CSC495

Richard Guo

Contents

1 Paper Summaries

1.1 Understanding the Average Impact of Microcredit Expansions

This paper studies the causal effects of expanding access to microcredit using results from seven RCTs conducted in different countries. While individual studies often report different findings, it is unclear whether these differences reflect variation across contexts or are cause by noise. The paper addresses this by combining evidence across experiments instead of interpreting each study by itself.

Data

Each study is a separate RCT where households were randomly offered access to microcredit. The data include treatment assignment, loan take-up, and outcomes such as business profits and household consumption. Because many households offered credit do not actually borrow, the analysis focuses on intent-to-treat (ITT) effects, which measure the causal effect of being offered credit, not the causal effect of actually usingn it.

Rather than pooling all household-level data together, each study produces a single ITT estimate and standard error. These study-level estimates are then used in the main analysis.

Methods

The paper uses a Bayesian hierarchical model to combine results across studies. Each estimated effect is modeled as

$$\hat{\tau}_s \sim \mathcal{N}(\tau_s, \sigma_s^2),$$

where τ_s is the true treatment effect in study s. The study specific effects are assumed to come from a shared distribution,

$$\tau_s \sim \mathcal{N}(\mu, \sigma_\tau^2).$$

This setup makes it so that noisy estimates are shrunk toward the overall mean while still allowing for real heterogeneity.

Conclusion and Takeaway

The paper finds that the average effect of microcredit on outcomes like profits and consumption is small and close to zero. Much of the apparent variation across studies can be explained by sampling noise rather than true causal differences. A key takeaway is that learning a prior over environments can prevent overinterpreting noisy experimental results.

1.2 Experimental Design for Policy Choice

This paper approaches causal inference from a policy perspective. Instead of focusing on estimating treatment effects as precisely as possible, it asks how experiments should be designed so that the resulting causal evidence leads to better policy decisions.

Causal Setup

What the outcomes will depend on is treatment and covariates; the causal parameters remain unknown. If known, the policymaker would have the ability to compute the counterfactual outcomes and then select the optimal policy. These causal effects could then be obtained through randomized experiments; however, the paper highlights that not all experiments will offer the same information.

Methods

The main contribution is framing experimental design as a decision problem. The experimenter chooses how treatment is assigned, which affects the variance of causal estimators:

$$\hat{\theta} \approx \mathcal{N}(\theta, V(\delta)),$$

where δ represents the experimental design. The welfare objective is modeled as a quadratic function of θ, which allows the experimenter to optimize the design based not on the error of estimation but on the expectation of the welfare loss.

Conclusion and Takeaway

The paper shows that experiments optimized for policy choice can look very different from standard designs focused on average treatment effects. It may be better to learn certain causal effects well, even if others are estimated poorly.

2 Datasets

Dataset	Domain	Design	Treatment (T)	Outcomes (Y)	Access / Notes
Moving to Opportunity (MTO)	Housing / Labor (US)	RCT	Housing voucher offer	Earnings, education, neighborhood exposure (long-run)	HUD + replication packages: `https://www.huduser.gov/portal/datasets/mto.html`
Oregon Health Insurance Experiment	Health (US)	Lottery RCT (encouragement)	Medicaid lottery selection	Health utilization, financial strain, health metrics	Study site + data docs: `https://www.nber.org/research/data/oregon-health-insurance-experiment-data`
RAND Health Insurance Experiment	Health (US)	RCT	Insurance cost-sharing level	Medical spending, health outcomes	RAND / ICPSR: `https://www.icpsr.umich.edu/web/ICPSR/studies/6439`
Project STAR (Class Size)	Education (US)	RCT	Small vs regular class	Test scores (multi-year)	Harvard Dataverse: `https://dataverse.harvard.edu/dataverse/STAR`
National Job Corps Study	Labor (US)	RCT	Job training offer	Employment, earnings (long-run)	openICPSR replication data: `https://www.openicpsr.org/openicpsr/project/116226`
National Supported Work (NSW)	Labor (US)	RCT	Subsidized job assignment	Employment, earnings	ICPSR public-use: `https://www.icpsr.umich.edu/web/ICPSR/studies/7763`
Head Start Impact Study (HSIS)	Education / Health (US)	RCT	Head Start access	Cognitive, social, health outcomes	Public-use files: `https://www.acf.hhs.gov/opre/project/head-start-impact-study`
Kenya School Deworming	Education / Health (Kenya)	Cluster RCT	School-level deworming	Schooling, long-run earnings proxies	Replication data: `https://www.povertyactionlab.org/evaluation/deworming-schools-western-kenya`
Progresa / Oportunidades	Development (Mexico)	RCT / Phase-in	Cash transfer eligibility	Schooling, health, consumption	IFPRI / evaluation data: `https://www.ifpri.org/publication/progresa`
Quarter-of-Birth Schooling	Education / Labor (US)	Quasi-experimental (IV)	Compulsory schooling laws	Earnings, education (Census)	IPUMS Census data: `https://ipums.org`
Medicare Part D (PDE) Public Use Files	Health / Public finance (US)	Policy change (quasi-exp)	Drug coverage exposure post-2006	Prescription utilization, spending (admin claims)	CMS PDE PUF landing: `https://www.cms.gov/data-research/statistics-trends-and-reports/medicare-part-d-prescriber-data`
Minimum Wage (US State Panels)	Labor (US)	Diff-in-diff style (quasi-exp)	State minimum wage increases	Employment, hours, wages over time	CPS via IPUMS-CPS: `https://cps.ipums.org/cps/`

Table 1: Datasets to benchmark causalPFN.

3 Moving to Oppoertunity (MTO)

We evaluated treatment effects in the MTO cell-level PUF using the randomized assignment arms to define three intent-to-treat contrasts: (i) any voucher (experimental or Section 8) vs control, (ii) experimental vs control, and (iii) Section 8 vs control. Outcomes `Y` were four standardized adult index measures (overall, economic, mental health, and physical health). For each contrast/outcome, we repeatedly performed random train/test splits (10 splits, 80/20), fit causalpfn on the training set, and summarized performance on the test set. The table reports the distribution of split-level estimates (mean +- sd), including the estimated ATE and summary statistics of the predicted CATEs on the test folds. Note: n sizes are low for cell-level PUFs.

Level	Contrast	Outcome (y_col)	n	p	Splits	ATE_test (mean±sd)	CATE mean_test (mean±sd)	CATE sd_test (mean±sd)
cell	any_voucher_vs_control	mn_f_all_idx_fix_z_ad	81	43	10	0.089 ± 0.122	0.021 ± 0.044	0.101 ± 0.031
cell	any_voucher_vs_control	mn_f_ec_idx_z_ad	81	43	10	0.027 ± 0.095	-0.038 ± 0.045	0.102 ± 0.016
cell	any_voucher_vs_control	mn_f_mh_idx_z_ad	81	43	10	0.081 ± 0.098	0.069 ± 0.028	0.092 ± 0.036
cell	any_voucher_vs_control	mn_f_ph_idx_fix_z_ad	81	43	10	0.097 ± 0.136	0.034 ± 0.038	0.109 ± 0.020
cell	exp_vs_control	mn_f_all_idx_fix_z_ad	56	43	10	0.061 ± 0.171	0.053 ± 0.037	0.110 ± 0.034
cell	exp_vs_control	mn_f_ec_idx_z_ad	56	43	10	-0.020 ± 0.143	0.002 ± 0.039	0.091 ± 0.029
cell	exp_vs_control	mn_f_mh_idx_z_ad	56	43	10	0.099 ± 0.140	0.067 ± 0.037	0.096 ± 0.029
cell	exp_vs_control	mn_f_ph_idx_fix_z_ad	56	43	10	0.074 ± 0.147	0.044 ± 0.053	0.097 ± 0.030
cell	s8_vs_control	mn_f_all_idx_fix_z_ad	47	43	10	-0.009 ± 0.168	0.008 ± 0.048	0.080 ± 0.014
cell	s8_vs_control	mn_f_ec_idx_z_ad	47	43	10	-0.095 ± 0.117	-0.059 ± 0.045	0.074 ± 0.024
cell	s8_vs_control	mn_f_mh_idx_z_ad	47	43	10	0.039 ± 0.130	0.066 ± 0.053	0.075 ± 0.021
cell	s8_vs_control	mn_f_ph_idx_fix_z_ad	47	43	10	0.064 ± 0.199	0.031 ± 0.054	0.113 ± 0.034

Table 2: MTO cell-level PUF results aggregated across 10 random 80/20 splits. Reported values are mean ± standard deviation across splits for ATE estimates and for summary statistics of predicted CATEs on the test folds.

3.1 Benchmark: Kling, Liebman, Katz (2007, Econometrica) — "Experimental Analysis of Neighborhood Effects"

Early analysis of MTO, reporting treatment effects by domain (economic self-sufficiency, mental health, physical health, etc.) and contrasts Experimental vs Control and Section 8 vs Control. Note: this was published in 2007, data is likely outdated.

Level	Contrast	y_{col}	Domain	n	p	ATE_{obs}	$\text{ATE}_{\text{CausalPFN}}$	CATE mean	CATE sd	ITT (Kling)	SE	Device	Seed
cell	exp_vs_control	mn_f_all_idx_fix_z_ad	overall	56	43	0.0621628240	0.0642868504	0.0642868504	0.1389464587	0.036	0.02	cpu	9
cell	s8_vs_control	mn_f_all_idx_fix_z_ad	overall	47	43	0.0492463931	-0.0019842105	-0.0019842105	0.1239116341	0.028	0.022	cpu	9
cell	any_voucher_vs_control	mn_f_all_idx_fix_z_ad	overall	81	43	0.0566897541	0.0108168945	0.0108168945	0.1515614837			cpu	9
cell	exp_vs_control	mn_f_ec_idx_z_ad	economic	56	43	0.0049872044	0.0011409213	0.0011409213	0.1419877857	0.017	0.031	cpu	9
cell	s8_vs_control	mn_f_ec_idx_z_ad	economic	47	43	-0.0417265929	-0.0768484920	-0.0768484920	0.1482894570	0.037	0.033	cpu	9
cell	any_voucher_vs_control	mn_f_ec_idx_z_ad	economic	81	43	-0.0148067810	-0.0352992602	-0.0352992602	0.1824342012			cpu	9
cell	exp_vs_control	mn_f_mh_idx_z_ad	mental_health	56	43	0.0886466578	0.0892872214	0.0892872214	0.1010343209	0.079	0.03	cpu	9
cell	s8_vs_control	mn_f_mh_idx_z_ad	mental_health	47	43	0.1101593524	0.0589399152	0.0589399152	0.1083177254	0.029	0.033	cpu	9
cell	any_voucher_vs_control	mn_f_mh_idx_z_ad	mental_health	81	43	0.0977622122	0.0657870024	0.0657870024	0.1239330769			cpu	9
cell	exp_vs_control	mn_f_ph_idx_fix_z_ad	physical_health	56	43	0.0528132990	0.0517719127	0.0517719127	0.1469904333	0.012	0.024	cpu	9
cell	s8_vs_control	mn_f_ph_idx_fix_z_ad	physical_health	47	43	0.0611598864	0.0125098573	0.0125098573	0.1447815150	0.019	0.026	cpu	9
cell	any_voucher_vs_control	mn_f_ph_idx_fix_z_ad	physical_health	81	43	0.0563499890	0.0161366444	0.0161366444	0.1396971643			cpu	9
pseudo	exp_vs_control	ps_f_all_idx_fix_z_ad	overall	2595	43	0.0632191896	0.0638474897	0.0638474897	0.1560544521	0.036	0.02	cpu	9
pseudo	s8_vs_control	ps_f_all_idx_fix_z_ad	overall	1817	43	0.0597817935	0.0141260298	0.0141260298	0.2515113056	0.028	0.022	cpu	9
pseudo	any_voucher_vs_control	ps_f_all_idx_fix_z_ad	overall	3273	43	0.0621270910	-0.0007078123	-0.0007078123	0.1737019867			cpu	9
pseudo	exp_vs_control	ps_f_ec_idx_z_ad	economic	2593	43	0.0019359384	-0.0283293221	-0.0283293221	0.1181795821	0.017	0.031	cpu	9
pseudo	s8_vs_control	ps_f_ec_idx_z_ad	economic	1815	43	-0.0425378121	-0.0930919200	-0.0930919200	0.1986617297	0.037	0.033	cpu	9
pseudo	any_voucher_vs_control	ps_f_ec_idx_z_ad	economic	3271	43	-0.0099142212	-0.0588981248	-0.0588981248	0.1117210761			cpu	9
pseudo	exp_vs_control	ps_f_mh_idx_z_ad	mental_health	2595	43	0.0846371874	0.1103355438	0.1103355438	0.1189373508	0.079	0.03	cpu	9
pseudo	s8_vs_control	ps_f_mh_idx_z_ad	mental_health	1817	43	0.1096196622	0.0685396716	0.0685396716	0.2110136002	0.029	0.033	cpu	9
pseudo	any_voucher_vs_control	ps_f_mh_idx_z_ad	mental_health	3273	43	0.0925744474	0.0571317375	0.0571317375	0.1905301660			cpu	9
pseudo	exp_vs_control	ps_f_ph_idx_fix_z_ad	physical_health	2595	43	0.0575969666	0.0368078053	0.0368078053	0.1436257511	0.012	0.024	cpu	9
pseudo	s8_vs_control	ps_f_ph_idx_fix_z_ad	physical_health	1817	43	0.0755528808	0.0076189372	0.0076189372	0.2907862067	0.019	0.026	cpu	9
pseudo	any_voucher_vs_control	ps_f_ph_idx_fix_z_ad	physical_health	3273	43	0.0633018017	0.0398856252	0.0398856252	0.2021370381			cpu	9

Table 3: Replication comparison for MTO adult indices using both the cell-level public-use file (cell) and the pseudo-individual public-use file (pseudo). Published ITT entries correspond to Kling, Liebman, and Katz (2007, *Econometrica*) Table II (All Adults) for Experimental vs Control and Section 8 vs Control.