X = x_i\right].
\end{aligned}$$

We note that $\mathbb{P}_{F_h}\left[|Y - y_i| > \delta \mid X = x_i\right] \to 0$ because $Y \mid X = x_i \xrightarrow{p} y_i$. Since ϵ is arbitrary, we have that

$$\sup_{t_i\in[0,\Delta]} |\mathbb{E}_{F_h}\left[L(t_i+Y) \mid X = x_i\right] - L(t_i+y_i)| \to 0,$$

and (30) must hold.

Now, we prove the desired claim. Denote the minimizer of (28) by (t_1^*, t_2^*). Let $r = \Delta/4$. Consider a ball about the minimizer in the feasible set $B_r((t_1^*, t_2^*))$. Note that all points in this ball violate weak equity because $(t_1^*, t_2^*) = (0, \Delta)$ and perturbing each coordinate by at most $r = \Delta/4$ still yields a point (t_1, t_2) where $t_2 > t_1$. Let $\mathcal{T} = \{(t_1, t_2) \in [0, \Delta]^2 \mid t_1 + t_2 \leq \Delta\}$. i.e. $K := \mathcal{T} \setminus B_r((t_1^*, t_2^*))$. Note that K must be compact. Since (t_1^*, t_2^*) is the unique minimizer of (28), then there exists $\delta > 0$ such that

$$J(t_1, t_2; 0) > J(t_1^*, t_2^*; 0) + \delta$$

for $(t_1, t_2) \in K$. This follows from the fact that J is continuous on a compact set K so it must attain a minimum on K, and this minimum value must be strictly larger than the objective value attained by (t_1^*, t_2^*) by the uniqueness of the solution to (28). By uniform convergence, there exists $b > 0$ so that $|h| < b$ implies that

$$|J(t_1, t_2; h) - J(t_1, t_2; 0)| \leq \frac{\delta}{3} \tag{31}$$

for all $(t_1, t_2) \in \mathcal{T}$. For $|h| < b$ and $(t_1, t_2) \in K$, we have that

$$J(t_1, t_2; h) \geq J(t_1, t_2; 0) - \frac{\delta}{3} \tag{32}$$

$$\geq J(t_1^*, t_2^*; 0) + \frac{2\delta}{3}. \tag{33}$$

If $|h| \leq b$, we also have that

$$J(t_1^*, t_2^*; h) \leq J(t_1^*, t_2^*; 0) + \frac{\delta}{3}. \tag{34}$$

Combining (33) and (34) imply that

$$J(t_1, t_2; h) \geq J(t_1^*, t_2^*; h) + \frac{\delta}{3}.$$

This implies that any point in K incurs strictly higher objective value than (t_1^*, t_2^*) under objective function $J(\cdot; h)$ for h sufficiently small. In this case, any minimizer of (29) cannot lie in K and must instead lie in $B_r((t_1^*, t_2^*))$. This implies that there exists $h > 0$ such that any optimal policy that solves (29) for this h must violate weak equity for $(B, B') = (\Delta/2, 0)$.

B.2.2 Non-Continuously Differentiable Case

Now, we consider the case where L is not continuously differentiable in the region where it is non-convex. In this case, we will be able to show that it is possible to construct a smoothed loss function $\tilde{L}$ that preserves the non-convexity of L but is also continuously differentiable by using the Gamma kernel Q_b for smoothing. For this smoothed loss function $\tilde{L}$, we can construct a distribution F_h that satisfies Assumption 1 using the above proof technique under which the optimal policy that solves (2) with population distribution F_h and loss function $\tilde{L}$. Using the smoothing kernel Q_b and the distribution F_h, we can show that there is a distribution F under which the optimal policy that solves (2) with the original loss function L is not weakly equitable.

We prove the following technical lemma.

Lemma 12. *If $L : \mathbb{R}_+ \to \mathbb{R}$ is a non-convex and monotone decreasing function, then there exist continuity points t_0, t_1, t_2 of L such that $t_0 = \lambda t_1 + (1 - \lambda) t_2$ for some $\lambda \in (0, 1)$ and*

$$L(t_0) - \lambda L(t_1) - (1 - \lambda) L(t_2) > 0. \tag{35}$$

We apply Lemma 12 and obtain continuity points t_0, t_1, t_2 of L that yield (35). Define

$$\epsilon := L(t_0) - \lambda L(t_1) - (1 - \lambda) L(t_2).$$

We define the smoothed loss function.

$$\tilde{L}_b(t) := \int_0^\infty L(t + z) dQ_b(z),$$

where Q_b is the Gamma kernel. As $b \to 0$, the Gamma kernel converges in distribution to the Dirac measure at 0, i.e. $dQ_b \xrightarrow{d} \delta_0$. First, we verify that $\tilde{L}_b(t)$ is continuously differentiable on $\mathbb{R}_+$ for any $b > 0$ in the following lemma.

Lemma 13. *Let L be a bounded and Lebesgue integrable function on $\mathbb{R}_+$. Then, $\tilde{L}_b$ is continuously differentiable.*

Second, we show that the non-convexity exhibited by L is preserved by $\tilde{L}_b$. By weak convergence, $b \to 0$, $\tilde{L}_b(t) \to L(t)$ for continuity points t of L. This implies that there exists b sufficiently small so that $|\tilde{L}_b(t_1) - L(t_1)| < \frac{\epsilon}{4}$, $|\tilde{L}_b(t_2) - L(t_2)| < \frac{\epsilon}{4}$, and lastly $|\tilde{L}_b(t_0) - L(t_0)| < \frac{\epsilon}{4}$. This implies that

$$\tilde{L}_b(t_0) - \lambda \tilde{L}_b(t_1) - (1-\lambda)\tilde{L}_b(t_2) \geq \frac{\epsilon}{2}. \tag{36}$$

Thus, $\tilde{L}_b$ is nonconvex in a region where it is continuously differentiable.

Since $\tilde{L}_b$ is nonconvex in a region where it is continuously differentiable, we can apply the proof in Section B.2.1 to show that there is a distribution F_h under which the optimal policy under $\tilde{L}$ is not weakly equitable. This distribution F_h has covariate distribution $F_{h,X}(x) = \sum_{i \in \{1,2\}} \frac{1}{2}\mathbb{I}(x = x_i)$ and $F_{h,Y|X=x_i} = K_h(\cdot; y_i)$. Define a population distribution F over (X, Y) where $F_X = \sum_{i \in \{1,2\}} \frac{1}{2}\mathbb{I}(x = x_i)$ and $F_{Y|X=x_i}$ is the distribution over $Z_1 + Z_{2,i}$, where $Z_1 \sim Q_b(\cdot; 0)$ and $Z_{2,i} \sim K_h(\cdot; y_i)$. We note that $F_{Y|X=x_1} \preceq_{\text{SD}} F_{Y|X=x_2}$ in this case. We observe that

$$\begin{aligned}
\mathbb{E}_F\left[L(t+Y) \mid X = x_i\right] &= \mathbb{E}\left[L(t + Z_1 + Z_{2,i}) \mid X = x_i\right] \\
&= \int_0^\infty \int_0^\infty L(t + z_1 + z_{2,i}) dQ_b(z_1) \cdot dF_{h,Y|X}(z_{2,i}) \\
&= \int_0^\infty \tilde{L}(t + z_{2,i}) dF_{h,Y|X}(z_{2,i}) \\
&= \mathbb{E}_{F_{h,Y|X}}\left[\tilde{L}(t+Y) \mid X = x_i\right].
\end{aligned}$$

Thus, any optimal policy t^* that solves (2) with loss function $\tilde{L}$ and population distribution F_h is also the optimal policy that solves (2) with loss function L and population F described above. Since t^* is not weakly equitable under F_h, t^* is also not weakly equitable under F.

B.3 Proof of Theorem 2

B.3.1 Existence

To establish existence of a minimizer, we show the following. Suppose L, B satisfy conditions (a) or (b). If a policy t^* solves the pointwise optimization problem

$$\min_{t(x) \in \mathbb{R}} \left\{\mathbb{E}_F\left[L(t(x) + Y) \mid X = x\right] + \lambda \cdot t(x) : t(x) \geq 0\right\} \tag{37}$$

for every $x \in \mathcal{X}$ and a fixed $\lambda > 0$ and satisfies $\mathbb{E}_F\left[t^*(X)\right] = B$, then t^* must be a solution to (2). Then, we show that a policy that solves the pointwise optimization problem (37) for $\lambda > 0$ and satisfies the budget constraint exists, concluding the proof.

We show the first part holds. Let $t^*(X)$ be a policy that solves (37) and satisfies the budget constraint. Then for any policy t in the feasible set of (2), we have that

$$\mathbb{E}_F\left[L(t^*(x) + Y) \mid X = x\right] + \lambda \cdot t^*(x) \leq \mathbb{E}_F\left[L(t(x) + Y) \mid X = x\right] + \lambda \cdot t(x)$$

for all $x \in \mathcal{X}$. Taking expectations of both sides, we have that

$$\mathbb{E}_F\left[L(t^*(X) + Y)\right] + \lambda \mathbb{E}_F\left[t^*(X)\right] \leq \mathbb{E}_F\left[L(t(X) + Y)\right] + \lambda \mathbb{E}_F\left[t(X)\right].$$

We note that $\mathbb{E}_F\left[t^*(X)\right] = B$ by assumption and $\mathbb{E}_F\left[t(X)\right] \leq B$ by feasibility. As a result, we

have that

$$\mathbb{E}_F\left[L(t^*(X)+Y)\right]+\lambda\cdot B\leq \mathbb{E}_F\left[L(t(X)+Y)\right]+\lambda\mathbb{E}_F\left[t(X)\right]\leq \mathbb{E}_F\left[L(t(X)+Y)\right]+\lambda\cdot B.$$

Then, we have that $\mathbb{E}_F\left[L(t^*(X)+Y)\right]\leq \mathbb{E}_F\left[L(t(X)+Y)\right]$, so t^* must be a minimizer of (2).

Now, we show that a solution t^* that solves (37) and satisfies the budget constraint exist. We start by proving the existence of the solution to (37) for any $\lambda>0$. We note that (37) can be re-written as

$$\min_{t(x)\in\mathbb{R}}\{-G_x(t(x))+\lambda t(x): t(x)\geq 0\}. \tag{38}$$

Suppose that the first set of conditions holds, i.e., L is decreasing, strictly convex, differentiable, bounded below by a constant $C>-\infty$ and $B>0$. In this case, we can apply Lemma 10 to see that $G_x(t)$ is differentiable. As a result, the objective of (38) is continuous. In addition, the objective of (38) is also coercive for $\lambda>0$ because L is bounded below, so $-G_x(t(x))+\lambda t(x)\to\infty$ as $t(x)\to\infty$. Since a continuous, coercive function must have at least one global minimizer on $[0,\infty)$, then (38) has at least one global minimizer. Lemma 10 also shows that $G_x(t)$ is strictly concave, so the objective of (38) is strictly convex. This implies that (38) has at most one minimizer on $[0,\infty)$. The unique solution that solves (38) must satisfy the KKT conditions that $G_x'(t(x))=\lambda$ or take value at the boundary $t(x)=0$. We also note that G_x' is strictly decreasing so is invertible on its image, i.e. $(0,G_x'(0)]$. The solution must be given by

$$t_\lambda(x)=\begin{cases}(G_x')^{-1}(\lambda) & 0<\lambda<G_x'(0)\\ 0 & \lambda\geq G_x'(0).\end{cases} \tag{39}$$

Suppose that the second set of conditions holds, i.e. L is an FGT index with $\alpha\geq 1$ and $0<B<c$. In this case, we can apply Lemma 11 to see that $G_x(t)$ is differentiable on $(0,\infty)$. Since L is an FGT index, the solution to (38) is equivalent to the solution to

$$\min_{t(x)\in\mathbb{R}_+}\{-G_x(t(x))+\lambda t(x): 0\leq t(x)\leq c\} \tag{40}$$

because $L(z)=0$ for $z\geq c$. Since the objective of (40) is continuous and the feasible set is compact, (40) must have at least one minimizer on $[0,c]$. By Lemma 11, we also have that $G_x(t)$ is strictly concave on $[0,c]$ for every $x\in\mathcal{X}$. Thus, (40) has at most one solution on $[0,c]$, so the minimizer must be unique. The unique solution that solves (40) must satisfy the KKT conditions that $G_x'(t)=\lambda$ or take value on the boundary $t(x)=0$. We also note that G_x' is strictly decreasing on $[0,c]$ so is invertible on this region. Thus, the optimal solution to (38) is also given by (39).

We argue that there exists $\lambda\geq 0$ so that $\mathbb{E}_F\left[t_\lambda(X)\right]=B$. To show this, we note that $t_\lambda(x)$ is continuous in λ, as is its expectation. Note that $\mathbb{E}_F\left[t_\lambda(X)\right]\to 0$ as $\lambda\to\infty$: To see this, note that

$$-G_x(t_\lambda(x))+\lambda t_\lambda(x)\leq -G_x(0)$$

because $t_\lambda(x)$ is the minimizer of (38). In addition, because L is bounded below

$$C+\lambda t_\lambda(x)\leq -G_x(t_\lambda(x))+\lambda t_\lambda(x).$$

Combining the two inequalities yields

$$t_\lambda(x)\leq\frac{-C-G_x(0)}{\lambda}.$$

We note that $-G_x(0) \geq C$, so as $\lambda \to \infty$, we have that $t_\lambda(x) \to 0$. Taking expectations yields the desired result.

Under conditions (a), we can also show that as $\lambda \to 0$, we have that $t_\lambda(x) \to \infty$. To see this, we note that L is strictly convex, decreasing, and bounded below, so $-G_x'(t) \to 0$ as $t \to \infty$. This implies that $(G_x')^{-1}(\lambda) \to \infty$ as $\lambda \to 0$. Combined with the above result, we can use the Intermediate Value Theorem to show that there exists $\lambda \geq 0$ that yields $\mathbb{E}_F[t_\lambda(X)] = B$. Similarly, under conditions (b), we have that $(G_x')^{-1}(0) = c$. Again, we can use the Intermediate Value Theorem to show that there exists $\lambda \geq 0$ that yields $\mathbb{E}_F[t_\lambda(X)] = B$ for $0 < B < c$.

B.3.2 Uniqueness

Under conditions (a), the objective of (2) is guaranteed to be strictly convex in t. The feasible set is convex, so (2) must have at most one global minimizer.

Under conditions (b), (2) can be rewritten as

$$\min_{t:\mathcal{X}\to\mathbb{R}} \{\mathbb{E}_F[L(t(X)+Y)] : \mathbb{E}_F[t(X)] \leq B, 0 \leq t(x) \leq c\}$$

because L is an FGT index, and no optimal policy will allocate a transfer larger than c. Applying Lemma 11 yields that the objective of this optimization problem is strictly convex on the feasible set.

B.3.3 Weak Equity

The existence and uniqueness results imply that the optimal policy has the form given in (39) for some $\lambda \geq 0$. We use this structure to verify weak equity.

Incremental transfer is monotone decreasing in post-transfer consumption. First, we show that the simpler property in (6) holds. Second, we show that the optimal policy that is obtained by solving (2) with budget B is the same as the cumulative optimal policy that is obtained by summing the optimal policy $t(x; B')$ that solves (2) with a small budget B' and the policy $\Delta t(x; B - B')$ that solves (2) with the remaining budget $B - B'$ after administering the transfers $t(x; B')$. In other words,

$$t(x; B) = t(x; B') + \Delta t(x; B - B').$$

The first result implies that

$$F_{Y+t(X;B')|X=x} \preceq_{\text{SD}} F_{Y+t(X;B')|X=x'} \implies \Delta t(x; B - B') \geq \Delta t(x'; B - B').$$

The second result implies that

$$\Delta t(x; B - B') = t(x; B) - t(x; B')$$

for all $x \in \mathcal{X}$. Taken together, these two results imply that (5) holds.

Transfer is monotone decreasing in consumption. We show that $t_\lambda(x) \geq t_\lambda(x')$ for any $\lambda \geq 0$. Since the optimal policy has the form in (39) for some choice of λ, then the optimal policy must be monotone decreasing in $Y \mid X = x$.

We note that L' exists almost everywhere and is monotone increasing because L is decreasing and convex. As a result, if $Y \mid X = x \preceq_{\text{SD}} Y \mid X = x'$, then $\mathbb{E}_F[L'(Y+t) \mid X = x] \leq$

$\mathbb{E}_F\left[L'(Y+t) \mid X = x'\right]$. This implies that $G'_x(t) \geq G'_{x'}(t)$ for any $t \in \mathbb{R}_+$. Since G'_x is decreasing in t, then $(G'_x)^{-1}$ is also decreasing. Since $G'_x(t) \geq G'_{x'}(t)$ for all t and both are decreasing in t, then we also have that $(G'_x)^{-1}(\lambda) \geq (G'_{x'})^{-1}(\lambda)$. Thus, we have that $t_\lambda(x) \geq t_\lambda(x')$. Thus, the optimal solution to (2) is monotone decreasing in consumption.

Cumulative policy is equal to sum of incremental policies. It is straightforward to see that the optimal policy under the former is given by $t_{\lambda(B)}$, where t_λ is given by (39) with $\lambda(B)$ being the unique choice of λ that satisfies the budget constraint at B. The optimal policy under budget B' is similarly given by $t_{\lambda(B')}$, where $\lambda(B')$ being the unique choice of λ that satisfies the budget constraint at B'. Since $B > B'$, we have that $\lambda(B) \leq \lambda(B')$.

We now consider solving

$$\min_{t:\mathcal{X}\to\mathbb{R}_+} \{\mathbb{E}_F\left[L(t(X) + t_{\lambda(B')}(X) + Y)\right] : \mathbb{E}_F\left[t(X)\right] \leq B - B'\}. \tag{41}$$

Let

$$\tilde{G}_x(t) = G_x(t + t_{\lambda(B')}(x)),$$
$$\tilde{G}'_x(t) = G'_x(t + t_{\lambda(B')}(x)).$$

The optimal policy that solves (41) is given by

$$\tilde{t}_\lambda(x) = \begin{cases} (\tilde{G}'_x)^{-1}(\lambda) & 0 < \lambda < \tilde{G}'_x(0) \\ 0 & \lambda \geq \tilde{G}'_x(0) \end{cases}.$$

for some unique choice of λ that satisfies $\mathbb{E}_F\left[\tilde{t}_\lambda(X)\right] = B - B'$. We observe that

$$\tilde{t}_\lambda(x) = \begin{cases} (G'_x)^{-1}(\lambda) - t_{\lambda(B')}(x) & 0 < \lambda < G'_x(t_{\lambda(B')}(x)) \\ 0 & \lambda \geq G'_x(t_{\lambda(B')}(x)) \end{cases}. \tag{42}$$

We show that λ that satisfies $\mathbb{E}_F\left[\tilde{t}_\lambda(X)\right] = B - B'$ must be equal to $\lambda(B)$. Consider two cases: (i) $t_{\lambda(B')}(x) > 0$ and (ii) $t_{\lambda(B')}(x) = 0$.

1. In the first case, $t_{\lambda(B')}(x) > 0$. This implies that $G'_x(t_{\lambda(B')}(x)) = \lambda(B')$. Since $\lambda(B) \leq \lambda(B')$, then we have that $\tilde{t}_{\lambda(B)}(x) = (G'_x)^{-1}(\lambda(B)) - t_{\lambda(B')}(x)$.
2. In the second case, $t_{\lambda(B')}(x) = 0$. This implies $G'_x(t_{\lambda(B')}(x)) = G'_x(0)$. As a result,

$$\tilde{t}_{\lambda(B)}(x) = \begin{cases} (G'_x)^{-1}(\lambda) & 0 < \lambda < G'_x(0), \\ 0 & \lambda \geq G'_x(0). \end{cases}$$

Combining these two cases, we have that

$$t_{\lambda(B')}(x) + \tilde{t}_{\lambda(B)}(x) = t_{\lambda(B)}(x),$$

as desired. Taking expectations, we can confirm that $\mathbb{E}_F\left[\tilde{t}_{\lambda(B)}(X)\right] = B - B'$. We observe that the optimal policy that solves (41) is given by (42) with $\lambda = \lambda(B)$, and $t_{\lambda(B')}(x) + \tilde{t}_{\lambda(B)}(x) = t_{\lambda(B)}(x)$. So, we have shown that our desired claim holds.

B.4 Proof of Lemma 3

Minimizing the worst-case conditional poverty rate is equivalent to minimizing an upper bound on the worst-case poverty rate, i.e.

$$\max_{x\in\mathcal{X}} \mathbb{P}_F\left[Y + t(X) < c \mid X = x\right] = \min\{\lambda \in [0,1] \mid \lambda \geq \mathbb{P}_F\left[Y + t(X) < c \mid X = x\right] \quad \forall x \in \mathcal{X}\}.$$

As a result, we can rewrite (7) as the following optimization problem with an auxiliary parameter λ introduced as an upper bound on $\max_{x\in\mathcal{X}} \mathbb{P}_F\left[Y + t(X) < c \mid X = x\right]$, i.e.

$$\min_{t:\mathcal{X}\to\mathbb{R}_+} \min_{\lambda\in[0,1]} \Big\{\lambda : \mathbb{E}_F\left[t(X)\right] \leq B, \quad \mathbb{P}_F\left[Y + t(X) < c \mid X = x\right] \leq \lambda \quad \forall x \in \mathcal{X}\Big\}. \tag{43}$$

Note that the second constraint of (43) is separable. Combined with the nonnegativity constraint yields that for all $x \in \mathcal{X}$, a feasible policy t must satisfy

$$t(x) \geq (c - F^{-1}_{Y|X=x}(\lambda))_+.$$

To obtain an optimal solution to (43), it is sufficient to consider (λ, t_λ) pairs, where $t_\lambda(x) = (c - F^{-1}_{Y|X=x}(\lambda))_+$ because a feasible pair (λ, t), where $t(x) > (c - F^{-1}_{Y|X=x}(\lambda))_+$ yields the same objective value as (λ, t_λ), which is also feasible. Thus, we can reduce (43) to the following optimization problem

$$\min_{\lambda\geq 0} \Big\{\lambda : \mathbb{E}_F\left[(c - F^{-1}_{Y|X=x}(\lambda))_+\right] \leq B\Big\}.$$

Under Assumption 1, $\mathbb{E}_F\left[(c - F^{-1}_{Y|X=x}(\lambda))_+\right]$ is strictly decreasing in λ when $0 < \mathbb{E}_F\left[(c - F^{-1}_{Y|X=x}(\lambda))_+\right] < c$, so the minimum value of λ is obtained when the budget constraint is satisfied with equality. Recall that any optimal gap-targeting policy when $0 < B < c$ must have the form given in Lemma 4 and satisfy the budget constraint with equality. These forms are identical, so an optimal gap minimizing policy also minimizes the worst-case conditional poverty rate.

B.5 Proof of Lemma 4

We compute the optimal policy when $L(z) = (c - z)_+$. Since L is convex, the optimal policy solves

$$\min_{t:\mathcal{X}\to\mathbb{R}_+} \{\mathbb{E}_F\left[L(t(X) + Y)\right] + \lambda(\mathbb{E}_F\left[t(X)\right] - B)\}$$

for some $\lambda \geq 0$. The first-order stationarity condition for a given x is

$$\frac{d}{dt}\mathbb{E}_F\left[L(t + Y) \mid X = x\right] + \lambda = 0. \tag{44}$$

We note that

$$\begin{aligned}\frac{d}{dt}\mathbb{E}_F\left[L(t+Y)\mid X=x\right] &= \frac{d}{dt}\mathbb{E}_F\left[(c-t-Y)\mathbb{I}(Y<c-t)\right]\\ &= \frac{d}{dt}\int_0^{c-t}(c-t)f_{Y|X=x}(y)dy - \frac{d}{dt}\int_0^{c-t} y f_{Y|X=x}(y)dy\\ &= \frac{d}{dt}((c-t)\cdot F_{Y|X=x}(c-t)) - (c-t)\cdot f_{Y|X=x}(c-t)\\ &= -(c-t)\cdot f_{Y|X=x}(c-t) - F_{Y|X=x}(c-t) + (c-t)\cdot f_{Y|X=x}(c-t)\\ &= -F_{Y|X=x}(c-t).\end{aligned}$$

The optimal policy will either satisfy the KKT condition, simplified below (44),

$$F_{Y|X=x}(c-t) = \lambda$$

or be equal to a boundary point $t=0$. The optimal policy is given by

$$t_\lambda(x) = \begin{cases} c - F^{-1}_{Y|X=x}(\lambda) & 0<\lambda<F_{Y|X=x}(c)\\ 0 & \lambda \geq F_{Y|X=x}(c).\end{cases} \tag{45}$$

We note that the policy must also be feasible, so $t\geq 0$. So, the optimal policy has the form in (8).

B.6 Proof of Theorem 5

Under Assumption 2 and that the maximum transfer size is c, the feasible set can be restricted to a compact space. Under Assumptions 1, (2) with $L(z)=\mathbb{I}(z<c)$ has a continuous objective. Thus, there exists a minimizer of (2) with $L(z)=\mathbb{I}(z<c)$ when the transfers are restricted to lie between 0 and c.

We note that any minimizer of (2) must satisfy the KKT conditions because (2) satisfies linear constraint qualification. We derive the KKT conditions below. We define the Lagrangian and its derivative below.

$$L(t,\lambda,\nu,\mu) = \mathbb{P}_F\left[t(X)+Y<c\right] + \lambda(\mathbb{E}_F\left[t(X)\right]-B) - \sum_{x\in\mathcal{X}}\nu_x\cdot t(x) + \sum_{x\in\mathcal{X}}\mu_x\cdot(t(x)-c)$$

$$\nabla_{t(x)}L(t,\lambda,\nu,\mu) = f_X(x)\cdot(-f_{Y|X=x}(c-t(x))+\lambda) - \nu_x + \mu_x = 0.$$

Let $\lambda\geq 0, \nu_x\geq 0, \mu_x\geq 0$ for all $x\in\mathcal{X}$. The KKT conditions of this optimization problem are as follows

$$\begin{aligned} f_{Y|X=x}(c-t(x)) - \lambda = \frac{\mu_x-\nu_x}{f_X(x)} \quad &\text{(Stationarity)},\\ \mu_x\cdot(t(x)-c) = 0 \quad &\text{(Complementary Slackness)},\\ \nu_x\cdot t(x) = 0 \quad &\text{(Complementary Slackness)},\\ \lambda\cdot(\mathbb{E}_F\left[t(X)\right]-B) = 0 \quad &\text{(Complementary Slackness)},\\ t(x)\geq 0 \quad &\text{(Primal Feasibility)},\\ t(x)\leq c \quad &\text{(Primal Feasibility)},\\ \mathbb{E}_F\left[t(X)\right]\leq B \quad &\text{(Primal Feasibility)}.\end{aligned}$$

The optimal solution must satisfy the KKT conditions for some $\lambda \geq 0$.

Now we show that any policy t that satisfies the KKT conditions with $\lambda \geq 0$ is α-valid for some $\alpha \geq 0$. Note that any t that satisfies the KKT conditions will satisfy primal feasibility, i.e. $t(x) \in [0, c]$.

1. Suppose t satisfies the KKT conditions and has $t(x) \in \{0, c\}$ for some $x \in \mathcal{X}$. Then we note that $t(x) \in \mathcal{T}_\alpha(x)$ because $\mathcal{T}_\alpha(x)$ contains $\{0, c\}$ for all $\alpha \geq 0$.

2. Suppose t satisfies the KKT conditions and has $0 < t(x) < c$ for some $x \in \mathcal{X}$. Then $\nu_x = \mu_x = 0$ to ensure complementary slackness is satisfied. Then we must have that $f_{Y|X=x}(c - t(x)) = \lambda$. So, $t(x) \in \mathcal{T}_\alpha(x)$ for $\lambda = \alpha$.

As a result, a policy t that satisfies the KKT conditions for a given value λ will be α-valid for some choice of $\alpha \geq 0$.

B.7 Proof of Corollary 6

By Theorem 5, an optimal deterministic policy must be α-valid for some $\alpha \geq 0$. So, for each value α, we can compute the policy t^*_α. Then, the optimal policy can be obtained by computing $\operatorname{argmin}_{\alpha:\alpha\geq 0} \mathbb{P}_F \left[t^*_\alpha(X) + Y < c\right]$.

B.8 Proof of Corollary 7

Recall that if a policy t is deterministic and α-valid then $t(x) \in \mathcal{T}_\alpha(x)$ for every $x \in \mathcal{X}$, and we can write that $t(x) = \langle \pi(x), z(x;\alpha)\rangle$ where $\pi(x) : \mathcal{X} \to \{0,1\}^K$ and $\sum_{k\in[K]} \pi_k(x) = 1$. Since α-valid policies have a highly structured form, the objective of (10) can be written as $\mathbb{E}_F\left[\langle \pi(X), p(X;\alpha)\rangle\right]$ and the constraint of (10) can be written as $\mathbb{E}_F\left[\langle \pi(X), z(X;\alpha)\rangle\right] \leq B$. This yields (11).

B.9 Proof of Lemma 8

Let $t(y; B)$ is continuous in B and satisfies weak equity (Definition 4). Suppose for the sake of contradiction that there exists two units $y_1, y_2 \in \mathcal{Y}$ and budgets $0 \leq B' \leq B$ where rank preservation does not hold, i.e.

$$y_1 + t(y_1; B') \leq y_2 + t(y_2; B'),$$
$$y_1 + t(y_1; B) > y_2 + t(y_2; B).$$

Since $t(y; B)$ is continuous in B, $t(y_2; B) - t(y_1; B)$ is continuous in B. The above conditions imply that $t(y_2; B') - t(y_1; B') \geq y_1 - y_2$ and $t(y_2; B) - t(y_1; B) < y_1 - y_2$. By the Intermediate Value Theorem, there exists $B_0 \in [B', B]$ so that

$$t(y_2; B_0) - t(y_1; B_0) = y_1 - y_2. \tag{46}$$

At B_0, we have that $y_1 + t(y_1; B_0) = y_2 + t(y_2; B_0)$, so by weak equity we have that both

$$t(y_1; B) - t(y_1; B_0) \geq t(y_2; B) - t(y_2; B_0)$$
$$t(y_2; B) - t(y_2; B_0) \geq t(y_1; B) - t(y_1; B_0).$$

for all $B \geq B_0$.

This implies that

$$t(y_1; B) - t(y_1; B_0) = t(y_2; B) - t(y_2; B_0) \tag{47}$$

for all $B \geq B_0$. Applying both (46), (47) implies

$$\begin{aligned} y_1 + t(y_1; B) - y_2 - t(y_2; B) &= y_1 - y_2 + (t(y_1; B) - t(y_2; B)) \\ &= y_1 - y_2 + (t(y_1; B_0) - t(y_2; B_0)) \\ &= y_1 + t(y_1; B_0) - y_2 - t(y_2; B_0) \\ &= 0. \end{aligned}$$

This is a contradiction because we assumed that $y_1 + t(y_1; B) > y_2 + t(y_2; B)$. Thus, if t is continuous in B and satisfies weak equity, it must also weakly preserve ranks.

B.10 Proof of Lemma 9

To solve this problem, we consider optimizing over a simpler class of policies $\mathcal{M}$. Denote post-transfer consumption as $s(y; B) = y + t(y; B)$. Let $\mathcal{M}$ be the class of policies where $t(y; B)$ is monotone decreasing and $s(y; B)$ is monotone increasing for every $B > 0$. We characterize the optimal policy over $\mathcal{M}$. Then, we verify that $\mathcal{W} \subset \mathcal{M}$ and that the optimal policy family over $\mathcal{M}$ also satisfies Definition 4, so it lies in $\mathcal{W}$. Thus, the optimal policy family over $\mathcal{M}$ is also the optimal policy family over $\mathcal{W}$.

We consider

$$\min_{t(\cdot;\cdot):\mathcal{Y}\times\mathbb{R}_+\to\mathbb{R}_+} \{\mathbb{P}_F\left[Y + t(Y; B) < c\right] : \mathbb{E}_F\left[t(Y; B)\right] \leq B,\, t \in \mathcal{M}\}. \tag{48}$$

This is equivalent to solving the following optimization problem for each $B > 0$

$$\min_{t:\mathcal{Y}\to\mathbb{R}_+, s:\mathcal{Y}\to\mathbb{R}_+} \{\mathbb{P}_F\left[Y + t(Y) < c\right] : \mathbb{E}_F\left[t(Y)\right] \leq B,\, s(y) = y + t(y),\, s \text{ mon. inc.}, t \text{ is mon. dec.}\}.$$

Fix $B > 0$. We can further restrict to policies that only allocate transfers to units with $y < c$ because units with $y \geq c$ lie above the poverty line already. Let $\mathcal{A}$ be the set of pre-transfer consumptions of units who are lifted to the poverty line under a policy t such that $s(y) = y + t(y)$ is monotone increasing, t is monotone decreasing, and $t(y) = 0$ for $y \geq c$, i.e.

$$\mathcal{A} = \{y < c \mid y + t(y) \geq c\}.$$

Since s is monotone increasing and t does not allocate transfers to units with $y \geq c$, $\mathcal{A}$ must have the form (z, c). An optimal policy t^* must allocate transfers $t(y) = c - y$ for $y \in \mathcal{A}$ because this is the minimum transfer amount that raises these units to the poverty line.

Let $\lambda := \inf \mathcal{A}$. Since the optimal policy t^* must be monotone decreasing, $t^*(y) \geq c - \lambda$ for $y \leq \lambda$. Since allocating transfers to these units will not decrease the poverty rate, the optimal policy t^* will allocate the lowest possible transfers to these units provided that the feasibility constraints are satisfied. We note that the policy

$$t(y; \lambda) = \begin{cases} c - \lambda & y \leq \lambda \\ c - y & \lambda < y \leq c \\ 0 & y > c \end{cases}. \tag{49}$$

guarantees that $s(y) = y + t(y)$ is monotone increasing because

$$s(y) = \begin{cases} c - \lambda + y & y \leq \lambda \\ c & y > \lambda \end{cases}.$$

As a result, the optimal policy family over $\mathcal{M}$ must have the form (21).

We show that $\mathcal{W} \subseteq \mathcal{M}$. First, if $t \in \mathcal{W}$, then t must be monotone decreasing for every $B > 0$. This can be seen by setting $B' = 0$ into Definition 4. Second, if $t \in \mathcal{W}$, then it weakly preserves ranks (Lemma 8), so $y + t(y; B)$ must be monotone increasing for every $B > 0$.

Any policy that weakly preserves ranks has the property that $y + t(y; B)$ is monotone increasing for every $B > 0$. Thus, if $t \in \mathcal{W}$, then $t \in \mathcal{M}$.

We can check that (21) satisfies weak equity. Note that

$$y + t(y; B) = \begin{cases} y + c - \lambda(B) & 0 \leq y < \lambda(B) \\ c & \lambda(B) \leq y < c \\ y & y \geq c. \end{cases}. \tag{50}$$

First, we note that $\lambda(B)$ that defines the oracle policy in (21) is decreasing in budget. To see this, we observe that the cost of a policy with form (49) is decreasing in λ. Policy cost as a function of λ is

$$\mathbb{E}_F\left[t(Y; \lambda)\right] = \int_0^{\lambda} (c - \lambda) dF_Y(y) dy + \int_{\lambda}^{c} (c - y) dF_Y(y) dy.$$

Using Leibniz rule, we can check that

$$\begin{aligned} \frac{\partial}{\partial \lambda} \mathbb{E}_F\left[t(Y; \lambda)\right] &= (c - \lambda) f(\lambda) - F(\lambda) - (c - \lambda) f(\lambda) \\ &= -F(\lambda). \end{aligned}$$

Since F is a cumulative density function and is positive when $\lambda > 0$, then policy cost is strictly decreasing in λ. This implies that $\lambda(B)$ is decreasing in budget B.

Let $B > B'$. Suppose that

$$y_1 + t(y_1; B') \leq y_2 + t(y_2; B'). \tag{51}$$

We compute that

$$t(y; B) - t(y; B') = \begin{cases} 0 & y \geq \lambda(B') \\ \lambda(B') - y & \lambda(B) \leq y < \lambda(B') \\ \lambda(B') - \lambda(B) & 0 \leq y < \lambda(B). \end{cases}$$

Note that $t(y; B) - t(y; B')$ is decreasing in y.

1. Case (i): $y_1, y_2 \in [0, \lambda(B'))$. Then (51) implies that $y_1 \leq y_2$. Since $t(y; B) - t(y; B')$ is decreasing in y, so $t(y_1; B) - t(y_1; B') \geq t(y_2; B) - t(y_2; B')$.

2. Case (ii): $y_1, y_2 \geq \lambda(B')$. Then, $t(y_i; B') = (c - y_i)_+$. Since $\lambda(B) \leq \lambda(B')$, then we also have that $t(y_i; B) = (c - y_i)_+$, so $t(y_1; B) - t(y_1; B') = t(y_2; B) - t(y_2; B')$.

3. Case (iii): $y_1 \in [0, \lambda(B'))$ and $y_2 \geq \lambda(B')$. This implies that $y_1 \leq y_2$ Since $t(y; B) - t(y; B')$ is decreasing in y, we have $t(y_1; B) - t(y_1; B') \geq t(y_2; B) - t(y_2; B')$.

4. Case (iv): $y_2 \in [0, \lambda(B'))$ and $y_1 \geq \lambda(B')$. In this case, we have that $c \leq y_2 + c - \lambda(B')$, so $y_2 \geq \lambda(B')$. This is a contradiction because we assumed that $y_2 \in [0, \lambda(B'))$, so this case does not occur.

Thus, the policy family in (21) satisfies weak equity, and it must also be the optimal solution to (20).

C Technical Proofs

C.1 Proof of Lemma 10

C.1.1 Once Differentiable

We apply Leibniz rule to show that $G_x(t)$ is differentiable on $[0, \infty)$ and find an explicit form for the derivative.

Leibniz Rule [Klenke, 2013]: Let $\mathcal{T}$ be an open subset of $\mathbb{R}$ and let $\mathcal{Y}$ be a measure space. Suppose that $g : \mathcal{T} \times \mathcal{Y} \to \mathbb{R}$ satisfies the following conditions.

1. $g(t, y)$ be a Lebesgue integrable function of y for each $t \in \mathcal{T}$.

2. For almost all $y \in \mathcal{Y}$, the partial derivative $\frac{\partial}{\partial t} g(t, y)$ exists for all $t \in \mathcal{T}$.

3. There is a Lebesgue integrable function $\theta : \mathcal{Y} \to \mathbb{R}_+$ for $|\frac{\partial g}{\partial t}(t, \cdot)| \leq \theta(y)$ almost everywhere for all $t \in \mathcal{T}$.

Then, for all $t \in \mathcal{T}$,

$$\frac{d}{dt} \int_{\mathcal{Y}} g(t, y) dy = \int_{\mathcal{Y}} \frac{\partial}{\partial t} g(t, y) dy.$$

Let $\mathcal{T} = (0, \infty)$. Recall that $F_{Y|X=x}$ has positive density on $\mathcal{Y}$. We can set $g(t, y) := L(y + t) \cdot f_{Y|X=x}(y)$. Since L is a real-valued decreasing function defined on $\mathbb{R}_+$, we have that $\sup_{z \geq 0} L(z) = L(0)$. In addition, we also have that L is bounded below by $C > -\infty$. Thus, we have that $|L(y + t)| \leq \max(|L(0)|, |C|)$. We use this fact to show that g is Lebesgue integrable as follows

$$\begin{aligned}
\int_{\mathcal{Y}} |g(t, y)| dy &= \int_{\mathcal{Y}} |L(y + t) \cdot f_{Y|X=x}(y)| dy \\
&= \int_{\mathcal{Y}} |L(y + t)| \cdot f_{Y|X=x}(y) dy \\
&\leq \max(|L(0)|, |C|) \int_{\mathcal{Y}} f_{Y|X=x}(y) dy \\
&< \infty.
\end{aligned}$$

Since L is monotone and differentiable, for all $y \in \mathcal{Y}$, $\frac{\partial g}{\partial t}(t, y) = L'(y + t) \cdot f_{Y|X=x}(y)$ exists for all $t \in \mathcal{T}$.

Since L is convex, differentiable, and decreasing, $L'(z) \leq 0$ and L' is an increasing function. Thus, $|L'(y + t)| \leq |L'(0)| < \infty$ for all $t \in \mathcal{T}$. Let $\theta(y) := |L'(0)| \cdot f_{Y|X=x}(y)$. We have that

$\theta(y)$ is Lebesgue integrable because

$$\int_{\mathcal{Y}} \theta(y)dy = |L'(0)| \int_{\mathcal{Y}} f_{Y|X=x}(y)dy = |L'(0)| < \infty.$$

Thus, $G_x(t)$ satisfies the conditions of Leibniz rule, and

$$G'_x(t) = -\int_{\mathcal{Y}} L'(t+y) f_{Y|X=x}(y)dy \tag{52}$$

for $t \in \mathcal{T}$.

It remains to check that G_x is right-differentiable at $t = 0$. Since L is right-differentiable at $z = 0$, we can denote $L'(0)$ as the right derivative of L at $z = 0$. We check that G_x is right-differentiable at 0.

$$\lim_{h \to 0^+} \frac{G_x(h) - G_x(0)}{h} = -\lim_{h \to 0^+} \int_{\mathcal{Y}} \frac{L(y+h) - L(y)}{h} \cdot dF_{Y|X=x}(y)dy.$$

By the Mean Value Theorem, we have that

$$L(y+h) - L(y) = h \cdot L'(z(y))$$

for $z(y) \in [y, y+h]$. As a result, we have that

$$\begin{aligned} \left| \frac{L(y+h) - L(y)}{h} \cdot dF_{Y|X=x}(y) \right| &= |L'(z(y))| \cdot dF_{Y|X=x}(y) \\ &\leq |L'(0)|. \end{aligned}$$

The last line follows from the fact that L is convex and decreasing, so $L'(z) \leq 0$ and L' is increasing. Thus the dominated convergence theorem can be applied to obtain that

$$\begin{aligned} \lim_{h \to 0^+} \frac{G_x(h) - G_x(0)}{h} &= -\int \lim_{h \to 0^+} \frac{L(y+h) - L(y)}{h} \cdot dF_{Y|X=x}(y)dy \\ &= -\int L'(y) dF_{Y|X=x}(y)dy. \end{aligned}$$

Thus, G_x is right-differentiable at $t = 0$.

C.1.2 Strict Concavity

Since L is strictly convex, it follows that $G_x(t)$ is strictly concave.

C.2 Proof of Lemma 11

C.2.1 Once Differentiable

We first consider $\alpha > 1$. Note that in this case L is differentiable on $\mathbb{R}_+$, so it is straightforward to see that $G_x(t)$ is also differentiable on $\mathbb{R}_+$.

When $\alpha = 1$, L is not differentiable on $\mathbb{R}_+$. Nevertheless, we have that

$$G_x(t) = -\int_0^{c-t} (c - t - y) \cdot dF_{Y|X=x}(y).$$

We can apply integration by parts to see that

$$\begin{aligned} G_x(t) &= -(c-t-y)\cdot F_{Y|X=x}(y)\Big|_0^{c-t} + \int_0^{c-t} F_{Y|X=x}(y)dy \\ &= -\int_0^{c-t} F_{Y|X=x}(y)dy. \end{aligned}$$

By Assumption 1, we have that $F_{Y|X}$ is continuous. Then, we have that $G_x'(t) = F_{Y|X=x}(c-t)$.

C.2.2 Strict Concavity

We observe that the FGT indices for $\alpha \geq 1$ are convex. In addition, they are strictly convex at c in the sense defined below.

Definition 9 (Strict convexity at a point)**.** A convex function $f : \mathbb{R} \to \mathbb{R}$ is strictly convex at a point c if for all $z_1, z_2 \in \mathbb{R}$ such that $z_1 < c < z_2$ and $\lambda \in (0,1)$,

$$f(\lambda z_1 + (1-\lambda)z_2) < \lambda f(z_1) + (1-\lambda)f(z_2).$$

Let $t_1, t_2 \in [0, c]$ such that $t_1 < t_2$. We define $y' = c - \lambda t_1 - (1-\lambda)t_2$. We note that $y' + t_1 < c < y' + t_2$. Applying Definition 9, we have that

$$L(y' + \lambda t_1 + (1-\lambda)t_2) < \lambda L(y' + t_1) + (1-\lambda)L(y' + t_2). \tag{53}$$

We note that this result also holds in a small neighborhood about y'. Let $\epsilon = \frac{1}{2}\cdot\min(y' + t_2 - c, c - y' - t_1)$. In this case, $y' + \epsilon + t_1 < c < y' + \epsilon + t_2$ and $y' - \epsilon + t_1 < c < y' - \epsilon + t_2$, so (53) holds when y' is replaced by $y' + \epsilon$ and $y' - \epsilon$. Let $I = [y' - \epsilon, y' + \epsilon]$. We note that under Assumption 1, $f_{Y|X}(y) > 0$ on I.

$$\begin{aligned} &G_x(\lambda t_1 + (1-\lambda)t_2) \\ &= -\int_{\mathcal{Y}} L(y + \lambda t_1 + (1-\lambda)t_2) f_{Y|X=x}(y)dy \\ &= -\int_I L(y + \lambda t_1 + (1-\lambda)t_2) f_{Y|X=x}(y)dy - \int_{\mathcal{Y}\setminus I} L(y + \lambda t_1 + (1-\lambda)t_2) f_{Y|X=x}(y)dy \\ &> -\int_I [\lambda L(y + t_1) + (1-\lambda)L(y + t_2)] f_{Y|X=x}(y)dy - \int_{\mathcal{Y}\setminus I} L(y + \lambda t_1 + (1-\lambda)t_2) f_{Y|X=x}(y)dy \\ &\geq -\int_I [\lambda L(y + t_1) + (1-\lambda)L(y + t_2)] f_{Y|X=x}(y)dy - \int_{\mathcal{Y}\setminus I} [\lambda L(y + t_1) + (1-\lambda)L(y + t_2)] f_{Y|X=x}(y)dy \\ &= \lambda G_x(t_1) + (1-\lambda)G_x(t_2). \end{aligned}$$

Thus, G_x is strictly concave on $[0, c]$.

C.3 Proof of Lemma 12

Since L is nonconvex there exist points $x, y \in \mathbb{R}_+$ and $\lambda \in (0,1)$ such that

$$L(\lambda x + (1-\lambda)y) > \lambda L(x) + (1-\lambda)L(y).$$

Let $z = \lambda x + (1-\lambda)y$. We aim to show that this relationship holds at three continuity points of L, as well.

Let $\mathcal{E}$ be the set of discontinuity points of L, and let $\mathcal{C} := \mathbb{R}_+ \setminus \mathcal{E}$ denote the set of continuity points of L. Since L is monotone, we observe that $\mathcal{C}$ must be dense in $\mathbb{R}_+$. For the sake of contradiction, suppose not. Then there exists some $p \in \mathbb{R}_+$ and $r > 0$ such that the neighborhood $B_r(p)$ about p with radius r does not intersect $\mathcal{C}$, i.e. $B_r(p) \cap \mathcal{C} = \emptyset$. This implies that $B_r(p) \subset \mathcal{E}$. However, this is a contradiction because $B_r(p)$ is an open set that contains an uncountable number of points, and L is monotone, so $\mathcal{E}$ must be a countable set. Thus, we must have that $\mathcal{C}$ is dense in $\mathbb{R}_+$.

Since $\mathcal{C}$ is dense in $\mathbb{R}_+$, we can construct sequences $\{x_n\}, \{y_n\}, \{z_n\}$ that lie in $\mathcal{C}$ such that $x_n \downarrow x$, $y_n \downarrow y$, and $z_n \uparrow z$. Since L is monotone decreasing, we note that

$$\begin{aligned}
\lim_{n\to\infty} L(x_n) &= C_x \leq L(x), \\
\lim_{n\to\infty} L(y_n) &= C_y \leq L(y), \\
\lim_{n\to\infty} L(z_n) &= C_z \geq L(z).
\end{aligned}$$

Note that for N sufficiently large, we have that $x_n < z_n < y_n$. For $n > N$, we define

$$\lambda_n = \frac{z_n - y_n}{x_n - y_n}$$

and for $n \leq N$, we set $\lambda_n = \frac{1}{2}$. We note that $\lambda_n \in (0,1)$ and $\lambda_n \to \lambda$. Note that the affine relationship $z_n = \lambda_n x_n + (1-\lambda_n) y_n$ is satisfied exactly for $n > N$.

We observe that

$$\begin{aligned}
\lim_{n\to\infty} L(z_n) - \lambda_n L(x_n) - (1-\lambda_n) L(y_n) &= C_z - \lambda C_x - (1-\lambda) C_y \\
&\geq L(z) - \lambda L(x) - (1-\lambda) L(y) \\
&> 0.
\end{aligned}$$

Thus, for sufficiently large n, we have that $L(z_n) - \lambda_n L(x_n) - (1-\lambda_n) L(y_n) > 0$. Since x_n, y_n, z_n are continuity points of L, $\lambda_n \in (0,1)$, and $z_n = \lambda_n x_n + (1-\lambda_n) y_n$, this proves the desired claim.

C.4 Proof of Lemma 13

To show that $\tilde{L}_b$ is continuously differentiable, we show that differentiation and integration can be interchanged using Leibniz rule and obtain an expression for $\tilde{L}_b'$. After that, we apply the dominated convergence theorem to conclude that $\tilde{L}_b'$ is continuous.

We verify that Leibniz rule applies to $\tilde{L}_b$. First, we apply a change of variables

$$\tilde{L}_b(t) = \int_t^\infty L(u) dQ_b(u-t) du.$$

Let $\mathcal{T} = (0, \infty)$. We note that $Q_b(z)$ has a positive density on $[0, \infty)$. We can show that $L(u) \cdot dQ_b(u-t)$ is Lebesgue integrable because

$$\begin{aligned}
\int_t^\infty |L(u) dQ_b(u-t)| dz &\leq ||L||_\infty \int_t^\infty dQ_b(u-t) dz \\
&< \infty.
\end{aligned}$$

In addition, the partial derivative $\frac{\partial}{\partial t} L(u) dQ_b(u-t) = -L(u) dQ_b'(u-t)$, which exists for all $t \in \mathcal{T}$. We can show that there is a Lebesgue integrable function $\theta(u)$ so that

$$\left| \frac{\partial}{\partial t} L(u) dQ_b(u-t) \cdot \mathbb{I}(u \geq t) \right| \leq \theta(u)$$

almost everywhere. Define $\theta(u) = |L(u)| \cdot \frac{1}{b^2}$. For $z \geq 0$, we have that

$$\begin{aligned}
\left| \frac{\partial}{\partial t} L(u) dQ_b(u-t) \cdot \mathbb{I}(u \geq t) \right| &= |L(u) \cdot dQ_b'(u-t) \mathbb{I}(u \geq t)| \\
&\leq \left| L(u) \cdot \exp(-(u-t)/b) \cdot \Big(- \frac{(u-t)}{b^3} + \frac{1}{b^2} \Big) \right| \\
&\leq |L(u)| \cdot \frac{\exp(-(u-t)/b)}{b^2} \cdot \left| 1 - \frac{u-t}{b} \right| \\
&\leq \theta(u).
\end{aligned}$$

The last line follows because $e^{-r}|1-r| \leq 1$ for $r \in \mathbb{R}_+$. We have that θ is Lebesgue integrable because

$$\int_0^\infty \theta(u) \leq \int_0^\infty |L(u)| \cdot \frac{1}{b^2} du$$

and L is Lebesgue integrable. Thus, we can apply Leibniz rule to conclude that

$$\tilde{L}_b'(t) = - \int_t^\infty L(u) \cdot dQ_b'(u-t).$$

We additionally note that pointwise convergence of the integrand holds, i.e. $t_n \to t$ implies that $L(u) \cdot dQ_b'(u-t_n) \cdot \mathbb{I}(u \geq t_n) \to L(u) \cdot dQ_b'(u-t) \cdot \mathbb{I}(u \geq t)$. Then pointwise convergence and the existence of the Lebesgue integrable dominating function imply that $\tilde{L}_b'(t_n) \to \tilde{L}_b'(t)$ by the dominated convergence theorem. Thus, $\tilde{L}_b'(t)$ is also continuous.

D Data

D.1 Country Selection

This section details our criteria for including countries in the analysis, and any exceptions or caveats thereto. Generally speaking, we aim to include countries that account for as large a share of the world's poor as possible subject to various feasibility constraints. The specific criteria we impose are that countries must:

1. Have an extreme poverty rate greater than 10%, or account for more than 1% of the world's extreme poor. This filter aims to focus our search effort on the countries that most influence aggregate costs. Globally, 94.4% of poor households live in countries that meet this criterion.[45]

[45] Aggregate poverty shares stated in this section are the authors' own calculations based on poverty estimates for 2023, from the World Bank's Poverty and Inequality Platform [World Bank, 2025a, Castañeda and Clarke, 2026], accessed April 9, 2026. Country inclusion/exclusion decisions were based on estimates for 2022, from the World Poverty Clock [World Data Lab, 2022], accessed September 25 2025.

- We make an exception at this stage to include Côte d'Ivoire and Senegal. They are covered by the EHCVM survey series, which also covers six other West African countries already in our sample, and the surveys are harmonized in such a way that including them is unusually easy.

2. Have a recent, high-quality, nationally-representative LSMS survey or equivalent. This is a prerequisite for our analysis. Whenever possible we use the same survey as that on which estimates in the World Bank's Poverty and Inequality Platform are based. 94.4% of poor households live in countries that additionally meet this criterion.

3. Are estimated to have at least 1,000 extremely poor households covered by this survey, given the survey sample size and published poverty rate. This filter aims to ensure that the survey contains enough poor households for reliable statistical learning; relaxing it would require countries to conduct larger surveys (or surveys more focused in the regions containing the most poor households). 94.1% of poor households live in countries that additionally meet this criterion.

    - We make an exception at this stage to include Bangladesh (which we anticipated would contain slightly fewer than 1,000 extremely poor households) because of its size.

4. Have made the relevant survey data, including a measure of aggregate household consumption, publicly available, or have provided it upon request. 81.3% of poor households live in countries that additionally meet this criterion.

5. Have usable currency conversion factors available. Specifically, we remove Afghanistan and Somalia on the basis of advice from World Bank staff that reliable PPP conversion factors for the years in which the relevant surveys were conducted are not available, and we remove South Sudan because we were unable to find a CPI adjustment factor temporally aligned with the survey's currency unit.

This leaves a final sample of 34 countries which collectively account for 76% of the world's extreme poor.

D.2 Outcome Construction

We calculate consumption per capita by dividing total household consumption by household size without an adult equivalence scale adjustment, following the World Bank Poverty and Inequality Platform's methodology [World Bank, 2025b]. Wherever possible we use the consumption aggregates provided along with the disaggregated data for this purpose. Accounting for transfers, housing, and other durable goods are among the thornier issues in measuring consumption [Amendola and Vecchi, 2022], and the sources vary somewhat in the ways they do this (see Table D.1). That said, large majorities include the value of both cash and in-kind transfers; include an estimate of the value of housing services consumed; and include an estimate of the value of services from non-housing durable goods.

We convert consumption aggregates from all surveys from local currency units (LCU) to 2017 PPP USD using Consumer Price Index and PPP conversion factors. In some cases we have access to precise conversion factors used by the Poverty and Inequality Platform (PIP) team to make international poverty rate estimates. In others, including when we are using a survey not used by PIP, we use factors drawn from the World Bank's development indicators [World Bank, 2023]. This requires identifying the base year of the LCUs in which the consumption aggregate is defined. The base year is clear from survey context and/or documentation in most cases. In

cases where a survey spans two years and no further information about currency base year is available, we average conversion factors across the two years. Table D.1 describes properties of the consumption aggregates we use and indicates deflator usage.[46]

Table E.2 displays the resulting poverty headcount rate estimates (column titled "Survey" under column group "Poverty Rate") alongside poverty rate estimates for the same country and year from PIP. Where the PIP estimate was derived from the same underlying survey as our estimate, we include it in the column labeled "World Bank (same source);" where it was derived from a different survey (including by interpolating between estimates from different surveys in different years) we include it in the column labeled "World Bank (other source)." When we use the same source we are generally able to match the corresponding PIP estimate to within a percentage point (sometimes using non-public deflators or, in Nigeria, omitting a region where survey enumeration was disrupted). In four cases we do not: Burundi (4.7 percentage points higher), Central African Republic (1.6 percentage points lower), South Africa (2.9 percentage points lower), and in the Democratic Republic of the Congo (5.2 percentage points higher). When we use a different source the average absolute difference between the estimates is 2.4 percentage points and the average difference is 0.1 percentage points.

[46]We thank Daniel Mahler and Elizabeth Foster in particular for their generous guidance and assistance obtaining and interpreting the deflators and conversion rates required for a number of these operations.

Table D.1: Consumption Aggregate Details

Country	Category of Consumption			Deflator Usage
	Grants and Transfers	Durable Goods[a]	Non-Rental Housing[b]	
Bangladesh	Unspecified	Unspecified	Imputed; method unspecified	None used
Benin	Included	User cost	Hedonic function[c]	Temporal (WB)[d]
Burkina Faso	Included	User cost	Hedonic function	Temporal (WB)
Burundi	Unspecified	User cost	Self-reported rental value	Spatial
Central African Republic	Included	User cost	Imputed; method unspecified	None used
Colombia	Included	Acquisition approach	Imputed; method unspecified	None used
Côte d'Ivoire	Included	User cost	Hedonic function	Temporal (WB)
Dem. Rep. of the Congo	Included	User cost	Hedonic function	Spatial (urban/rural)
Ethiopia[e]	Included	Acquisition approach	Unspecified	Spatial
Ghana	Included	User cost	Hedonic function	None used
Guinea-Bissau	Included	User cost	Hedonic function	Temporal (WB)
India	Included	Acquisition approach[f]	All housing consumption excluded[g]	Spatial and temporal (WB)
Indonesia	Unspecified	Unspecified	Unspecified	None used
Kenya	Food included; non-food excluded	Unspecified	All housing consumption excluded[h]	None used
Liberia	Food included; non-food excluded	Unspecified	Self-reported rental value	None used
Madagascar	Food included; non-food excluded	User cost	Hedonic function	None used
Malawi	Food included; non-food excluded	User cost[i]	Self-reported rental value[j]	None used
Mali	Included	User cost	Hedonic function	Temporal (WB)
Mexico	Included	Acquisition approach	Imputed; method unspecified	None used
Namibia	Unspecified	Included; handling unspecified	Imputed; method unspecified	Spatial; temporal unspecified
Niger	Included	User cost	Hedonic function	Temporal, (WB)
Nigeria[k]	Food included; non-food unspecified	Acquisition approach[l]	Hedonic function	Spatial and temporal
Pakistan	Included	Included; handling unspecified	Self-reported rental value	Unspecified
Rwanda	Included	User cost	Self-reported rental value[m]	Temporal
Senegal	Included	User cost	Hedonic function	Temporal (WB)
Sierra Leone	Food included; non-food unspecified	User cost	Hedonic function	Spatial and temporal
South Africa	Included	Acquisition approach	Average rental yield[n]	Temporal
Sudan	Food included; non-food unspecified	Acquisition approach	Self-reported rental value	None used
Tanzania	Food included; non-food excluded	User cost	Hedonic function	Spatial and temporal
Timor-Leste	Food included; non-food unspecified	User cost	Hedonic function	None used
Togo	Included	User cost	Hedonic function	Temporal (WB)
Uganda	Included	Acquisition approach	Market prices (details unspecified)	None used
Yemen	Included	User cost	Self-reported[o]	Spatial
Zimbabwe	Food included; non-food excluded	User cost	Self-reported rental value	Spatial and temporal

This table lists characteristics of each consumption aggregate, including how each one handles often-tricky consumption categories.

[a] This column uses terminology introduced by Amendola and Vecchi [2022].

[b] Includes any housing not rented on the market: Owned housing, family- or employer-provided housing, etc.

[c] A function predicting rental value based on observable housing characteristics and amenities; fit using self-reported rents of respondents living in rental housing.

[d] Here "(WB)" indicates deflator data obtained from the World Bank which differs from, or is not included in, the corresponding survey dataset.

[e] Responses from Tigray Region are not included in survey microdata.

[f] Long-term ("lumpy") durables excluded.

[g] All (rental or non-rental) housing consumption is excluded, per WB methodology.

[h] All (rental or non-rental) housing consumption is excluded, per WB methodology.

[i] Assumes linear depreciation and disregards interest rate.

[j] Outliers imputed using hedonic function.

[k] Omitting responses from state of Borno per WB methodology.

[l] Long-term ("lumpy") durables excluded.

[m] Outliers imputed by means unspecified.

[n] Disaggregated by housing type and province.

[o] Counterfactual rent estimated using self-reported capital value of home

D.3 Predictor Selection

This section describes our rubric for selecting covariates from household surveys for inclusion as predictors when learning transfer policies. Our goal is to select only characteristics that are plausibly verifiable. To that end, we build a rubric from an initial list of covariates actually used in PMTs implemented in low- and middle-income countries, as described in Section 3. In particular, we define categories of covariates, generalizing where appropriate from individual covariates that have been used. For example, if previous PMTs used the material out of which the house's walls are made, then we take that as justification for also using the material out of which other parts of the house, such as the roof, are made. The resulting categories (labelled (a), (b), (c), etc. below) are exhaustive of the categories of variables within each top-level group (e.g., household demographics) that we include.

We also exercise judgment in excluding some variables that have been used in real-world PMTs but that we deem too difficult to plausibly verify at large scale. Examples include: Age of dwelling; indicators of food security, such as whether household members had skipped meals in the past week; use of fertilizer; use of an internet connection; income and private transfers (like foreign remittance); and ownership of a mobile phone or usage of someone else's mobile phone. We also exclude questions with unstructured free-entry text responses.

We use the resulting rubric as a guide when selecting predictors to use from each survey. The rubric, edited for clarity, is as follows:

1. Household demographics
    (a) Counts of household members by type, including the overall number of household members, the number of male and female members, the numbers in various age-based subcategories such as children, adults, elderly adults / senior citizens, and the number of specially-abled members. If the survey does not pre-define age-based categories then we will ourselves count the number of children 17 and under, and the number of older adults 65 and older. Examples of prior usage for the number of elderly adults: 60 and over for Below Poverty Line (BPL) classification in India [Planning Commission, 2011] or 65 and over [Alatas et al., 2012, Hanna and Olken, 2018].
    (b) Household head: Age, gender, and marital status. Examples of prior usage: Alatas et al. [2012] include married/unmarried; Fernandez and Hadiwidjaja [2018] include single/married/divorced; Kidd and Wylde [2011] include widow status.
2. Human capital
    (a) Educational attainment of the household head. Depending on the source this may take the form of years completed, highest academic level completed, or simply a literacy indicator. Examples of prior usage: Alatas et al. [2012], Planning Commission [2011, 2012].
    (b) Maximum educational attainment of all adult members. Again, this may be measured in years, in level completed, etc.[47] Example of prior usage: Kidd and Wylde [2011].
    (c) Maximum educational attainment of all female adult members. Example of prior usage: Kidd and Wylde [2011].
    (d) Number of children currently enrolled in school.

[47] We take the maximum rather than including the attainment of all members individually in order to produce a per-household predictor.

3. Household asset: presence and number owned

   (a) We do *not* include ownership-structure details beyond the fact of asset ownership.

   (b) Dwelling characteristics such as per capita number of rooms [Kidd and Wylde, 2011] or floor space [Alatas et al., 2012], material used to construct floors, walls, or roof, etc., as well as an indicator for homeownership itself. Examples of prior usage: Alatas et al. [2012] use binary indicators for home-ownership and roof materials. Hanna and Olken [2018] consider more granular categories.

   (c) Presence and physical characteristics of amenities such as: Type of latrine, water source, lighting source, drainage system, waste collection, access to electricity and gas, type of cooking fuel, cable connection. Examples of prior usage: Alatas et al. [2012] use a simple binary indicator for availability of clean water within the house. Hanna and Olken [2018] use more granular categories of water source.

   (d) Ownership of consumer durables such as appliances (e.g. radio, television, refrigerator, generator, cooker, heater, fan, air conditioner); transportation (e.g. car, bicycle, motorbike); furniture (e.g. sofa, bed etc); and devices (e.g. computers).

   (e) Productive agricultural assets including land, livestock, irrigation facilities, and farm machinery. We exclude cultivation details, such as crop types and amounts, as they may not be verifiable. We include the amount of land owned. Examples of prior usage: Planning Commission [2011, 2012] include farm machinery.

   (f) Productive non-agricultural assets, e.g a sewing machine.

   (g) We generally do *not* include financial assets, as we expect the kinds of financial assets held by poor households to be hard to verify, but we include any that they hold as part of a government scheme for which the government might plausibly hold records. Example of prior usage: Planning Commission [2011, 2012] include usage of the Kisan credit card scheme in India.

   (h) We *exclude* crop stores, as they may not be easy to verify.

4. Livelihood activities

   (a) Primary sector of employment of the household head (e.g., agriculture, manufacturing, services).

   (b) Primary occupation of the household head (e.g., self-employed, salaried employee, casual laborer, etc.).

   (c) We do *not* include the occupations of other members of the household.

   (d) Ownership of enterprises.

   (e) Receipt of other public transfers, including amount, if this could plausibly be verified by merging in other administrative records. We omit transfers when it is unclear whether they are administratively documented, such as child support or transfers from non-government institutions. Example of prior usage: Camacho and Conover [2011] describes a poverty census which includes social security information.

   Note. *Note: In cases where livelihood data is encoded using coding systems like ISIC, ISCO, etc., we truncate the codes to the first two digits, to control the eventual number of categories. In cases where this data is not recorded using a coding system admitting truncation, we limit to 50 categories including a catchall "other".*

5. Geographic indicators

(a) Urban/rural status, including any further available classifications such as peri-urban.

(b) Administrative geographic information: Which administrative division a household is in. Note that some care is required in the treatment of these indicators, as surveys may include region identifiers at a finer granularity than that at which they are representative. To obtain results which reflect expected performance across households anywhere in each country, we include geographic identifiers only at granularities at which the survey is representative. For example, if a survey were to sample households from all districts but only a subset of subdistricts, we would include district identifiers but not subdistrict identifiers as predictors.

(c) Distance to important locations such as district centers, markets, or public facilities such as post offices. Examples of prior usage: Alatas et al. [2012] use distance to district centers and markets. Hanna and Olken [2018] use distance to a post office.

(d) Environmental conditions such as rainfall history. Del Ninno and Mills [2015] provides a few case studies that illustrate feasibility.

6. Community characteristics

(a) Presence of publicly or privately provided services such as healthcare. Examples of prior usage: Alatas et al. [2012] uses an indicator for presence of a doctor. Kidd and Wylde [2011] uses an indicator for presence of a midwife.

(b) Presence of publicly or privately provided infrastructure such as paved roads, banking facilities, or regional government offices. Examples of prior work: Alatas et al. [2012], Kidd and Wylde [2011] use an indicator for presence of banking facilities. Kidd and Wylde [2011] use an indicator for presence of regional government offices.

(c) Population characteristics such as headcount or population density. Example of prior work: Kidd and Wylde [2011].

D.4 Secondary Data

This section compiles, for convenient reference, the full list of secondary data sources (i.e., those other than LSMS-style survey data) referenced elsewhere in the text.

- Country-level poverty data from the World Bank Poverty and Inequality Platform [World Bank, 2025a, Castañeda and Clarke, 2026]:
  - Poverty headcount rate and poverty gap index for each country for which we use a survey, from the year of that survey, interpolated by the authors if necessary.
  - Poverty headcount rate and poverty gap index for all countries in the world in 2023, when necessary using interpolated data provided in the dataset.
- Country-level poverty headcount data from the World Poverty Clock [World Data Lab, 2022].
- Country-level population estimates from the World Bank Development Indicators [World Bank, 2023].
- Effective currency exchange rates between local currency units (LCU) and US dollars. These rates are drawn from the World Bank's Development Indicators [World Bank, 2023]. We use the "DEC alternative conversion factor" series, designed to reflect actual effective exchange rates, on advice from World Bank staff, in particular to match local

currency units used in PPP conversion factors. Table E.5 documents how dollar costs for the affected countries would change if we instead used official exchange rates from the World Bank's Development Indicators [World Bank, 2023].

- Conversion factors between LCU and USD PPP (international dollars). These factors are drawn from multiple sources, depending on data availability and purpose:
  - The World Bank's Statistics Online portal [World Bank, 2025c].
  - The World Bank's Development Indicators [World Bank, 2023].
  - Private correspondence with the World Bank (Rwanda).
  - We fill some gaps in the World Bank dataset using data from the IMF[48] [IMF, 2025]: Taiwan (2017, 2021); Venezuela (2021); Yemen (2017, 2021).
  - Both World Bank and IMF datasets were missing Venezuela's 2017 conversion factor at the time of access. We estimate this value by assuming that the nominal currency exchange rate and the conversion factor from LCU to USD PPP (international dollars) have the same ratio as in the nearest year with complete data. By assuming that ratio, we are able to obtain an estimate for the conversion factor based on the nominal exchange rate for that year, which is present in IMF data.
- Country-level Consumer Price Index data from the World Bank's Statistics Online portal [World Bank, 2025c], the World Bank's Development Indicators [World Bank, 2023], and private correspondence with the World Bank (Rwanda) to adjust for inflation.
- Net official development assistance and official aid received, by country [World Bank, 2023].
- GDP by country and region [World Bank, 2023].
- Revenue as share of GDP by country [IMF, 2025] and by region [World Bank, 2023].

[48] Indicator name: Rate, Domestic currency per international dollar in PPP terms, ICP benchmarks 2017-2021

E Additional Exhibits

Table E.1: Survey data sources

Country	Survey Name	Citation	Survey Year
Bangladesh	HIES	Bangladesh Bureau of Statistics (BBS) [2022]	2022
Benin	EHCVM	Institut National de la Statistique et de l'Analyse Économique (INSAE) [2022]	2018
Burkina Faso	EHCVM	Institut National de la Statistique et de la Démographie (INSD) [2022]	2018
Burundi	EICVMB	Institut de Statistiques et d'Etudes Economiques du Burundi (ISTEEBU) [2020]	2020
Central African Republic	EHCVM	Institut Centrafricain des Statistiques et des Etudes Economiques et Sociales (ICASEES) [2021]	2021
Colombia	ENPH	Departamento Administrativo Nacional de Estadística (DANE) [2020]	2016
Congo - Kinshasa	EGI-ODD	Institut National de la Statistique (INS), DRC [2020]	2020
Côte d'Ivoire	EHCVM	Institut National de la Statistique (INS) [2022]	2018
Ethiopia	Socio-Economic Panel Survey	Ethiopian Statistical Service (ESS) [2024]	2021
Ghana	Living Standards Survey 7	Ghana Statistical Service (GSS) [2018]	2016
Guinea-Bissau	EHCVM	WAEMU Commission [2022a]	2018
India	Household Consumption Expenditure Survey	Ministry of Statistics and Programme Implementation (MoSPI) [2024]	2022
Indonesia	SUSENAS	Badan Pusat Statistik (BPS) [2018]	2018
Kenya	Continuous Household Survey	Kenya National Bureau of Statistics (KNBS) [2023]	2021
Liberia	HIES	Liberia Institute for Statistics and Geo-Information Services (LISGIS) [2024]	2014
Madagascar	Permanent Household Survey	Institut National de la Statistique (INSTAT), Madagascar [2021]	2021
Malawi	Fifth Integrated Household Survey	National Statistical Office (NSO) [2020]	2019
Mali	EHCVM	Institut National de la Statistique (INSTAT) [2022]	2018
Mexico	ENIGH	Instituto Nacional de Estadística y Geografía (INEGI) [2025]	2024
Namibia	Income and Expenditure Survey 2015-16	Namibia Statistics Agency (NSA) [2018]	2015
Niger	EHCVM	WAEMU Commission [2022b]	2018
Nigeria	Living Standards Survey	National Bureau of Statistics (NBS) [2021]	2018
Pakistan	HIES	Pakistan Bureau of Statistics (PBS) [2020]	2018
Rwanda	EICV7	National Institute of Statistics of Rwanda (NISR) [2025]	2023
Senegal	EHCVM	Agence Nationale de la Statistique et de la Démographie (ANSD) [2022]	2018
Sierra Leone	Integrated Household Survey 2018	Statistics Sierra Leone (Stats SL) [2018]	2018
South Africa	Income and Expenditure Survey	Statistics South Africa [2013]	2010
Sudan	NBHS	Central Bureau of Statistics (CBS), Sudan [2009]	2009
Tanzania	National Panel Survey, Wave 5	National Bureau of Statistics (NBS), Ministry of Finance and Planning [2023]	2020
Timor-Leste	Survey of Living Standards and Extension	National Statistics Directorate (NSD), Timor-Leste [2011]	2007
Togo	HSHLS	WAEMU Commission [2022c]	2018
Uganda	National Panel Survey	Uganda Bureau of Statistics (UBOS) [2021]	2019
Yemen	HBS	Central Statistical Organization (CSO), Yemen [2014]	2014
Zimbabwe	PICES	Zimbabwe National Statistics Agency (ZIMSTAT) [2017]	2017

We note that EHCVM is an abbreviation for "Enquête Harmonisée sur les Conditions de Vie des Ménages," HSHLS is an abbreviation for "Enquête Harmonisée sur les Conditions de Vie des Ménages" (Togo's WAEMU-harmonized survey, referenced in English as the Harmonized Survey on Households Living Standards), EICV7 is an abbreviation for "Enquête Intégrale sur les Conditions de Vie des Ménages," HIES is an abbreviation for "Household Income and Expenditure Survey," EICVMB is an abbreviation for "Enquête Intégrée sur les Conditions de Vie des Ménages au Burundi," EGI-ODD is an abbreviation for "Enquête par Grappes à Indicateurs des Objectifs de Développement Durable," ENPH is an abbreviation for "Encuesta Nacional de Presupuestos de los Hogares," ENIGH is an abbreviation for "Encuesta Nacional de Ingresos y Gastos de los Hogares," NBHS is an abbreviation for "National Baseline Household Survey," HBS is an abbreviation for "Household Budget Survey," PICES is an abbreviation for "Poverty, Income, Consumption and Expenditure Survey," and SUSENAS is an abbreviation for "Survei Sosial Ekonomi Nasional."

Table E.2: Survey characteristics

Country	n	d	Poverty Rate		
			World Bank (same source)	World Bank (other source)	Survey
Bangladesh	14268	193	3.6	-	3.6
Benin	8012	321	20.8	-	20.8
Burkina Faso	7010	379	31.2	-	31.2
Burundi	8358	82	62.1	-	66.8
Central African Republic	6411	157	65.7	-	64.1
Colombia	87201	109	-	4.9	2.8
Congo - Kinshasa	8993	64	78.9	-	84.1
Côte d'Ivoire	12992	366	11.5	-	11.5
Ethiopia	4959	148	-	32.0	35.6
Ghana	14009	327	25.2	-	25.2
Guinea-Bissau	5351	417	21.7	-	21.1
India	261696	40	2.3	-	2.3
Indonesia	295155	99	4.4	-	4.1
Kenya	16963	40	36.1	-	36.1
Liberia	4085	306	25.9	-	26.0
Madagascar	16540	171	52.3	-	52.3
Malawi	11434	152	70.1	-	70.1
Mali	6602	420	15.2	-	15.2
Mexico	91414	208	1.2	-	1.1
Namibia	10090	160	15.6	-	15.6
Niger	6024	215	50.9	-	50.9
Nigeria	21580	129	30.9	-	30.9
Pakistan	24809	134	4.9	-	4.9
Rwanda	15054	151	24.2	-	24.2
Senegal	7156	411	9.2	-	9.2
Sierra Leone	6810	177	26.1	-	26.1
South Africa	25328	102	18.0	-	15.1
Sudan	7913	58	18.6	-	19.1
Tanzania	4709	165	-	44.9	46.2
Timor-Leste	4477	208	40.8	-	40.8
Togo	6171	160	28.4	-	28.4
Uganda	3074	99	-	42.1	39.5
Yemen	9376	252	19.8	-	19.8
Zimbabwe	30155	222	34.2	-	34.3

This table describes the countries and surveys used in the empirical analysis. The column n is the number of households in the survey data. The column d is the number of covariates that we use from the survey for learning transfer policies. The "Poverty Rate" columns report the headcount poverty rates we estimate ("Survey") and those reported by the World Bank, with the latter differentiated into two cases: those in which the World Bank estimate is based on the same survey as ours, and those in which it is based on or extrapolated from other sources (World Bank's Poverty and Inequality Platform, accessed 14 July 2025).

Table E.3: Cost accounting (Togo)

	Survey	**Satellite**
Total program cost	**$1,523**	**$1,779**
Cost of transfers	$1,505	$1,777
Administrative costs	**$18**	**$3**
Screening	$16	$0
Training data	$3	$3

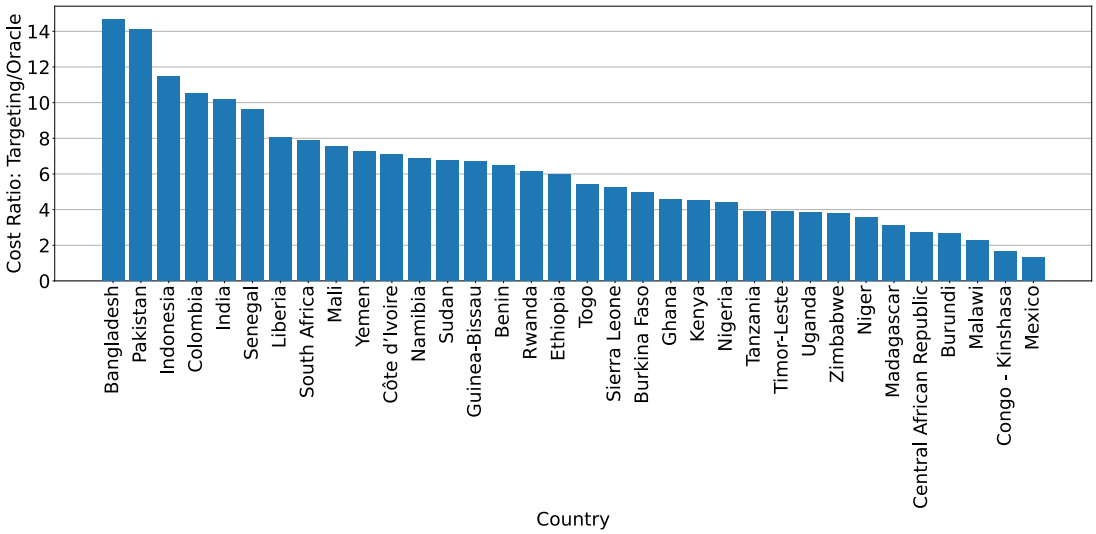
This table lists estimated costs, in millions of nominal 2023 USD, of two hypothetical year-long programs in Togo. One is targeted using survey data, and the other using satellite imagery.

Table E.4: Comparison of Policy Costs for In-Sample Countries to Country GDP and Government Revenue.

Country	Reference Year	Policy Cost	GDP	Gov't Revenue	Policy Cost / GDP	Policy Cost / Gov't Revenue
Bangladesh	2022	4.66 ± 0.52	479.07	42.72	0.01	0.11
Benin	2018	1.61 ± 0.17	17.31	2.35	0.09	0.68
Burkina Faso	2018	2.84 ± 0.15	19.28	3.81	0.15	0.74
Burundi	2020	2.78 ± 0.07	3.12	0.72	0.89	3.85
Central African Republic	2021	2.08 ± 0.07	2.83	0.39	0.73	5.38
Colombia	2016	1.53 ± 0.13	358.93	99.49	0.00	0.02
Congo - Kinshasa	2020	31.59 ± 0.33	57.36	5.37	0.55	5.88
Côte d'Ivoire	2018	1.87 ± 0.26	71.01	10.41	0.03	0.18
Ethiopia	2021	25.57 ± 1.78	125.11	13.80	0.20	1.85
Ghana	2016	4.89 ± 0.25	71.28	9.37	0.07	0.52
Guinea-Bissau	2018	0.21 ± 0.03	1.89	0.28	0.11	0.77
India	2022	15.80 ± 1.98	3483.84	701.39	0.00	0.02
Indonesia	2018	7.93 ± 0.24	1264.73	188.71	0.01	0.04
Kenya	2021	9.66 ± 0.46	123.36	20.75	0.08	0.47
Liberia	2014	0.98 ± 0.09	4.15	1.18	0.24	0.83
Madagascar	2021	4.94 ± 0.14	16.14	1.79	0.31	2.76
Malawi	2019	4.27 ± 0.13	13.17	1.94	0.32	2.19
Mali	2018	1.76 ± 0.22	24.72	3.85	0.07	0.46
Mexico	2024	0.22 ± 0.21	1799.64	443.51	0.00	0.00
Namibia	2015	0.44 ± 0.03	14.57	5.15	0.03	0.09
Niger	2018	4.72 ± 0.21	15.58	2.83	0.30	1.67
Nigeria	2018	29.50 ± 1.48	511.75	43.52	0.06	0.68
Pakistan	2018	5.63 ± 0.60	432.14	58.06	0.01	0.10
Rwanda	2023	1.68 ± 0.06	14.33	3.15	0.12	0.53
Senegal	2018	1.11 ± 0.19	28.05	5.30	0.04	0.21
Sierra Leone	2018	0.75 ± 0.05	7.75	0.78	0.10	0.95
South Africa	2010	9.13 ± 0.52	583.21	138.96	0.02	0.07
Sudan	2009	3.40 ± 0.25	73.32	10.97	0.05	0.31
Tanzania	2020	13.98 ± 1.30	77.78	11.58	0.18	1.21
Timor-Leste	2007	0.14 ± 0.01	0.80	0.37	0.18	0.39
Togo	2018	1.50 ± 0.10	8.53	1.55	0.18	0.97
Uganda	2019	6.69 ± 0.46	42.14	5.67	0.16	1.18
Yemen	2014	4.51 ± 0.32	55.64	13.16	0.08	0.34
Zimbabwe	2017	2.56 ± 0.08	63.49	11.11	0.04	0.23

This table reports the cost of the gap-minimizing policy that reduces the poverty rate to 1% in each country, and compares it to national GDP and government revenue. "Reference Year" corresponds to the year that the Policy Cost, GDP, and government revenue are reported and is taken to be the survey year. "GDP" corresponds to the country GDP in the survey year in units of billions of 2023 nominal USD. "Policy Cost" is the cost of the policy, in billions of 2023 nominal USD, to attain a 1% poverty rate in the country. We report 95% confidence intervals for the country-specific policy cost under a fixed transfer rule. For each country, we generate 50 bootstrap samples of the evaluation set, computing the cost to attain a 1% poverty rate on each bootstrap sample, and compute the standard error across the bootstrapped policy costs."Gov't Revenue" corresponds to country government revenue in the survey year in units of billions of 2023 nominal USD. We obtain the survey year GDP from the World Bank (data source, accessed 13 August 25). We obtain the survey year Gov't Revenue by multiplying survey year government revenue percentages from the IMF(data source, accessed 14 July 25) by the country GDP in the preceding column."Policy Cost / GDP" is the ratio of the policy cost and GDP. "Policy Cost / Gov't Revenue" is the ratio of policy cost and government revenue.

Figure E.1: Cost Ratio of Feasible Gap-Minimizing Policy to Aggregate Poverty Gap



This figure reports, for each country in Table E.2, the cost ratio between the gap-minimizing policy that achieves a 1% post-transfer poverty rate and the aggregate poverty gap.

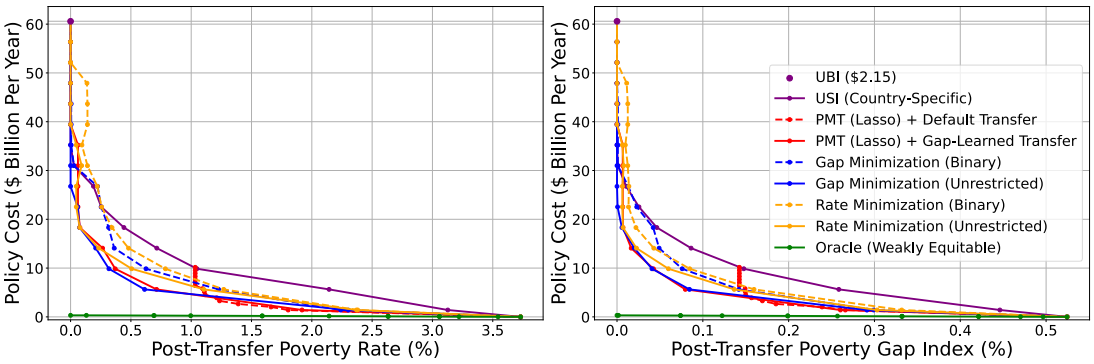
Table E.5: Cost ratios when using official exchange rates

Country	Cost Multiplier Using Official FX Rate
Sudan	3.14
Angola	1.66
Yemen, Rep.	1.32
Burundi	1.16
Syrian Arab Republic	1.04
Iran, Islamic Rep.	1.03
Nepal	1.02
Australia	1.02
Nauru	1.02
Haiti	1.01
India	0.99
Bangladesh	0.98
Uganda	0.98
Egypt, Arab Rep.	0.83
South Sudan	0.75

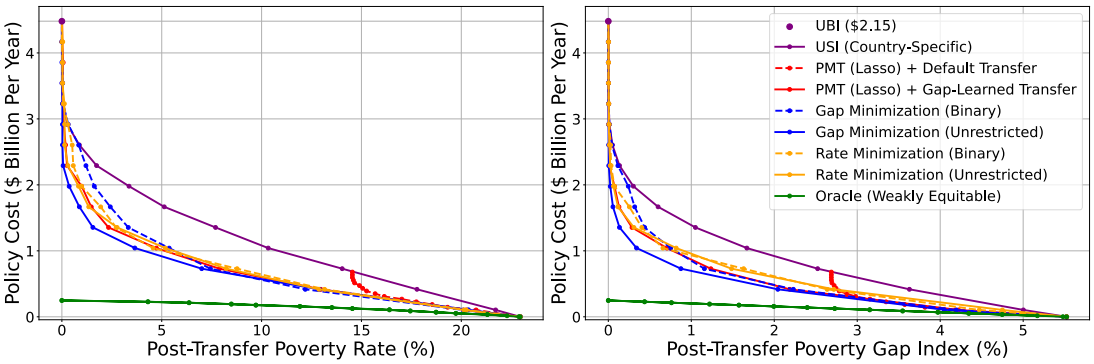
This table lists how much country-level costs would change using official, rather than effective, exchange rate estimates. Only countries with usable official exchange rate data available, and where switching rates changes the cost by at least 1%, are included.

Figure E.2: Policy Cost vs. Post-Transfer Poverty Measures by Country

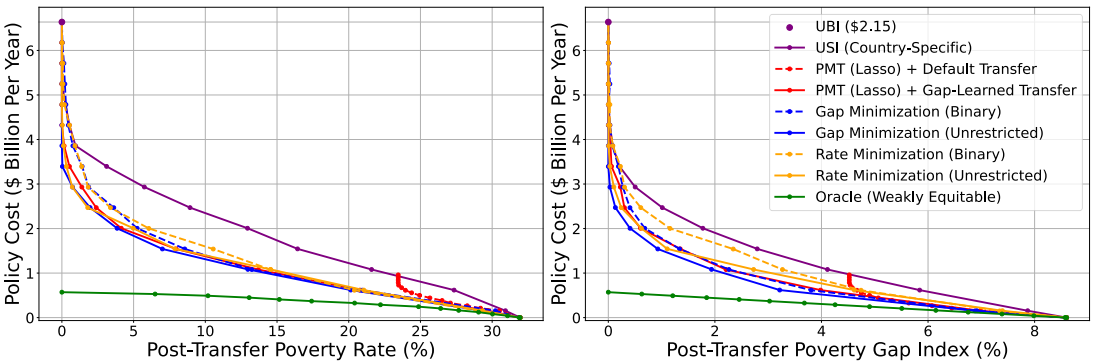
(1) Bangladesh



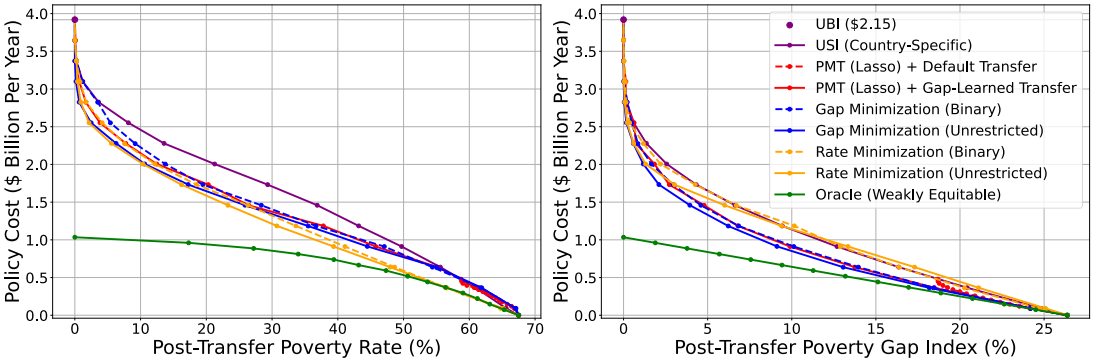
(2) Benin



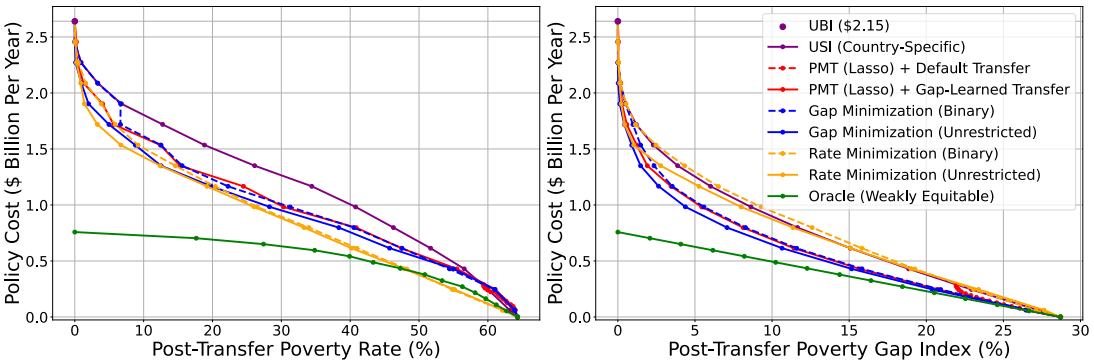
(3) Burkina Faso



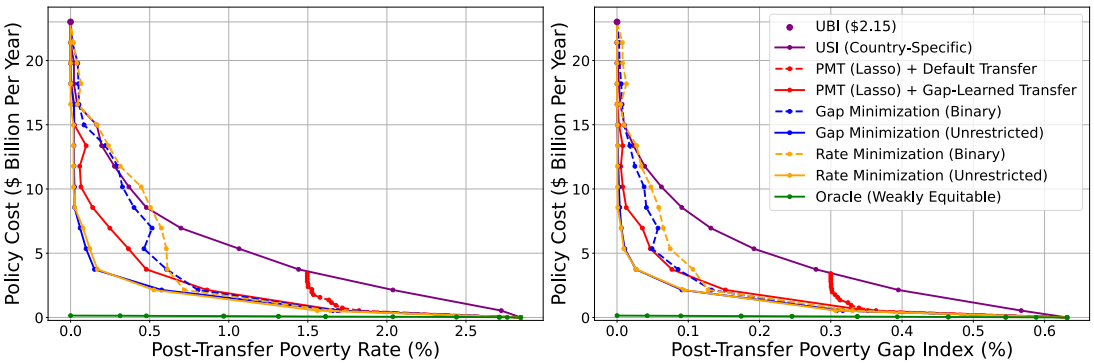
(4) Burundi



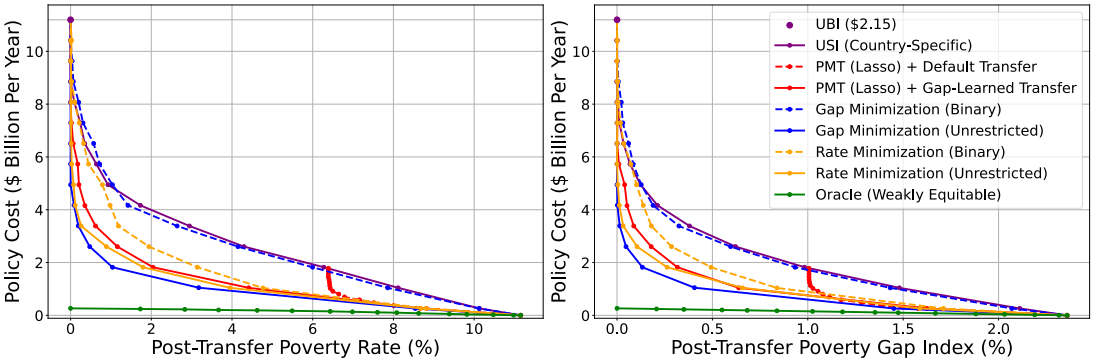
(5) Central African Republic



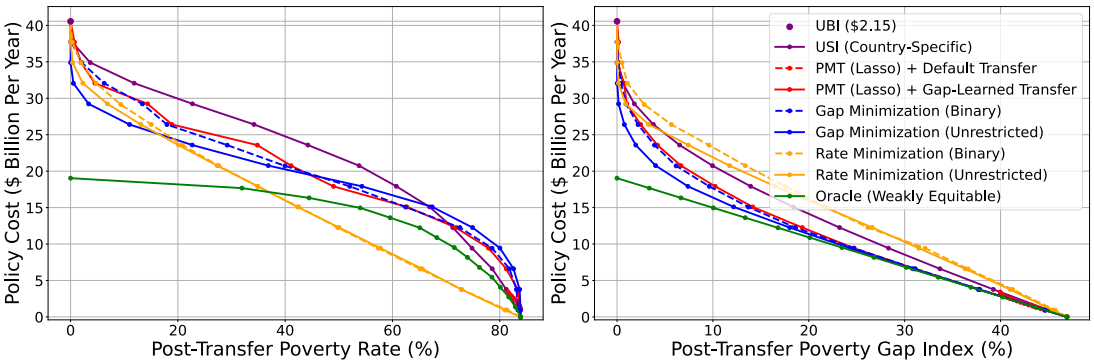
(6) Colombia



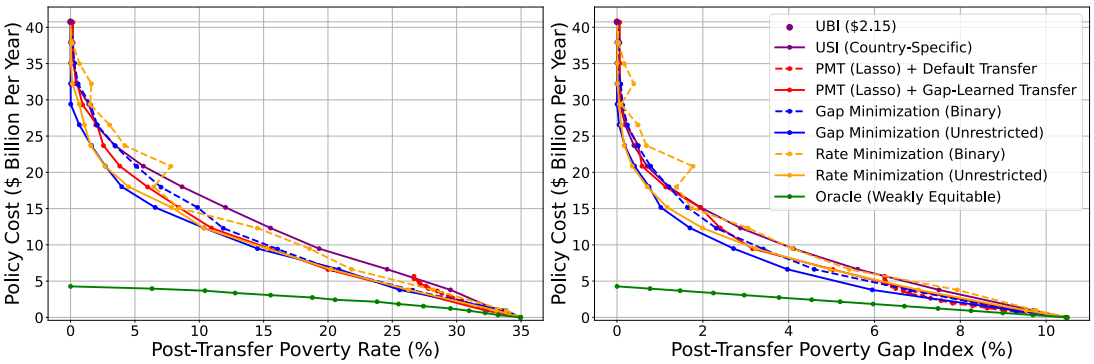
(7) Côte d'Ivoire



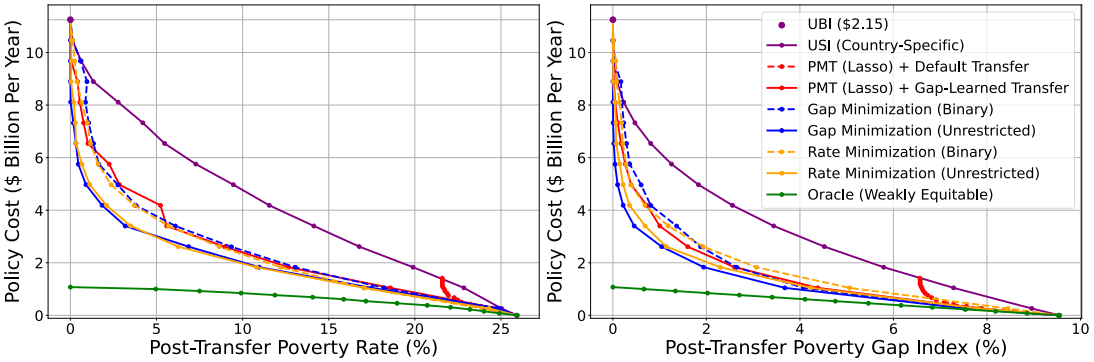
(8) Democratic Republic of Congo



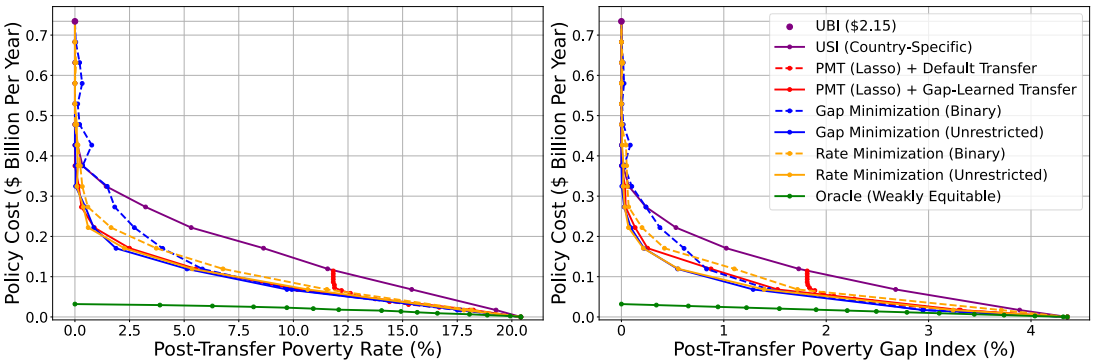
(9) Ethiopia



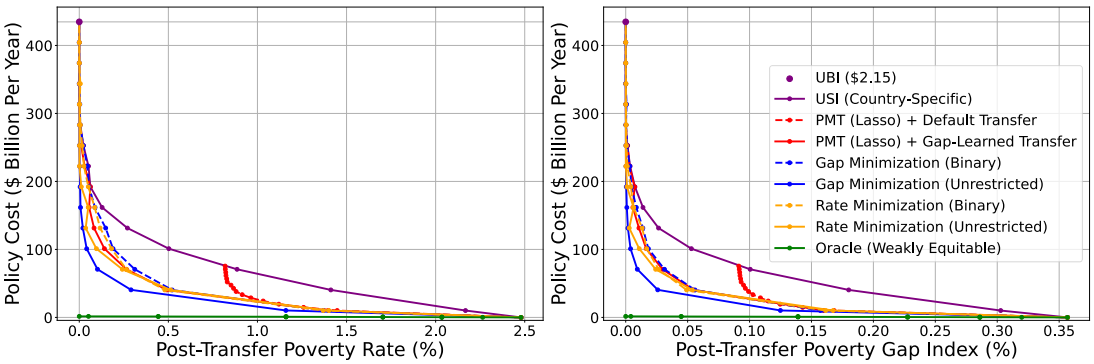
(10) Ghana



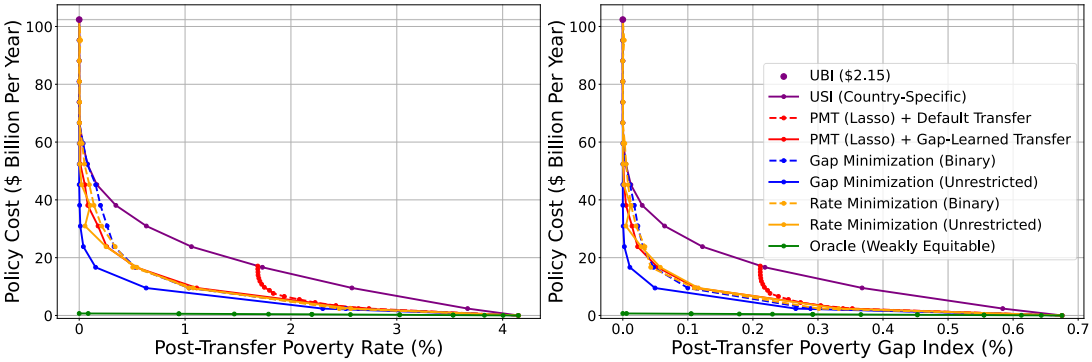
(11) Guinea-Bissau



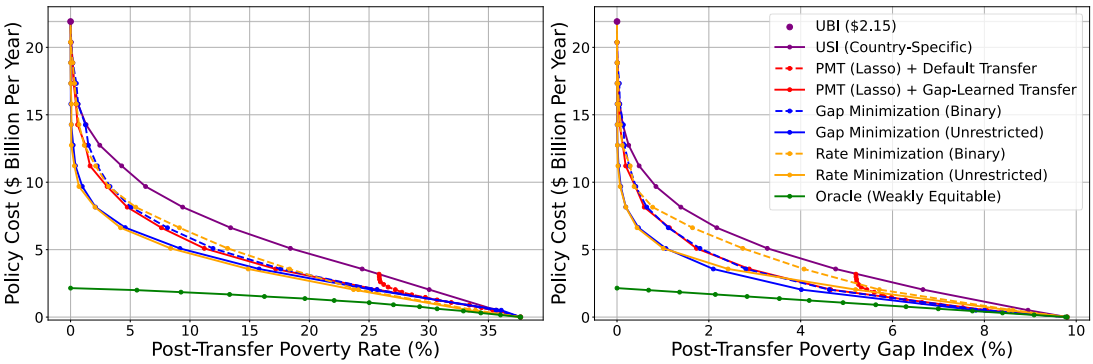
(12) India



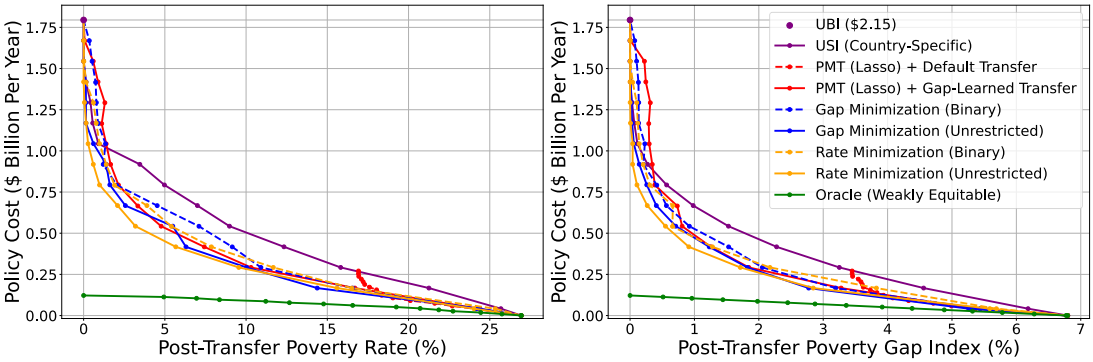
(13) Indonesia



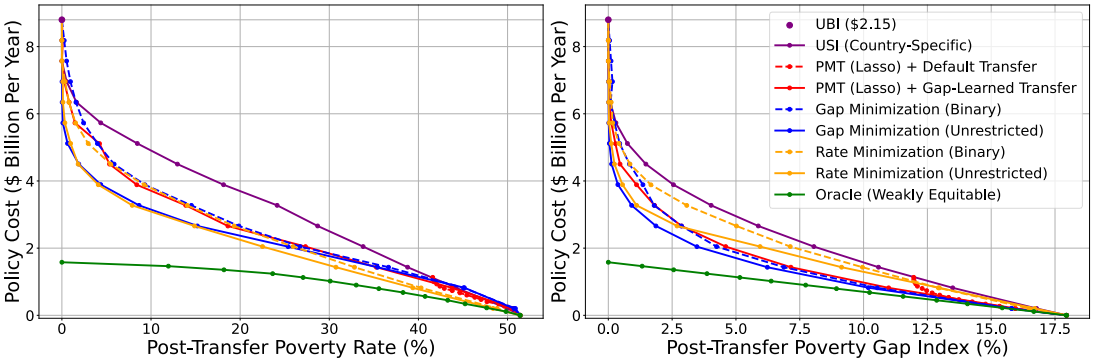
(14) Kenya



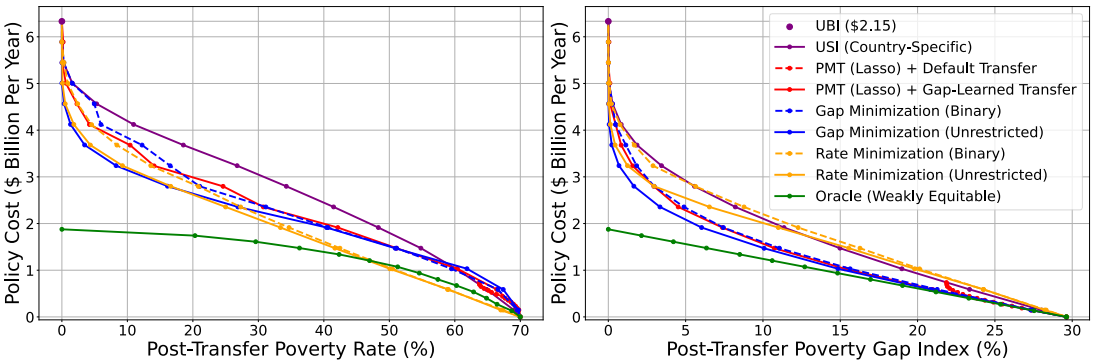
(15) Liberia



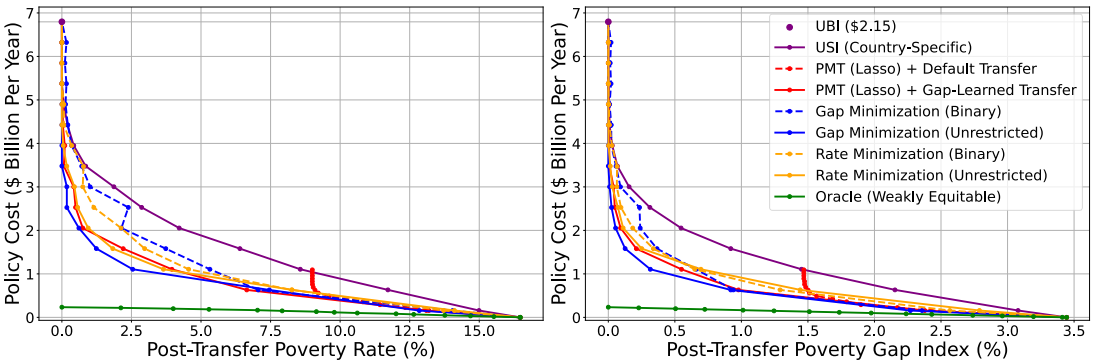
(16) Madagascar



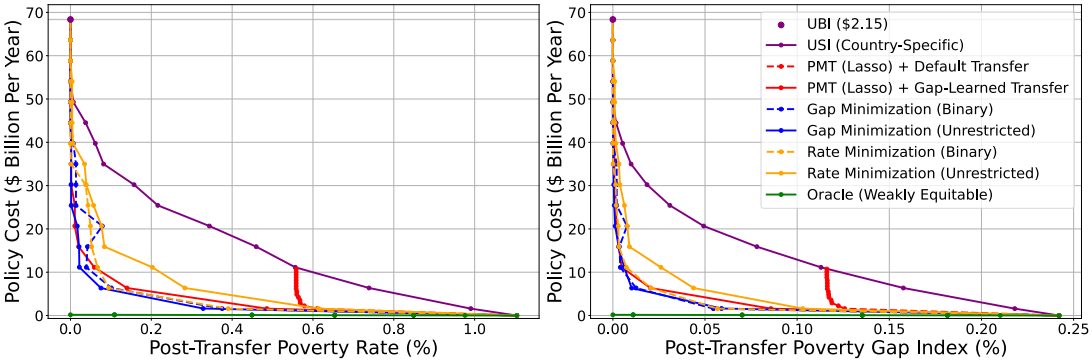
(17) Malawi



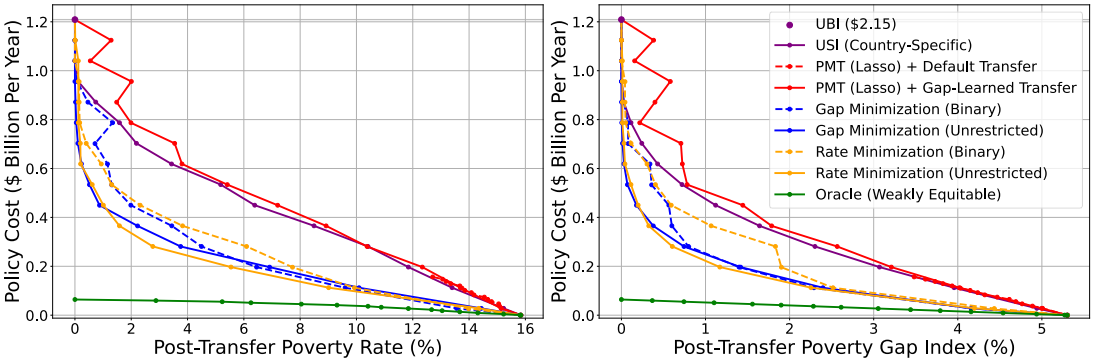
(18) Mali



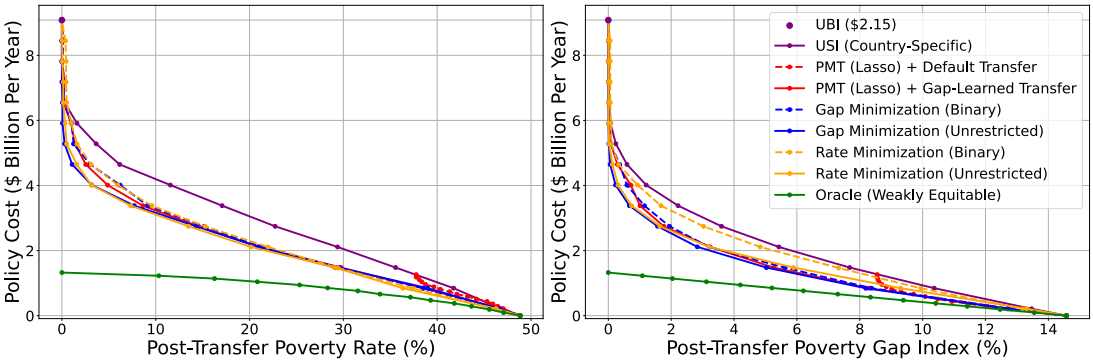
(19) Mexico



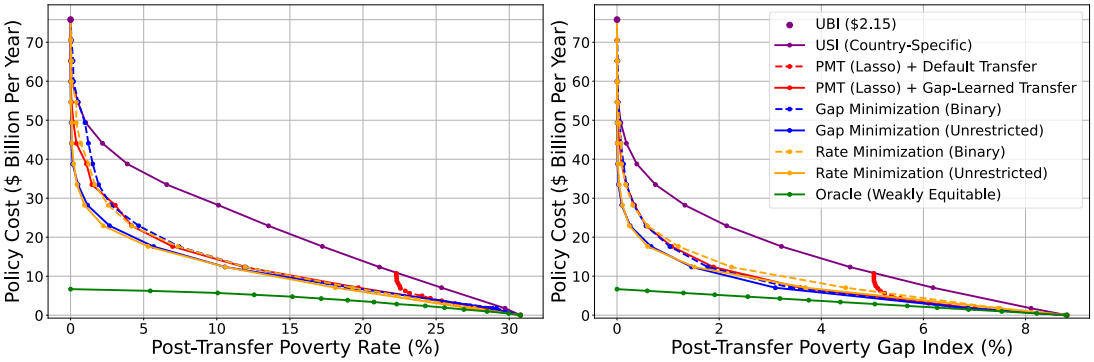
(20) Namibia



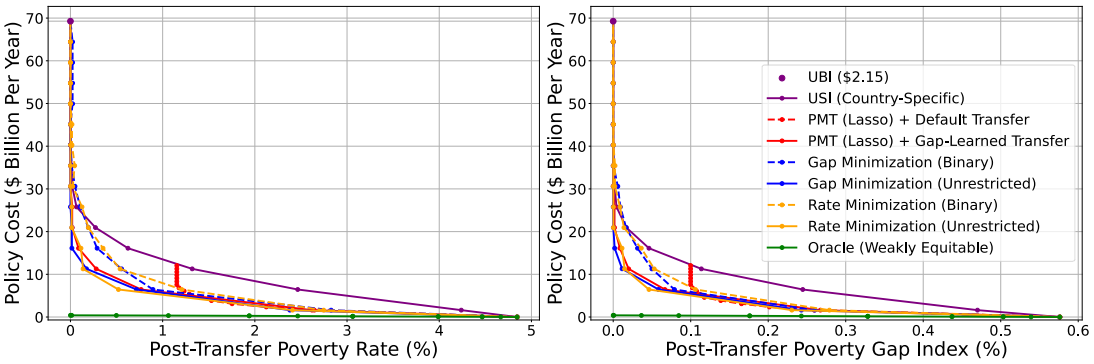
(21) Niger



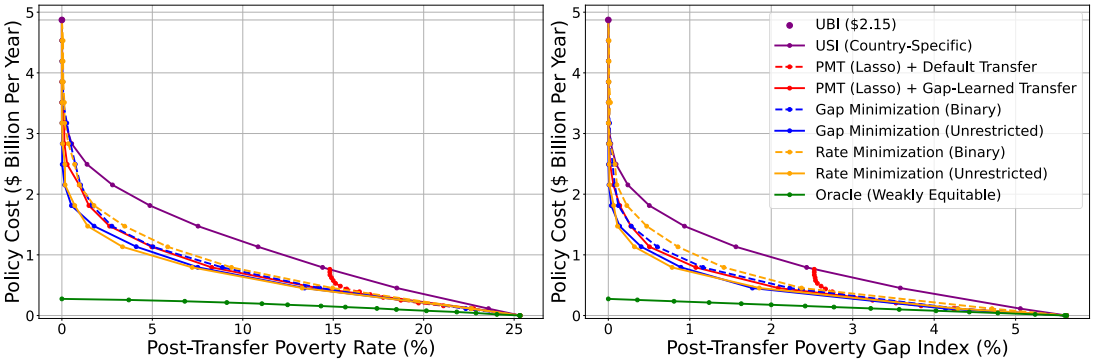
(22) Nigeria



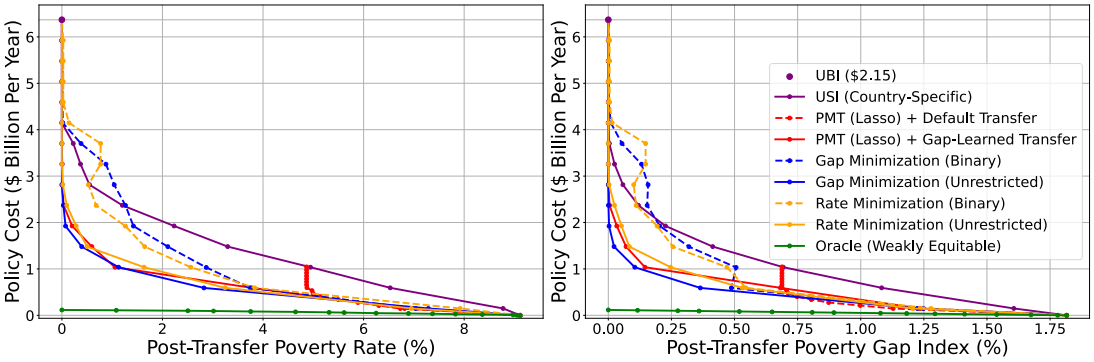
(23) Pakistan



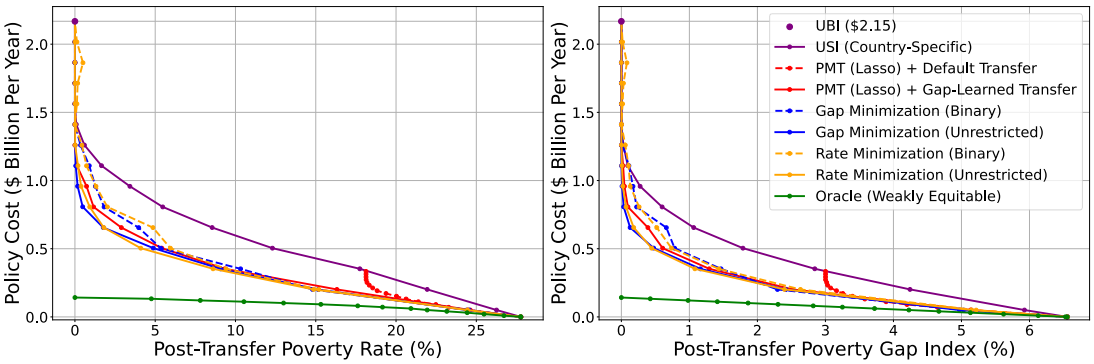
(24) Rwanda



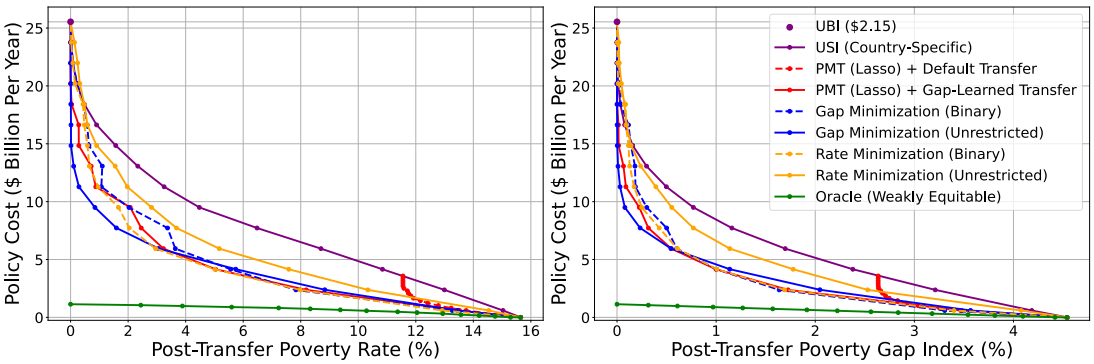
(25) Senegal



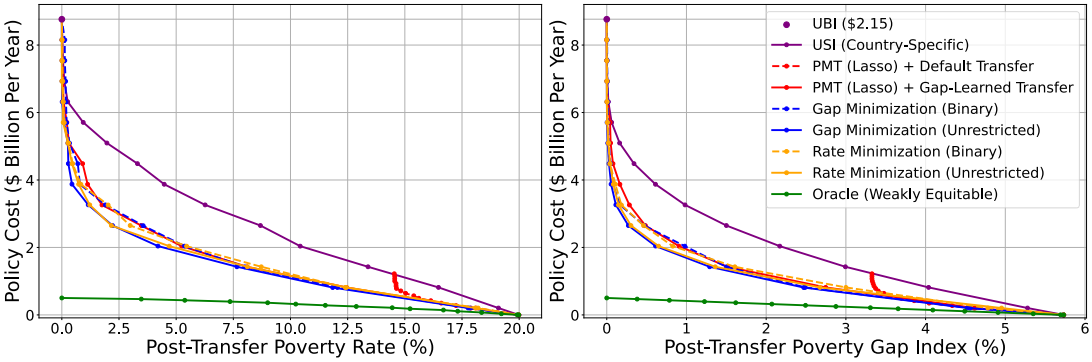
(26) Sierra Leone



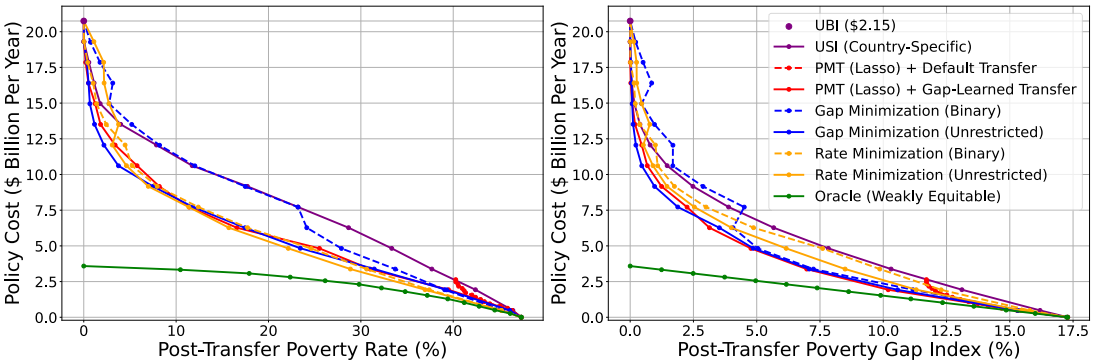
(27) South Africa



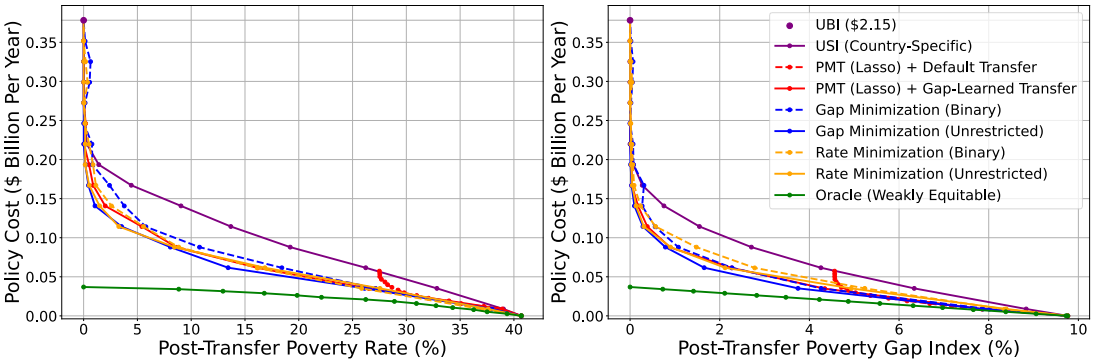
(28) Sudan



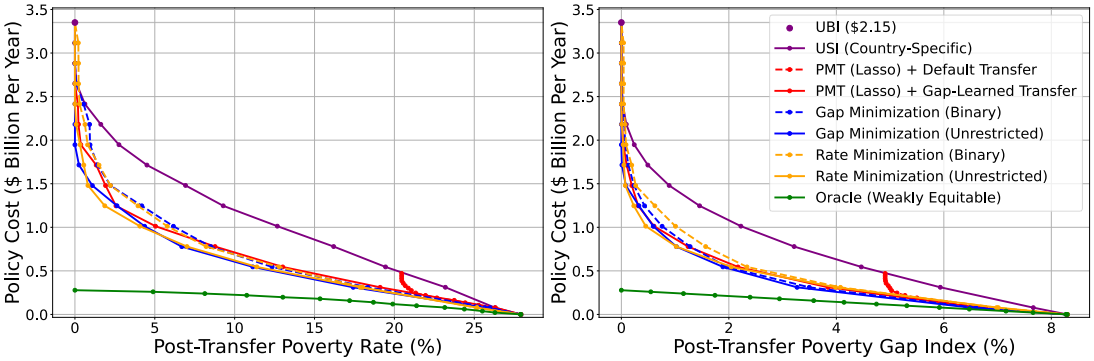
(29) Tanzania



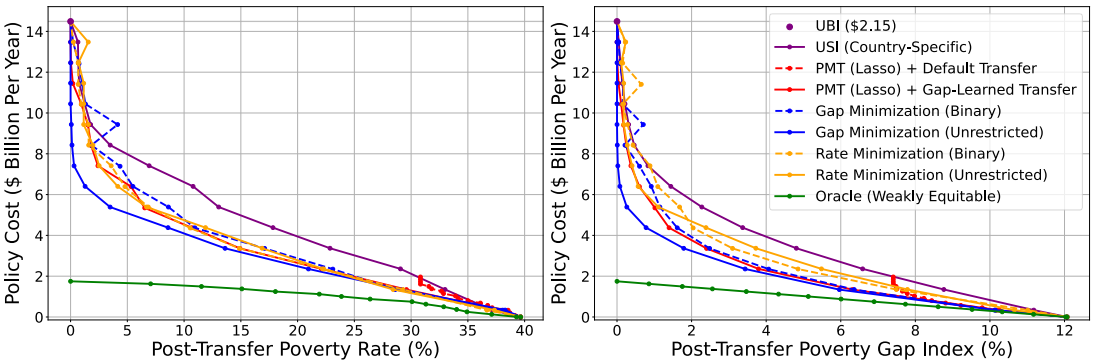
(30) Timor-Leste



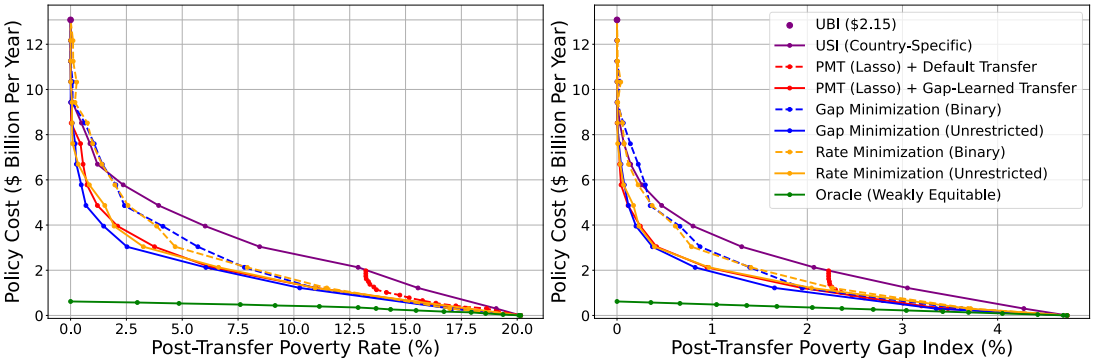
(31) Togo



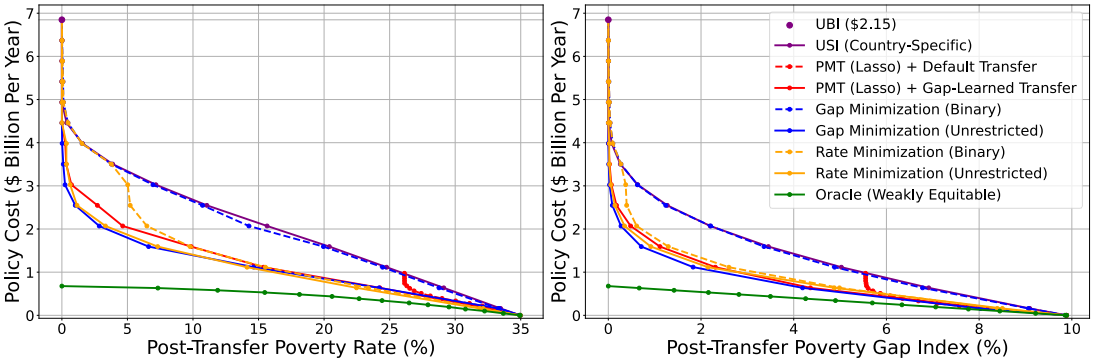
(32) Uganda



(33) Yemen

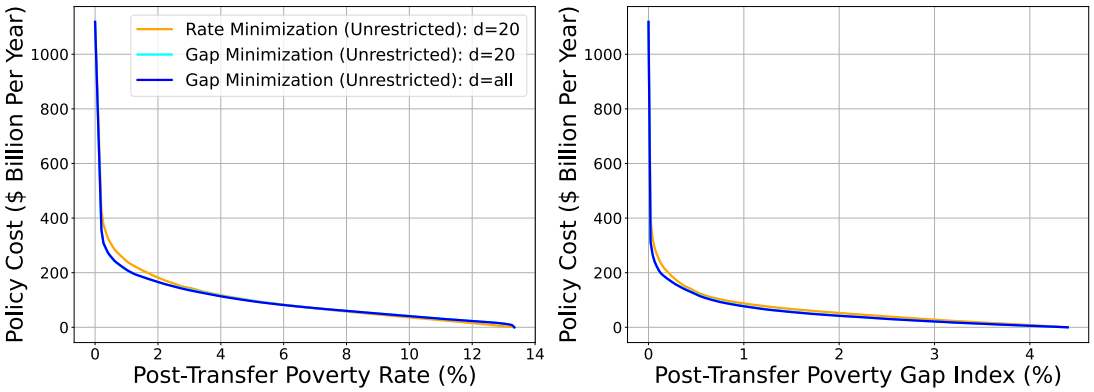


(34) Zimbabwe



This figure reports country-by-country results for gap minimization (binary/unrestricted), rate minimization (binary/unrestricted), PMTs learned using lasso and with default transfer sizes v.s. transfers learned to minimize the poverty objective, UBI at $2.15 per person per day, and Universal Supplemental Income.

Figure E.3: Comparison of Rate Minimization, Gap Minimization with Restricted Feature Set



This figure reports the costs, in billions of nominal 2023 USD annually, of the indicated optimal transfer policies learned for the full sample of countries. We compare the performance of rate minimization with the restricted feature set ($d = 20$), gap minimization with the restricted feature set ($d = 20$), and gap minimization with the full feature set. We find that the policy costs of gap minimization with the full and restricted feature sets coincide.