

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Essays on the Economics of Public Service Delivery

Permalink

<https://escholarship.org/uc/item/32r8v6rp>

ISBN

9798189270475

Author

Tamim, Abdulrazzak

Publication Date

2026-05-01

Peer reviewed|Thesis/dissertation

Essays on the Economics of Public Service Delivery

by

Abdulrazzak Tamim

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Edward Miguel, Chair

Professor Nano Barahona

Professor Kelsey Jack

Professor Nick Tsivanidis

Spring 2026

Essays on the Economics of Public Service Delivery

Copyright 2026
by
Abdulrazzak Tamim

Abstract

Essays on the Economics of Public Service Delivery

by

Abdulrazzak Tamim

Doctor of Philosophy in Economics

University of California, Berkeley

Professor Edward Miguel, Chair

This dissertation studies how the design of allocation and delivery systems shapes access to essential goods and services in low- and middle-income countries. Across three chapters, I examine the welfare consequences of non-price rationing, the effects of housing subsidies on forced migrants' well-being, and the barriers to digital payment adoption for utilities. In Chapter 2, I quantify the welfare cost of non-price rationing in a water market. Using the universe of meter-level billing data from Amman, Jordan, I develop and estimate a structural model of consumer demand under rationing. The analysis shows that seemingly inefficient allocation mechanisms such as access restrictions are a consequence of governments' inability to implement the optimal mechanism of pricing and redistribution of the surplus. Chapter 3 provides experimental evidence on the effects of housing subsidies for Syrian refugees in Mafrq and Irbid, Jordan. The results offer a cautionary tale: the program had minimal impacts on consumption and expenditure, and many households reported that they would have preferred cash transfers. These findings are among the first experimental evidence on how housing policy shapes the welfare of displaced populations. Chapter 4 investigates barriers to the adoption of digital payments for water services in Addis Ababa, Ethiopia. Despite the potential of digital payments to reduce transaction costs, adoption remains low. Using a randomized controlled trial that separately varies information provision and incentives to pay digitally, we find that both interventions enhance adoption, but incentives are more cost-effective.

To the academics and students of Palestine

Contents

Contents	ii
List of Figures	iv
List of Tables	vi
1 Introduction	1
2 Rationing by Intermittency	3
2.1 Introduction	3
2.2 Institutional Setting	7
2.3 Data and Descriptives	8
2.4 A Model of the Water Market	11
2.5 Quasi-Experimental Evidence	25
2.6 Structural Estimation	29
2.7 Counterfactual Analysis	39
2.8 Conclusion	46
2.9 Figures and Tables	48
3 Housing Subsidies for Forced Migrants	60
3.1 Introduction	60
3.2 Background and Context	64
3.3 Experimental Design and Data	67
3.4 Empirical Strategy	72
3.5 Housing Subsidy Program Impacts on Refugee Life Outcomes	74
3.6 Impacts on Neighbor Attitudes	80
3.7 Addressing Low Take-up Rates	82
3.8 Comparing Forecasts to Estimated Program Impacts	83
3.9 Conclusion	84
3.10 Figures and Tables	87
4 Digital Payments Adoption	97
4.1 Introduction	97

4.2	Institutional Setting	100
4.3	Experimental Design	101
4.4	Data, Descriptives, and Internal Validity	103
4.5	Empirical Strategy	105
4.6	Results	106
4.7	Mechanisms	111
4.8	Expert Predictions	113
4.9	Cost Effectiveness	114
4.10	Conclusion	115
4.11	Figures and Tables	117
Bibliography		133
A Appendix to Rationing by Intermittency		143
A.1	Solution Algorithm	143
A.2	Additional Figures and Tables	144
A.3	Omitted Proofs	151
A.4	More Details on the RDD	169
A.5	More Details on the Price Instrument	174
B Appendix to Housing Subsidies for Forced Migrants		178
B.1	Estimates Correcting for Non-Random Attrition	179
B.2	Heterogeneous Effects Analysis	184
B.3	Additional Results	186
B.4	Addressing Low Take-up Rates	188
C Appendix to Digital Payments Adoption		197
C.1	Sample Representativeness	197
C.2	Additional Figures and Tables	198
C.3	Statistical Power	201

List of Figures

2.1	Spatial Variation in Rationing	48
2.2	Patterns by Income Quintile	50
2.3	Impact of Supply Intensity on Consumption	51
2.4	Estimation Architecture	52
2.5	Regime Shares by Expenditure Quintile	53
2.6	Welfare Decomposition: Gains Relative to Status Quo	57
2.7	Distribution of Winners and Losers	57
2.8	Distributional Incidence by Expenditure Quintile	57
2.9	The Hidden Cost of Intermittency	59
3.1	Geographic Location of Treated and Control Communities	87
3.2	RCT, Data Collection, and Forecasts Timeline	88
3.3	Program Impacts vs. Expert Predictions	96
4.1	Experimental Timeline	122
4.2	Payment Trends	123
4.3	Digital Adoption Rates by Arm	124
4.4	Treatment Effect by Payment Channel	125
4.5	Heterogeneous Effects on Primary Outcomes (First Dose)	126
4.6	Heterogeneous Effects on Primary Outcomes (Second Dose)	127
4.7	Heterogeneous Effects on Primary Outcomes (Post-Treatment)	128
4.8	Heterogeneous Effects on Secondary Outcomes	129
4.9	Treatment Effects on Barriers to Digital Payment	130
4.10	Forecasts vs. Realized Treatment Effects	131
4.11	Distribution of Incentive Transfers Among Incentive Arm Redeemers	132
A.1	Patterns by Expenditure Quintile	144
A.2	Distribution of Water Supply Days	145
A.3	Change in Water Consumption at the Supply Boundary	145
A.4	Observed and Model-Optimal Storage Capacity	146
A.5	Storage, Shadow Value, and Supply Response	147
A.6	Private Marginal Value of an On-Day	150
A.7	Manipulation Density Test	171

A.8	Impact of Supply Intensity on Consumption (Donut)	173
B.1	Heterogeneity Analysis: LASSO-Based Predicted Compliance	191
C.1	Administrative Universe vs. Experimental Sample	197

List of Tables

2.1	Structural MLE Estimates	49
2.2	Model Comparison with RDD Estimate	49
2.3	Summary of Structural Parameter Calibration	54
2.4	Social Marginal Cost Decomposition	55
2.5	Counterfactual Equilibrium Outcomes	56
2.6	Distributional Effects of Pricing Reform	58
3.1	Average Respondent, Household, and Shelter Characteristics	89
3.2	Forecasts of Treatment Effects	90
3.3	Panel Retention and Compliance	91
3.4	Pooled Treatment Effects on Primary Outcomes	92
3.5	Statistically Significant Impacts of the Program at Midline	93
3.6	Statistically Significant Impacts of the Program at Endline and Follow-up	94
3.7	Impacts of the Program on Neighbor's Attitudes Toward Refugees	95
4.1	Transition of Payment Modes	117
4.2	Primary Effects (Admin Data)	118
4.3	Secondary Effects (Survey Data)	119
4.4	Payment Transitions from January to February	120
4.5	Cost Effectiveness Estimates	121
A.1	Econometric Regimes	146
A.2	Data Sources and Identification	147
A.3	Determinants of Observed Storage Capacity	148
A.4	Observed vs. Model-Implied Storage Capacity	148
A.5	Oversized Storage by Income Quintile	149
A.6	Common Tank Sizes	149
A.7	Observed vs. Predicted Moments	149
A.8	Household Classification	152
A.9	Shadow Prices by Regime	153
A.10	Private Marginal Value of an Additional Day by Regime	153
A.11	Household Classification	157
A.12	Instrumental Variables Estimates of the Demand Slope	176

A.13	Robustness to Alternative Lag Lengths for the Instrument	177
B.1	Pooled Treatment Effects on Primary Outcomes (with Lee Bounds)	180
B.2	Significant Impacts at Midline (with Lee Bounds)	181
B.3	Significant Impacts at Endline and Follow-up (with Lee bounds)	182
B.4	Impacts on Neighbor's Attitudes Toward Refugees (with Lee Bounds)	183
B.5	Gender Heterogeneity at Endline	184
B.6	Neighbor Impacts on Social Attitudes: Heterogeneity	185
B.7	Impacts of the Program on Earnings, Labor, & Occupational Choice	186
B.8	Robustness of Baseline Balance	187
B.9	Baseline Averages Interacting Treatment with Compliance	188
B.10	Variables Selected for LASSO	189
B.11	Primary Outcomes	192
B.12	Primary Outcomes and Selected Social Attitudes Index Components	192
B.13	Pooled Treatment Effects on Primary Outcomes	193
B.14	Significant Impacts of the Program at Midline	194
B.15	Significant Impacts of the Program at Endline and Follow-up	195
B.16	Impacts on Neighbor's Attitudes Toward Refugees	196
C.1	Balance of Baseline Covariates Across Treatment Arms	198
C.2	Attrition Rates	199
C.3	Baseline Means, Attriters vs. Stayers	199
C.4	Forecasts and Realized Treatment Effects	200
C.5	Ex-Post Statistical Power	201

Acknowledgments

I am indebted to my dissertation committee for their mentorship throughout the PhD journey. Ted Miguel reminded me to never lose sight of the big picture. His genuine excitement helped draw me to the PhD and kept me focused on answering important, policy-relevant questions. Nano Barahona expanded how I think about identification in economics and generously helped me throughout, including with building and estimating my model. Kelsey Jack supported me since the beginning, helped me transform my ideas from a very vague pitch to a full paper, and provided detailed and thoughtful comments. Nick Tsivanidis challenged and inspired me to understand what tradeoffs policymakers face and put them into words before getting into equilibrium equations. I also thank Lucas Davis, Ben Faber, Meredith Fowlie, Marco Gonzalez-Navarro, Emmanuel Saez, Jim Saltee, Joe Shapiro, and Reed Walker for comments and discussions.

Chapter 2 benefited from data generously shared by Miyahuna, Jordan's water utility in Amman. I thank Nick Tsivanidis for introducing me to the International Growth Centre team who facilitated access and to Ghadeer Noufal for making it happen. I am also grateful to Sofia Belhouari who provided excellent research assistance through the Economics Department's Undergraduate Research Assistantship Program.

Chapter 3 is part of the Syrian Refugee Life Study (SRLS) and is co-authored with Sam Leone, Edward Miguel, Bailey Palmer, Sandra Rozo, Emma Smith, and Sarah Stillman. Mindset did an excellent job collecting the data. Working with vulnerable populations through SRLS reinforced for me how much lived human experience lies behind every data point, a reminder I carry into all of my research.

Chapter 4 is co-authored with Joseph Levine and Tsega Adego. I thank Kelsey Jack for connecting me with the team, the Policy Studies Institute in Ethiopia for leading data collection, and the Addis Ababa Water and Sewerage Authority for sharing the universe of billing data that constituted our sampling frame and primary outcomes.

Finally, I would like to thank my family and friends for unconditionally supporting me throughout my life and during my PhD. To my wife, Tamara, I cannot put into words what an impact you have had on my life. Thank you for believing in me even when I did not, for showing me that the sky is the limit, and for inspiring me to become a better researcher and a much better human being.

Chapter 1

Introduction

Governments in low- and middle-income countries face a fundamental challenge: how to deliver essential goods and services to their populations when resources are scarce and administrative capacity is limited. Water has to be allocated among competing users, housing assistance needs to reach displaced populations, and payment systems must function at scale. The design of these systems including how prices are set, how transfers are structured, and how transactions are processed, has direct consequences for household welfare. Yet despite the central role that policy design plays in service delivery, we have limited evidence on the welfare costs of common allocation mechanisms and on what works in expanding access.

This gap matters because the stakes are large. Over two billion people live in water-stressed countries (World Health Organization, 2023), and the allocation rules that governments adopt (whether pricing, rationing, or access restrictions) shape who gets water and at what cost. Over 100 million people were forcibly displaced worldwide and depend on assistance programs whose design determines whether aid translates into improved welfare (UNHCR, 2025). And across the developing world, utilities and governments spend substantial resources managing in-person payment systems that digitization could make more efficient. In each case, the design of the policy is not incidental to its success but is rather central.

A common difficulty in evaluating these policies is that allocation rules, program participation, and payment behavior are all endogenous. Households that face rationing differ from those that do not; refugees who receive housing subsidies may differ from those who do not; and citizens who pay digitally may differ from those who pay in person. Each chapter of this dissertation addresses this identification challenge using a different methodology. The first chapter combines a regression discontinuity design with a structural model of consumer demand under rationing, using the universe of meter-level billing data to recover willingness to pay and simulate counterfactual allocations. The second and third chapters use randomized controlled trials to isolate the causal effects of specific policy interventions.

The first chapter examines how governments allocate scarce water resources using a combination of prices, quantity caps, and access restrictions. I build a structural model that quantifies the welfare costs of these allocation mechanisms and apply it to the water market in Amman, Jordan, one of the most water-scarce countries in the world. The main finding is that seemingly inefficient

policies such as rationing are a consequence of governments' inability to raise prices and redistribute surplus. In a sense, the Second Welfare Theorem fails to hold in these low-state-capacity settings where the political and administrative tools needed to implement transfers are unavailable.

The second chapter turns to another question: what are the impacts of housing assistance to forcibly displaced populations on their life outcomes? A growing literature examines cash transfers for low-income households, but little is known about the effects of housing subsidies on recipient welfare. We explore this question among Syrian refugees in Mafraq and Irbid, Jordan. Our results indicate that even though the program we evaluated enhance housing quality, it had minimal impacts on consumption and expenditure, and many households reported that they would have preferred cash transfers.

The third chapter focuses on a different but related policy design question: how to push digitization in bill payment systems. Governments around the world spend significant resources managing in-person payments for essential services, and digitization promises to reduce these costs. Yet many citizens continue to pay in person despite the availability of digital alternatives. To understand the barriers to digital payment adoption, we collaborated with the water utility in Addis Ababa, Ethiopia, and tested both incentives and information interventions. We find that both enhance digital payment adoption, but incentives are more cost-effective.

Taken together, these three chapters show that the design of allocation and delivery systems has first-order consequences for welfare, and that common policy choices including rationing water, delivering in-kind housing assistance, and maintaining in-person payment systems can be improved with careful analysis and experimentation.

Chapter 2

Rationing by Intermittency

2.1 Introduction

How to allocate scarce resources when aggregate demand exceeds supply? The standard economic prescription is to set the price at the social marginal cost, let the market clear, and redistribute through transfers. When transfers are unavailable, however, the planner must use the price itself as a redistribution device, accepting allocative distortions for distributional gains (Feldstein, 1972). Even though this may seem straightforward in principle, governments especially in the developing world overwhelmingly choose various forms of rationing over the price mechanism with redistribution.

Rationing is itself a redistribution device, and economists have long studied how it works when price is not used. The classical cases fall into two categories: quantity caps restrict how much each consumer can buy, and queuing mechanisms ration through waiting time (Tobin, 1952; Barzel, 1974; Weitzman, 1991). Both sets of models start from a good that remains continuously available; what is rationed is how much or how fast one can obtain it. In this paper, I explore a new form of rationing which works along a different dimension: access itself. This policy shifts the locus of adjustment from a flow margin (time foregone, consumption cut) to a stock margin: households invest in private capital to smooth consumption across supply interruptions. The resulting capital stock is deadweight loss, and one that likely falls regressively on the households the rationing was meant to protect. Despite its prominence, this storage is a welfare cost that neither the quota nor the queuing frameworks are set up to measure.

Quantifying the welfare costs of rationing by intermittency and its equity implications is also important from the perspective of governments and policymakers. Estimates indicate that intermittent water access is widespread: over one billion people receive piped water on an intermittent schedule (Bivins et al., 2017). Rather than flowing continuously, the pipe is pressurized for only a few hours or days per week. Unreliable availability also remains a prominent feature of water service in low- and middle-income countries (WHO and UNICEF, 2025). Yet the economics literature has largely treated intermittency as an exogenous feature of the environment rather than as a policy instrument. To fill this gap, I ask: what is the welfare cost of this form of rationing, and

what prevents governments from implementing the pricing mechanism with redistribution?

To answer this question, I develop a structural model of the water market in which a regulator chooses from three instruments: a volumetric price, a uniform quota, and the number of supply days. Heterogeneous households respond by investing in private storage capacity before observing idiosyncratic preference shocks, and then choose consumption given their constraints. I estimate the model using administrative billing data covering around 910,000 households in Amman, Jordan, one of the most water-scarce cities in the world, and validate it using a spatial regression discontinuity design at rationing zone boundaries.

The analysis yields three findings. First, transitioning from the intermittent status quo (2 supply days per week at an administered price of 0.76 JD/m³) to counterfactual market-clearing pricing with standard redistributive preferences generates a social welfare gain of 3.1% of status quo welfare, with optimal prices of 6.14 JD/m³ (roughly eight times the current rate) and supply expanding to 3.75 days per week. The gain comes at the expense of consumer welfare: without redistribution, every household is worse off. With a flat lump-sum rebate, the loss reverses partially and 49.3% of households gain. Mean household storage investment also falls by 38%, from 3.19 to 1.98 m³, confirming that the private storage capital stock under intermittency is predominantly deadweight loss.

Moreover, this deadweight capital is regressive: storage costs consume roughly 7.5% of weekly expenditure for the poorest quintile (Q1) but only 2.3% for the wealthiest (Q5) under the status quo. The actual direct bill payment (excluding storage) as implied by the model constitutes around 5% for Q1 and less than 1% for Q5 households. These results have two implications. First, intermittency effectively operates as a regressive implicit tax collected outside the fiscal system through forced private capital investment. Second, storage costs are substantial share of total costs and they have been unquantified in the water allocation literature to the best of my knowledge.

Second, consider pricing on a rehabilitated network, where comprehensive pipe replacement reduces the network's non-revenue water from 46% to 11%, consistent with results achieved in pilot rehabilitation zones in Amman (USAID Water Governance Activity, 2022). Because the iceberg multiplier is lower at every supply duration, the planner can extend supply much closer to continuous service before the marginal leakage cost binds. Combined with the same lump-sum transfers, this generates welfare gains of 6.4% relative to the status quo. Storage drops to 0.5 m³, an 85% decline from the status quo. Even though these gains are larger than the degraded network results, the allocative efficiency of the price instrument does not require a costly network upgrade; it arises from reallocating water toward higher-valued uses and from extending supply duration. This finding is important because it implies that the welfare case for pricing reform does not depend solely on the government's ability to finance large infrastructure investments but rather on its ability to redistribute the revenue.

The third counterfactual I consider is a uniform quota (capping each household's consumption at a fixed volume while keeping the administered price fixed). This yields a similar consumer surplus relative to the status quo. At the feasible supply duration on degraded pipes, scarcity is already managed by the physical pipe constraint, so the optimal quota of 4.5 m³ per week does not reduce consumption by much. Overall, the quota instrument yields allocative outcomes similar to the status quo while imposing the additional fixed cost of smart metering infrastructure, making

quotas comparable or slightly worse than the status quo in terms of aggregate social welfare.

Results up to here establish that pricing with redistribution increases consumer surplus, but why is intermittency inefficient, even for standard redistributive preferences? Intermittent supply reduces welfare through two channels. The first is a *quantity* channel: the pipe's limited pressurization window strictly bounds how much water any household can extract, compressing the consumption distribution from above. The second is a *smoothing* channel: finite storage capacity forces households to consume unevenly across on-days and off-days, generating a temporal distortion even when total volume is unconstrained. Private storage investment partially offsets the smoothing channel but constitutes a deadweight loss, representing a stock-channel analogue of classical rent dissipation in queuing models (Barzel, 1974; Deacon and Sonstelie, 1985).

In the final set of results, I compute summary statistics of the welfare weight distribution that govern optimal policy. Two summary statistics emerge from the planner's first-order conditions. The first, the consumption-weighted welfare weight $\bar{g}_q$, governs the optimal price: it is essentially a function of the covariance between social welfare weights and consumption. When $\bar{g}_q > 1$ (< 1), heavier water consumers tend to be poorer (richer) households, so a volumetric subsidy (tax) is progressive. The second summary statistic, $\bar{g}_\pi$, is to my knowledge a new welfare-relevant object in the optimal taxation literature. It governs optimal supply duration: it is a function of the covariance between welfare weights and a household's private marginal value of an additional supply day π_i . Because wealthier households tend to have higher demand and therefore bind more frequently against physical constraints, π_i is generally increasing in income, pushing $\bar{g}_\pi < 1$. In this case, the households who benefit most from extended supply are precisely those the planner down-weights, weakening the equity case for expanding supply duration.

Empirically, both summary statistics are close to but strictly below unity ($\bar{g}_q = 0.97$, $\bar{g}_\pi = 0.96$), reflecting flat Engel curves in both consumption volume and the marginal value of supply: poor and rich households consume similar quantities of water, and wealthier households benefit only modestly more from an additional supply day. Two implications follow. First, neither the price nor the supply-duration instrument is itself a strong redistributive tool: the price change and the supply change that accompany pricing reform are each roughly distributionally neutral in their direct incidence. The progressivity of the reform that emerges in the counterfactuals therefore comes almost entirely from the *rebate of the fiscal surplus*, not from who bears the higher price or who gains the extra supply day. Second, because $\bar{g}_\pi < \bar{g}_q$, supply duration is if anything a slightly *worse* redistributive instrument than pricing: extending supply disproportionately benefits high-demand wealthier households. Together these results reframe the policy lever. Pricing reform is valuable in this setting not because the price itself progressively targets the poor, but because it generates revenue that the state can then redistribute through cash transfers if feasible.

This paper makes four contributions. First, I contribute to the literature on water allocation and pricing (Timmins, 2002; Szabó, 2015; Rafey, 2023; Ryan and Sudarshan, 2022; Donna and Espín-Sánchez, 2023; Abajian et al., 2025). Within this literature, Abajian et al. (2025) study drought adaptation and inequality in Cape Town's water system, Baisa et al. (2010) estimate the welfare costs of unreliable water service, and Hagerty (2023) quantifies transaction costs in water markets. My contribution is to model intermittent supply as a rationing technology and to quantify the welfare cost of replacing pricing with physical rationing, departing from the existing literature that

treats water rationing as a quantity cap. Moreover, unlike settings where time rationing substitutes for externality pricing, intermittent residential water supply involves no commons problem.

Second, I bridge the public finance literature on optimal commodity taxation with the applied setting of regulated utilities in developing countries. The Ramsey–Feldstein framework (Ramsey, 1927; Feldstein, 1972; Diamond, 1975) and its applications (Borenstein, 2012; Levinson and Silva, 2022; Allcott et al., 2019; Jacobsen et al., 2020) consider settings where the good is delivered continuously and the price is the sole instrument. Related work on infrastructure pricing in developing countries includes McRae (2015) on subsidy traps in electricity and Burgess et al. (2020) on the consequences of treating electricity as a right. I extend the Ramsey–Feldstein framework to an environment where rationing by intermittency creates a second instrument (supply duration) with its own distributional characteristic ($\bar{q}_\pi$), so that the planner’s problem involves two distinct equity-efficiency tradeoffs.

Third, I contribute to the literature on non-price rationing and its welfare consequences. This literature typically treats rationing as an exogenous constraint (Tobin, 1952; Neary and Roberts, 1980) or a clearing mechanism via either waiting (Barzel, 1974; Lindsay and Feigenbaum, 1984; Deacon and Sonstelie, 1985; Frech and Lee, 1987) or search and inventory (Weitzman, 1991). The welfare costs of non-price allocation have also been quantified in many markets such as housing (Glaeser and Luttmer, 2003; Diamond et al., 2019), energy (Davis and Kilian, 2011), fertility (Guo et al., 2025), and bans (Davis, 2008; Barahona et al., 2020). My contribution relative to this literature is to identify a novel *stock* channel of rent dissipation: whereas queuing models emphasize flow costs (time spent waiting), intermittent water supply forces households to accumulate a deadweight capital stock that constitutes the ongoing welfare cost of the rationing regime.

Fourth, I connect to the literature on private adaptation to environmental risk. Within this literature, Barreca et al. (2016) document the diffusion of residential air conditioning as the principal margin of adaptation to extreme heat, Carleton et al. (2022) develop a revealed-preference framework to recover the unobserved costs of adaptation at global scale, and Balboni et al. (2026) and Hsiao (2024) study relocation and coastal development in response to flood risk. My contribution is to study adaptation to a constraint that is itself a policy choice. Closest in spirit is Deschênes et al. (2017), who formalizes defensive investments in response to environmental disamenities; I depart from this framing by studying investment against a policy-created rather than a physical threat. Storage capital in this setting does not mitigate physical scarcity; it compensates for the supply schedule itself. Private storage is therefore deadweight in a stronger sense than climate-adaptation capital as removing the policy eliminates both the cost and the need for the capital, a counterfactual that is unavailable in the climate setting.

The paper proceeds as follows. Section 2.2 describes the institutional setting. Section 4.4 details the data. Section 2.4 develops the structural model. Section 2.5 presents quasi-experimental reduced-form evidence. Section 2.6 estimates the structural parameters. Section 2.7 presents counterfactual policy simulations, and Section 2.8 concludes.

2.2 Institutional Setting

In the face of water scarcity, rising urban populations, and limited infrastructure, governments around the world have implemented various forms of water rationing. These schemes differ fundamentally in how they constrain households. Some recommend and aim to impose use restrictions or volumetric caps while maintaining continuous pressurization, such as odd-even lawn watering in Los Angeles or Cape Town's 50-liter daily limit during its 2018 drought. Others physically depressurize the network on a rotating schedule, so that no water flows from the tap for days at a time.

Such intermittent supply schemes are widespread: Bogotá implemented 24-hour rotating shut-offs across nine zones during its 2024 drought; cities across Mexico, India, Pakistan, and much of the Middle East operate under supply schedules. Jordan exemplifies this more severe form of rationing, making it a valuable setting to study how households respond to supply constraints, the welfare costs of intermittent service, and the efficiency implications of non-price allocation mechanisms.

Jordan is one of the most water scarce countries in the world, with per capita supply averaging around 33 gallons per day based on the Ministry of Water and Irrigation data; the analogous US estimates are at 82 gallons (Dieter et al., 2018). The water network is fed in part by local groundwater wells, whose proximity to a given zone affects hydraulic pressure and effective supply duration. Total municipal consumption (residential, commercial, and institutional users served by the piped network) in 2021 stood at 519 million cubic meters (MCM), of which roughly 73% is groundwater and 27% is surface water. Total groundwater extraction exceeds the estimated safe yield by more than a factor of two, implying that the aquifer is being depleted even under the current rationing regime.

Water is also heavily subsidized and the sector faces deep deficits, which are financed through concessional loans from international donors and the accumulation of arrears to the national electricity utility. The government ultimately clears some of these debts through direct treasury advances and sovereign debt guarantees. The situation is exacerbated by the high non-revenue water: around 40% of the water that gets pumped does not get billed, driven by a combination of aging infrastructure leakage and likely unauthorized use.

In the face of scarcity and leakage, as in many countries around the world, the government has responded to demand exceeding supply by rationing water. The Ministry has implemented a rotation scheme where Amman is divided into "water zones", and in each zone households are assigned water from a specific time to another, for example Wednesday 10 am to Thursday 4 pm. Supply windows vary across zones, and the schedule is publicly posted and predictable. For a household in a zone receiving for example two days of supply, the pipe is physically depressurized for the remaining five days, during which the household must rely entirely on privately stored water. This rotating schedule will be the base of the regression discontinuity design, which I explain later in more detail (Section 2.5).

2.3 Data and Descriptives

Data Sources

Universe of Amman’s Billing Data. The Ministry of Water and Irrigation (MWI) provided access to the utility’s universe of billing data, which cover all billing cycles over 2020-2024 in Amman, Jordan.¹ Meters are read physically by utility staff typically at the end of each billing cycle, which was quarterly from 2020 through 2023 and switched to monthly in 2024. The variables included are the meter number, water consumption, total bill, and the location of the meter. The original dataset includes residential and non-residential users, but I restrict the analysis to approximately 910,000 residential consumers.

Water Supply Data. To define zone-level treatment, I use administrative shapefiles provided by Miyahuna, Amman’s water utility. These shapefiles delineate the exact boundaries of the utility’s rationing districts and include the official supply schedule for each zone (start day/time and end day/time, e.g., “Sunday evening to Tuesday morning”).

To compute supply duration, I extract the start and end day and time for each zone. For example, a schedule reading “Sunday evening to Tuesday morning” corresponds to approximately 1.5 days of supply per week. For the two zones with multiple supply windows (less than 1% of zones), I sum their disjoint intervals. Zones labeled as “continuous supply” are coded as 7 days per week. I then locate each household within the water zone network. Using Google’s Geocoding API, I convert the detailed address recorded in the billing data into latitude and longitude coordinates. Each geocoded household is spatially joined to the utility’s rationing district shapefiles using a point-in-polygon algorithm, assigning it the supply schedule of its containing zone.

To construct the running variable for the regression discontinuity design, I compute the shortest perpendicular distance from each household to the nearest boundary separating zones with *different* supply intensities. I exclude all boundary segments between adjacent zones that share the same supply schedule. Because treatment status depends on the supply *relative* to the adjacent zone, a single zone may contain both treated and control units. Figure 2.1a illustrates this variation. Households near the boundary of a zone are classified as “low-supply” (control) relative to a higher-supply neighbor (treatment) if they have fewer water days. Figure 2.1b also displays the borders that are included in a regression discontinuity design to identify the causal effect of supply intensity on consumption. To compare similar households, I control for border fixed effects in all regressions in the estimation. Section 2.5 also presents results separately by gap magnitude (e.g., a 1-day difference, 2-day difference, etc.) and by baseline supply level.

Groundwater Production Data. Data on groundwater extraction was provided by MWI. The raw administrative database contains approximately 65,000 historical well-year observations na-

¹Note that in this context, there is a private alternative: tanker truck water. This substitute is about 4.8 times as expensive as piped water, but it accounted for only 4.6% of household’s water consumption in 2015 (Klassert et al., 2023). Because the private market represents a minority of residential consumption, I do not model it explicitly. I come back to this point later when providing descriptives.

tionwide. To construct the relevant cross-section for this analysis, I first restrict the sample to records from 2020 onward and spatially intersect the coordinates with the administrative boundaries of the Amman water distribution zones. I then collapse the panel to the unique well level, averaging production volumes across the observed years. This yields a final sample of 294 unique, active wells within the study area. I link groundwater extraction to household consumption by examining how local well production interacts with the spatial discontinuity in piped water supply.

Household-Level Survey. The structural model also employs household-level panel data from the Impact Evaluation of the Millennium Challenge Account (MCA) Jordan Compact (Jeuland et al., 2020). The sampling frame is based on the population census and consists of a geographically stratified urban households distributed across two primary regions: the Zarqa Governorate and East Amman. There are four rounds of data (2014, 2015, 2016, and 2018) with roughly 3,500 households in each round. Thus, even though the data is not representative of a randomly drawn household in Amman, it can shed light on some of the key channels needed for the structural estimation.

During the survey rounds, enumerators collected highly granular data to construct measures of household welfare and water security. To quantify water storage capacity, the survey explicitly recorded the physical volumetric capacity of all active storage infrastructure. This includes distinguishing between small containers ($\leq 0.02 \text{ m}^3$) and large ground-level or rooftop tanks ($> 0.02 \text{ m}^3$), measured and aggregated. Furthermore, household income and general expenditures were constructed by systematically probing and aggregating specific sub-categories rather than relying on single top-line estimates. Income encompasses formal salaries, informal business activities, and direct transfers, while expenditures track broad consumption alongside highly specific water-related costs, capturing shifts between standard utility bills and costly alternative sources such as vended tanker water.

Descriptive Facts

To motivate the structural model of rationing developed in Section 2.4, I document four stylized facts regarding how the burden of this rationing is distributed across the income spectrum. All of the facts are qualitatively similar if we use total expenditure instead of income (see Figure A.1 in the Appendix).

Fact #1: The financial burden of water is deeply regressive. Despite heavy subsidization, the water budget share falls steeply from the poorest income quintile to the wealthiest (Figure 2.2, top-left). This empirical gradient establishes why managing scarcity through prices is socially costly. Because this budget share is roughly three times higher for the poorest quintile than for the wealthiest, resolving aquifer scarcity through market-clearing prices would act as a highly regressive tax, disproportionately reducing the consumer surplus of the poor.

Fact #2: Water consumption is highly income-inelastic. The data show a weak consumption-income gradient (Figure 2.2, top-right): across a fivefold increase in income from Q1 to Q5, mean monthly consumption rises by only $\sim 10\%$ (12.2 to 13.6 m³) and median consumption by $\sim 37\%$ (8.0 to 11.0 m³), implying an income elasticity of 0.11 at sample means. Consistent with Fact #1, this means the regressivity of water expenditure arises not because the poor consume much less water, but because a similar bill absorbs a much larger share of their income. Because the volumetric subsidy transfers roughly similar absolute amounts across the income distribution, the tariff has limited distributive power relative to direct cash transfers.

Fact #3: Storage scales with income much faster than consumption. Fact #2 shows that physical water consumption rises only modestly with income. Private storage, by contrast, scales more: mean storage capacity rises from $4,942$ liters in Q1 to $6,489$ liters in Q5 (31% increase) with a much steeper slope at the top of the distribution (Figure 2.2, bottom-left). Richer households are not primarily buying more total water from the network; they are buying the ability to smooth consumption across days when the pipe is shut off. In the structural model, the physical pipe capacity acts as a hard ceiling on total extraction during on-days. A larger tank cannot absorb more water from a depressurized pipe; it allows the household to transfer their fixed allotment from “on-days” to “off-days.” This distinction matters for the welfare decomposition. Because tank investment is a private choice, households buffer themselves against the smoothing distortion, so the residual temporal smoothing loss in equilibrium is small. The pipe ceiling, however, is a public-capacity constraint that no private investment can relax: no amount of storage lets a household extract more water than what the pipe delivers over the cycle. High-demand households therefore remain bound by absolute volume, and it is this quantity channel (not the smoothing channel) that dominates the welfare cost of intermittency in the structural results.

Fact #4: Households have limited alternatives to the piped network. A potential concern when modeling intermittent piped water is whether households have access to alternative sources that would make the rationing constraint less binding. The most prominent private alternative is tanker truck delivery. The data suggest that this alternative plays a minimal role, with tankers functioning as an emergency backstop as in Baisa et al. (2010). As shown in the bottom-right panel of Figure 2.2, the extensive margin of tanker usage is remarkably low: across all income groups, only around 5% of households purchase any tanker water in a given week. Furthermore, the intensive margin (volume conditional on buying) is relatively flat across quintiles, indicating a fixed emergency requirement rather than income-elastic demand. Because over 94% of households rely exclusively on the municipal pipe and their private storage capital, omitting the private tanker market from the structural model allows for a cleaner focus on the primary physical constraints without sacrificing the core economic mechanisms of the rationing environment.

2.4 A Model of the Water Market

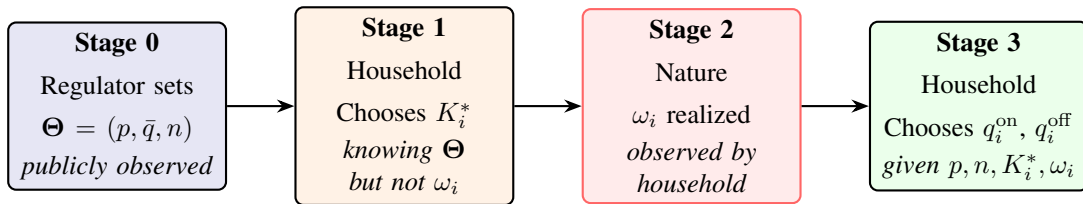
Setup

Let $t = 1, \dots, T$ index days in the billing cycle. Each day t has a supply state $\sigma_t \in \{0, 1\}$, where $\sigma_t = 1$ if the pipe is pressurized and $\sigma_t = 0$ otherwise. The total number of on-days is $n \equiv \sum_{t=1}^T \sigma_t$.

The economy is populated with a large number of N heterogeneous households indexed by $i \in \{1, 2, \dots, N\}$. In the first stage, they decide how much to invest in storage before observing their idiosyncratic preference shocks. The economic logic is that storage infrastructure is lumpy and durable; an agent installs it once and lives with it across many billing cycles. Because the preference intensity varies across households and is correlated with income, optimal storage K_i^* also varies across households, with higher-demand households investing in larger tanks.

In the second stage, these households demand water for on- and off-period consumption after realizing their demand shocks. Each household observes the realization of its idiosyncratic shock ω_i prior to making its consumption decision. I assume that N is sufficiently large that mean-zero transitory shocks cancel out in aggregation.

There is a public utility that supplies water. Throughout the discussion, I assume the aquifer is scarce, meaning that if water were priced strictly at pumping cost c , demand would exhaust the sustainable yield ($\sum q_i(p = c) > S$). This assumption describes the context and many regions around the world and implies a strictly positive capacity rent. I consider a water market in which the utility supplies water at a price p , which I allow to vary by regime. The timing of the model is illustrated in the schematic below.



Preferences and Demand

Household i chooses water consumption sequence $\{q_{i,t}\}_{t=1}^T$ to maximize a quasilinear utility over an instantaneous utility and an aggregate numéraire:

$$\begin{aligned}
 U_i &= \underbrace{\sum_{t=1}^T u_i(q_{i,t})}_{\equiv \text{GU}_i \text{ (gross water utility)}} + \underbrace{c_i}_{\text{numéraire}} \\
 &\quad (2.1)
 \end{aligned}$$

subject to budget constraint:

$$y_i = p \sum_{t=1}^T q_{i,t} + c_i + \Gamma_i(K_i) \quad (2.2)$$

and the storage law of motion:

$$I_{i,t+1} = \min\{I_{i,t} + \phi \sigma_t - q_{i,t}, K_i\}, \quad I_{i,t} \geq 0, \quad I_{i,1} = 0 \quad (2.3)$$

and the feasibility constraint:

$$0 \leq q_{i,t} \leq \phi \sigma_t + I_{i,t} \quad (2.4)$$

where ϕ is the daily flow rate when the pipe is open, $I_{i,t}$ is the water stored in the tank at the start of day t , and K_i is the household's tank capacity in gallons. In order to make sure that I have an interior solution for the storage investment decision, I make this assumption:

Assumption 1 (Storage Cost Technology). The annualized storage cost function $\Gamma_i : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ satisfies:

1. $\Gamma_i(0) = 0$ (no cost if no storage is installed);
2. $\Gamma'_i(K) > 0$ for all $K > 0$ (strictly increasing marginal cost);
3. $\Gamma''_i(K) \geq 0$ (weakly convex).

On an on-day ($\sigma_t = 1$), the household draws water from the pressurized pipe at rate $q_{i,t} \leq \phi$; any excess flow $\phi - q_{i,t} > 0$ accumulates in the storage tank. On an off-day ($\sigma_t = 0$), the pipe is depressurized and the household consumes exclusively from stored water. Thus, on-days serve a dual purpose: direct consumption and inventory accumulation for the upcoming off-period. To simplify the dynamic model, I make the following assumption, which matches the institutional setting, where each zone receives water in a single continuous window per billing cycle:

Assumption 2 (Contiguous Supply). All on-days are scheduled consecutively: $\sigma_t = 1$ for $t = 1, \dots, n$ and $\sigma_t = 0$ for $t = n + 1, \dots, T$.

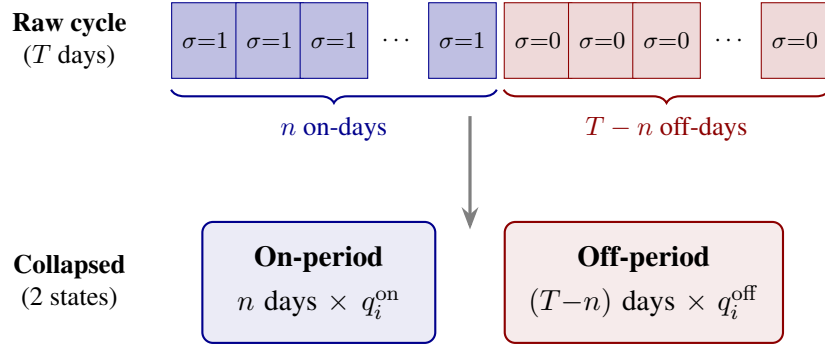
This assumption simplifies the finite-time problem to a static optimization. The following proposition makes this idea formally:

Proposition 1 (Within-State Equalization). *Suppose u_i is strictly concave, the supply schedule is non-stochastic, and on-days are contiguous (Assumption 2). Then at the optimum, $q_{i,t} = q_{i,t'}$ for any two on-days t, t' , and $q_{i,t} = q_{i,t'}$ for any two off-days t, t' .*

Proof. See Appendix A.3. □

The main idea of the proposition is that we can simplify a complex dynamic problem into a two-period model with “on” and “off” periods, as detailed below:

This setup rules out differences in sequencing: two zones with the same total n but different spacing of on-days are treated identically, even though split schedules would reduce the maximum off-spell length and thus the storage requirement. In practice, only 2 zones (less than 1%) receive water in split windows; all others receive a single contiguous block.



Second Stage: Consumption and Storage Allocation

Second-Stage Household Problem. Given pre-installed tank capacity K_i and the realization of ω_i , the household chooses the daily consumption rates in each state ($q_i^{\text{on}}, q_i^{\text{off}}$) to maximize:

$$\max_{q_i^{\text{on}}, q_i^{\text{off}}} U_i = \text{GU}_i + \underbrace{y_i - p \cdot n \cdot q_i^{\text{on}} - p \cdot (T-n) \cdot q_i^{\text{off}} - \Gamma_i(K_i)}_{\text{numéraire } c_i}$$

where $\text{GU}_i = n \cdot u_i(q_i^{\text{on}}) + (T-n) \cdot u_i(q_i^{\text{off}})$ is gross utility, subject to the total volume definition $q_i \equiv nq_i^{\text{on}} + (T-n)q_i^{\text{off}}$ and the physical constraints:

$$\begin{aligned} q_i &\leq \phi \cdot n && \text{(Pipe)} \\ (T-n)q_i^{\text{off}} &\leq K_i && \text{(Tank)} \\ q_i^{\text{on}}, q_i^{\text{off}} &\geq 0 && \text{(Feasibility)} \end{aligned}$$

Thus, this stage's problem is a household optimization problem with two constraints: the first says that the household cannot consume in total more than what the pipe delivers, and the second says that off-period consumption cannot exceed the tank capacity.

First Stage: Storage Investment

Before introducing the first stage problem, it is useful to define consumer surplus and the value function, which will be used for the welfare analysis.

Definition 1 (Second-Stage Consumer Surplus). For a household with installed tank capacity K_i facing preference shock realization ω_i , the second-stage consumer surplus $W_i(K_i, \omega_i)$ is the maximized value of the consumption problem, holding storage fixed:

$$W_i(K_i, \omega_i) \equiv \max_{q_i^{\text{on}}, q_i^{\text{off}}} \text{GU}_i - p q_i \quad (2.5)$$

subject to the tank, pipe, and feasibility constraints. Thus W_i measures realized consumer surplus from water: it is net of water expenditure $p q_i$, but gross of storage costs $\Gamma_i(K_i)$, which are treated as sunk in the second stage.²

²Note that $\text{GU}_i - p q_i$ is the standard consumer surplus concept under quasilinear preferences where income effects are absent and utility is transferable through the numeraire.

Definition 2 (Ex Ante Value Function). The ex ante value function integrates the expected second-stage consumer surplus over the distribution of preference shocks and subtracts the annualized cost of storage capital:

$$V_i(K_i) \equiv \int_{-\infty}^{\infty} W_i(K_i, \omega) dF(\omega) - \underbrace{\Gamma_i(K_i)}_{\text{storage cost}} \quad (2.6)$$

where $F(\cdot)$ is the CDF of $\omega \sim \mathcal{N}(0, \sigma_\omega^2)$, and $\Gamma_i(K_i)$ satisfies Assumption 1.

First-Stage Household Problem. In the first stage, the household chooses its storage capacity K_i to maximize $V_i(K_i)$ *before* observing their preference shock ω_i . Because the second-stage consumer surplus W_i is concave in K_i (the marginal value of storage is decreasing in each regime) and the storage cost Γ_i is convex, the objective function V_i is globally concave. This guarantees a unique optimal storage investment K_i^* . The household's problem is:³

$$\max_{K_i \geq 0} V_i(K_i) \equiv \mathbb{E}_\omega [W_i(K_i, \omega)] - \Gamma_i(K_i) \quad (2.7)$$

Finally, I define another key object, the value of having an additional supply day:

Definition 3 (Private Marginal Value of an On-Day). For a household with optimal second-stage allocation $(q_i^{\text{on}}, q_i^{\text{off}})$ under supply schedule n , define the private marginal value of one additional on-day as:

$$\pi_i(n) \equiv \left. \frac{dV_i}{dn} \right|_{K_i} = \mathbb{E}_\omega \left[\underbrace{L_i(\omega)}_{\text{smoothing}} + \underbrace{\mu_{\phi,i}(\omega) \cdot \phi}_{\text{quantity}} \right] \quad (2.8)$$

where L_i is the smoothing loss (defined later in equation (2.16)), $\mu_{\phi,i} \geq 0$ is the shadow price of the pipe constraint, and ϕ is the daily network flow rate, which I treat as a constant engineering parameter. Equation (2.8) follows from the envelope theorem applied to the household's second-stage Lagrangian (see Appendix A.3). This object plays a central role in the planner's problem that balances the costs and benefits of supply duration.

Comparative Statics

With the demand side of the model now fully specified, we can assess how optimal storage responds to supply days and prices, which will validate the empirical welfare quantification.

Proposition 2 (Effect of Duration on Storage). *For a household with interior optimal storage, $dK_i^*/dn < 0$: more frequent supply reduces storage investment.*

³To see this expression mathematically, note that the household's total utility is $U_i = GU_i + c_i$. Substituting back the budget constraint $y_i = pq_i + c_i + \Gamma_i(K_i)$ gives indirect utility $U_i^* = (GU_i^* - pq_i) - \Gamma_i(K_i) + y_i$. Because the grouped term is exactly the Second-Stage Consumer Surplus (W_i), total indirect utility simplifies to $U_i^* = W_i(K_i, \omega_i) - \Gamma_i(K_i) + y_i$. Since y_i is exogenous, maximizing this expression with respect to K_i is identical to maximizing expected consumer surplus net of storage costs.

Proof. See Appendix A.3. □

Proposition 3 (Effect of Price on Storage). *For a household with interior optimal storage, $dK_i^*/dp < 0$: higher prices reduce storage investment.*

Proof. See Appendix A.3. □

The intuition is that a higher price reduces desired consumption, so the household needs less storage capacity to smooth a smaller quantity across on- and off-days.

Corollary 1 (Storage Under Continuous Supply). *At $n = T$, $K_i^* = 0$ for all households. Under continuous supply, no household invests in storage.*

Proof. See Appendix A.3. □

Supply and Cost

I assume total water supply is fixed at $S > 0$. Supply is subject to two distinct physical constraints, each generating a component of the shadow price. The first constraint is scarcity (ν), which is the intertemporal stock constraint (see Appendix A.3). It represents the opportunity cost of depleting the aquifer, or the marginal user cost, as extraction rates exceed the estimated safe yield. I treat ν as an exogenous parameter component of the social marginal cost. The second constraint is capacity, where aggregate demand cannot exceed the pipe capacity S : $\sum q_i < S$. The multiplier on this constraint, $\mu \geq 0$, represents the congestion cost.

Assumption 3 (Subsidized Pricing and Iceberg Costs). To capture network inefficiency (physical leakage), I define the *iceberg cost multiplier* $\iota(n) \geq 1$. To deliver a fixed net quantity Q to households, the utility must produce a gross quantity $\iota(n) \cdot Q$.

I assume $\iota(n)$ is strictly increasing in supply duration n ($\iota'(n) > 0$), reflecting that longer pressurization durations increase physical water losses. This assumption is supported by engineering models; for example, Taylor et al. (2018) model leakage as increasing in supply duration for a given pipe condition. In Amman, the evidence supports the rising leakage in pressurization time. Recent data confirm the pattern: when Miyahuna extended supply from 36 to 38 hours per week, NRW rose from 42.8% to 44.7% (Al-Rkebat, 2026). The idea is that on degraded infrastructure, longer pressurization pushes more water through deteriorated joints and aging mains, so that gross extraction must rise to deliver the same net quantity. Without this assumption, extending supply would be costless and the planner would always prefer continuous service regardless of network condition, eliminating the tradeoff between service reliability and physical losses that is central to the policy problem.

The total cost function is:

$$C(Q, n) = \underbrace{c}_{\text{Marginal Cost}} \cdot \underbrace{\iota(n) \cdot Q}_{\text{Gross Production}} + \underbrace{F}_{\text{Fixed Cost}} \quad (2.9)$$

The iceberg cost $\iota(n)$ acts as a fiscal multiplier. Because ι is a function of supply duration rather than throughput, increasing supply days n raises the leakage multiplier. This implies that to deliver the *same* fixed quantity Q , the utility must extract and treat significantly more water. This inflates the total cost C , which in turn increases the required fiscal subsidy, creating a vicious cycle where higher supply frequency targets exacerbate the deficit.

This cost structure also formally establishes the utility as a natural monopoly: it is strictly more efficient for a single entity to supply the aggregate quantity Q than to fragment production across multiple agents (e.g., households producing longer supply via private storage). Given the significant fixed infrastructure costs ($F > 0$) and constant marginal costs, the average cost function $AC = C/Q$ is strictly decreasing in output Q . Following Baumol (1977), strictly declining average cost is a sufficient condition for the strict subadditivity of the cost function in the single-product context, implying the utility is a natural monopoly.

Social Welfare Weights

A central question in water policy is how to balance allocative efficiency against the distributional goal of ensuring access for low-income households. The Second Welfare Theorem guarantees that this tradeoff is illusory when lump-sum transfers are available: the planner sets p to match the social marginal cost and redistributes through transfers. In the absence of distribution instruments, the planner must use the price of water itself as a redistribution device, accepting allocative distortions for distributional gains. To formalize this tradeoff, I introduce welfare weights into the planner's objective.

Definition 4 (Social Marginal Welfare Weight). Following the standard public finance approach to modeling redistributive preferences (e.g., Saez (2002); Saez and Stantcheva (2016)), I define the social marginal welfare weight $g_i(y_i)$ as the planner's valuation of a marginal unit of income transferred to household i .

Under our quasilinear utility specification, the private marginal utility of income is unity. Therefore, the weight g_i is driven entirely by the planner's redistributive tastes over baseline income y_i . I normalize these weights (similar to the the marginal value of public funds literature) such that the population average is exactly one: $\frac{1}{N} \sum_{i=1}^N g_i(y_i) = 1$. This normalization ensures that $g_i(y_i)$ measures the social value of transferring \$1 to household i relative to distributing that dollar uniformly across the population. Two summary statistics of the weight distribution play a central role in optimal policy:

Definition 5 (Consumption-Weighted Social Welfare Weight).

$$\bar{g}_q \equiv \frac{\sum_{i=1}^N g_i \mathbb{E}[q_i]}{\sum_{i=1}^N \mathbb{E}[q_i]} \quad (2.10)$$

Definition 6 (Duration-Benefit-Weighted Social Welfare Weight).

$$\bar{g}_\pi \equiv \frac{\sum_{i=1}^N g_i \pi_i}{\sum_{i=1}^N \pi_i} \quad (2.11)$$

The two summary weights arise because price and supply duration affect different margins of household welfare, and the households who benefit most from each instrument are generally different. The first weight $\bar{g}_q$ governs the optimal price: a price change affects consumer surplus proportionally to consumption volume, so the relevant average weight is consumption-weighted. On the other hand, $\bar{g}_\pi$ governs optimal supply duration: a supply frequency increase benefits households proportionally to π_i , which is large for households whose existing capacity cannot smooth demand across off-days when evaluated at the current storage allocation $K_i^*(n)$.

Two features of $\bar{g}_\pi$ deserve emphasis. First, $\bar{g}_\pi$ is not a primitive: π_i is positive only for households whose tank currently binds, and which households are tank-constrained depends on the installed storage K_i^* . Second, K_i^* is itself decreasing in n (Proposition 2), so as the planner extends supply duration, households shed storage and the set of tank-constrained households shrinks. The value of $\bar{g}_\pi$ therefore moves with n : the summary weight is defined at a specific storage allocation, and the planner's optimality condition jointly determines n^* and the equilibrium $\{K_i^*\}$ that pins down $\bar{g}_\pi$ at the optimum.

This storage-adjustment channel also has a direct fiscal implication that enters the planner's FOC separately. When households shed storage in response to higher n , they lose the ability to draw subsidized water during off-days, which reduces aggregate subsidized consumption. Each unit of shed storage therefore saves the utility money, partially offsetting the fiscal wedge from extending supply, which is included in the planner's FOC as an adaptation term. The empirical values reported in Section 2.7 are evaluated at the status quo storage allocation $K_i^*(n^{SQ})$.

Remark 1 (Sign of the Summary Weights). The summary weights $\bar{g}_q$ and $\bar{g}_\pi$ can be decomposed as:

$$\bar{g}_q = 1 + \frac{\text{cov}(g_i, \mathbb{E}[q_i])}{\bar{g} \cdot \bar{q}}, \quad \bar{g}_\pi = 1 + \frac{\text{cov}(g_i, \pi_i)}{\bar{g} \cdot \bar{\pi}}.$$

The signs of these covariances are ambiguous in general. If water is a normal good (absolute consumption q_i increasing in income), the welfare weight g_i is decreasing in income for $\eta > 0$, so $\text{cov}(g_i, q_i) < 0$ and $\bar{g}_q < 1$. However, the magnitude depends on how steeply consumption rises with income.

If the volumetric Engel curve is approximately flat as the data suggest (Fact #2), the covariance is small and $\bar{g}_q$ is close to unity. The *redistributive welfare gain* from the price instrument is therefore limited, since it scales with $1 - \bar{g}_q$. The main welfare gain from pricing reform in this setting accordingly operates through the revenue raised and redistributed, rather than through the price change itself being progressive. Similarly, $\bar{g}_\pi$ depends on whether high-income households benefit more from additional supply days. If wealthier households have higher demand and thus bind more frequently against physical constraints, π_i is increasing in y_i , pushing $\text{cov}(g_i, \pi_i) < 0$ and $\bar{g}_\pi < 1$.

The Regulator's Problem

The planner maximizes aggregate welfare, accounting for the temporal smoothing structure. The instruments are the price p , a uniform quota $\bar{q}$, and supply duration n . The planner's objective is:⁴

$$\max_{p, \bar{q}, n} \mathcal{W}(\Theta) = \sum_i g_i V_i + (p - \Lambda) \sum_{i=1}^N \mathbb{E}[q_i] - F \quad (2.12)$$

subject to the aggregate resource constraint:

$$\iota(n) \cdot \sum_{i=1}^N \mathbb{E}[q_i] \leq S \quad (\text{multiplier } \mu) \quad (2.13)$$

where $V_i = \mathbb{E}[\text{GU}_i - p q_i] - \Gamma_i(K_i^*)$ is consumer surplus net of storage costs and $\text{GU}_i = n u_i(q_i^{\text{on}}) + (T - n) u_i(q_i^{\text{off}})$ is gross utility, and the full marginal cost $\Lambda = (c + \lambda)\iota(n)$, with λ is the scarcity rent.⁵ The cost term $\iota(n)$ reflects the fact that the aquifer depletes by the *gross* extracted quantity $\iota(n) \sum q_i$: both pumping cost c and scarcity rent ν are incurred per unit of gross extraction, not per unit of net delivery. Observe that the planner takes into account K_i^* , which is the household response to the policy choice.

In addition to the status quo that uses observed price and water supply days, I present three distinct counterfactual regimes based on the available policy instruments, where $\bar{q} = \infty$ denotes the absence of a binding quota:

1. **Status Quo:** $\Theta^{SQ} = (\bar{p}, \infty, n^{SQ})$, where $\bar{p}$ is the observed administered price and n^{SQ} is the observed median supply duration. No quota is enforced and no instrument is optimized; this regime serves as the baseline for all welfare comparisons.
2. **First Best:** $\Theta^* = (p^*, \infty, n^*)$, characterized by social-marginal-cost pricing and optimal supply duration. I evaluate this regime under two iceberg schedules, $\iota(n; \delta_t)$ and $\iota(n; \delta_r)$, corresponding to the degraded status quo network and a fully rehabilitated network respectively.
3. **Second Best:** $\Theta^{SB} = (\bar{p}, \bar{q}^*, n^{SB})$, where the planner faces a fixed price ceiling $\bar{p}$ and uses both uniform quotas and supply duration as rationing instruments.

⁴This is equivalent to maximizing gross utility directly:

$$\max_{p, \bar{q}, n} \mathcal{W}(\Theta) = \sum_{i=1}^N g_i [\mathbb{E}[\text{GU}_i] - \Gamma_i(K_i^*)] + \sum_i (1 - g_i) p \mathbb{E}[q_i] - (c + \nu) \iota(n) \sum_{i=1}^N \mathbb{E}[q_i] - F$$

⁵This cost is $\lambda = \nu + \mu$, where $\mu \geq 0$ is the capacity rent and the scarcity rent ν is derived from the aquifer's Bellman equation (Appendix A.3), where the state transition is $S_{t+1} = S_t - Q_t^{\text{gross}}$. Because leaked water is irreversibly lost from the aquifer, ν applies to every unit pumped, including leakage.

4. **Third Best:** $\Theta^{TB} = (\bar{p}, \infty, n^{TB})$, where the only available instrument is supply duration since prices are held fixed and there is no quota enforced.

In all three regimes, the planner's instrument set is limited to $(p, \bar{q}, n)$; lump-sum transfers are not available as a separate policy tool. In Section 2.7, I separately examine distributional incidence under alternative assumptions about how net revenue is rebated to households, but these transfer rules are not part of the planner's optimization.

First-Best Allocation

First, observe that a planner that can price at the social marginal cost would never choose a rigid quota $\bar{q}$, so the relevant choice variables for the planner are p and n . The policy space is therefore $\Theta^* = (p^*, \infty, n^*)$.

Theorem 1 (First-Best Allocation). *The socially optimal policy (p^*, n^*) satisfies:*

(i) **Optimal Price.**

$$p^* = \frac{\Lambda^*}{1 + \frac{\bar{g}_q - 1}{|\varepsilon_{Q,p}|}} \quad (2.14)$$

where $\Lambda^* \equiv (c + \lambda) \iota(n^*)$ is the social marginal cost, $\bar{g}_q$ is the consumption-weighted welfare weight, and $\varepsilon_{Q,p} < 0$ is the aggregate price elasticity of demand.

(ii) **Optimal Supply Duration.**

$$\underbrace{\sum_{i=1}^N g_i \pi_i(n^*)}_{\text{Weighted Marginal Benefit}} = \underbrace{(c + \lambda) \iota'(n^*) \sum_{i=1}^N \mathbb{E}[q_i]}_{\text{Leakage Cost}} + \underbrace{\Delta_p^* \sum_{i=1}^N \mathbb{E}\left[\frac{dq_i}{dn}\right]}_{\text{Fiscal Wedge}} \quad (2.15)$$

where $\Delta_p^* \equiv \Lambda^* - p^*$ and $\pi_i = \mathbb{E}[L_i + \mu_{\phi,i} \phi]$ is the private marginal value of an on-day (Definition 3).

Proof. See Appendix A.3. □

Equation (2.14) reveals that when $\bar{g}_q = 1$, $p^* = \Lambda^*$: the optimal price equals the social marginal cost (inclusive of the scarcity rent λ). There is also an equity motive for price distortion. When $\bar{g}_q \neq 1$, $p^* \neq \Lambda^*$: the planner distorts the price as a redistribution instrument. The direction depends on $\bar{g}_q - 1$: If $\bar{g}_q > 1$, $p^* < \Lambda^*$: the planner subsidizes water. If $\bar{g}_q < 1$, $p^* > \Lambda^*$: the planner taxes water above marginal cost. The magnitude of the distortion $|p^* - \Lambda^*|$ is inversely proportional to $|\varepsilon_{Q,p}|$: when demand is inelastic, the Harberger triangle is small relative to the distributional gain (or loss avoided), so the planner exploits the price instrument aggressively.⁶

⁶Overall, the pricing rule (2.14) follows the same economic logic as Ramsey (1927) and Feldstein (1972) in that the optimal departure from marginal-cost pricing reflects both an inverse-elasticity efficiency term and a distributional

As for equation (2.15), the left-hand side sums household-level benefits from an additional on-day: Pipe-constrained households, contributing $\mu_{\phi,i} \phi > 0$ (quantity expansion); and tank-binding households, contributing $L_i > 0$ (smoothing dividend). Unconstrained households contribute nothing ($\pi_i = 0$). For more details, see Appendix A.3.⁷ The right-hand side is the marginal leakage cost of extending supply by one day: longer pressurization raises the iceberg multiplier, increasing the gross extraction required to deliver the existing aggregate consumption. The planner balances these: extend supply duration until the aggregate marginal benefit equals the marginal leakage cost, plus a fiscal wedge.

Second-Best Allocation

Now suppose the planner faces some friction to pricing but can choose a quantity cap on consumption, making the policy space in this case $\Theta^{SB} = (\bar{p}, \bar{q}^*, n^{SB})$.

Theorem 2 (Second-Best Rationing). *Under a fixed, subsidized price $\bar{p} < p^*$ with instruments $(\bar{q}, n)$, the optimal policy satisfies:*

(i) *Optimal Uniform Quota.*

$$\underbrace{\frac{1}{|\Omega_{\bar{q}}|} \sum_{i \in \Omega_{\bar{q}}} g_i \mathbb{E}[u'_i(\bar{q}^*/T) - \bar{p}]}_{\text{Weighted Net Marginal Value}} = \underbrace{\Lambda^{SB} - \bar{p}}_{\text{Fiscal Cost per Unit}} \quad (2.17)$$

where $\Lambda^{SB} \equiv (c + \lambda)\iota(n^{SB})$ is the social marginal cost and $\Omega_{\bar{q}}$ is the set of quota-constrained households.

(ii) *Optimal Supply Duration.*

$$\underbrace{\sum_{i=1}^N g_i \pi_i(n^{SB}) + \mathcal{S}(n^{SB})}_{\text{Effective Marginal Benefit}} = \underbrace{(c + \lambda) \iota'(n^{SB}) \sum_{i=1}^N \mathbb{E}[q_i]}_{\text{Leakage Cost}} + \underbrace{\Delta_p \sum_{i=1}^N \mathbb{E}\left[\frac{dq_i}{dn} \Big|_{K_i}\right]}_{\text{Fiscal Wedge}} \quad (2.18)$$

where: $\Delta_p \equiv \Lambda^{SB} - \bar{p}$ and $\mathcal{S}(n) \equiv -\Delta_p \sum_{i=1}^N \mathbb{P}(i \in \Omega_K) \cdot \frac{dK_i^*}{dn} > 0$. The adaptation correction is positive because $dK_i^*/dn < 0$ (Proposition 2): more frequent supply induces disinvestment in

term. In a way, the current model is the single-good, resource-constrained analogue of their multi-good, budget-constrained results, extended to incorporate a supply duration instrument.

⁷The smoothing loss for household i is defined as

$$L_i \equiv \left[u_i(q_i^{\text{on}}) - u'_i(q_i^{\text{on}}) q_i^{\text{on}} \right] - \left[u_i(q_i^{\text{off}}) - u'_i(q_i^{\text{off}}) q_i^{\text{off}} \right] \geq 0 \quad (2.16)$$

which is zero if and only if $q_i^{\text{on}} = q_i^{\text{off}}$ and strictly positive when the tank binds. This expression is formally derived Appendix A.3.

storage, which reduces subsidized consumption. The correction is weighted by $\mathbb{P}(i \in \Omega_K)$, the probability that a household is storage-constrained.

Proof. See Appendix A.3. □

Equation (2.17) shows that the planner sets the quota so that the *g-weighted average net marginal value* of water (marginal utility minus price) among constrained households equals the fiscal cost per unit of delivering subsidized water. When g_i varies across quota-constrained households, the optimal quota tilts toward the preferences of high-weight (low-income) households in this group.

Equation (2.18) show how supply days choice balances cost with benefits. The right-hand side is the social marginal cost of extending supply by one day: longer pressurization raises the iceberg multiplier $\iota(n)$, so gross extraction must rise to deliver the same aggregate consumption. The second term, $\Delta_p \sum_i \mathbb{E}[dq_i/dn|_{K_i}]$, is the *fiscal wedge*: holding storage fixed, extending supply directly raises subsidized consumption among pipe- and tank-constrained households, each unit of which costs the utility Δ_p . The left-hand side collects the benefits. The first term, $\sum_i g_i \pi_i(n^{SB})$, aggregates household-level gains from an additional on-day: smoothing benefits L_i and quantity-expansion benefits $\mu_{\phi,i} \phi$, weighted by the planner's redistributive preferences. The second term, $\mathcal{S}(n^{SB})$, is the *adaptation correction*: when n rises, households shed storage (Proposition 2), and each unit of storage shed eliminates a corresponding unit of subsidized consumption, saving the utility Δ_p per unit.

In the numerical counterfactuals (Section 2.7), K_i^* is re-solved from the household FOC at each candidate policy $(p, n, \bar{q})$, so the adaptation correction is fully endogenous. The net effect is that $n^{SB} < n^*$: the planner provides a lower supply frequency than the first-best.⁸

Third-Best Allocation

Now consider a planner that can neither change prices nor implement a quota. In this case, $\Theta^{TB} = (\bar{p}, \infty, n^{TB})$.

Proposition 4 (Third-Best Allocation). *Under $\bar{q} = \infty$ (no quota enforcement), the optimal supply duration n^{TB} satisfies:*

⁸The Second-Best and Third-Best FOCs condition dq_i/dn on fixed storage K_i and include the separate adaptation correction $\mathcal{S}(n)$, whereas the First-Best FOC (2.15) uses the unconditional total derivative and has no $\mathcal{S}$ term. The reason is that the subsidized price $\bar{p} < \Lambda$ drives a wedge between the private and social value of storage: the household's FOC $\mathbb{E}[\mu_K] = \Gamma'(K^*)$ is privately optimal but socially suboptimal, because each unit of storage-enabled consumption generates a fiscal cost Δ_p that the household ignores. When the planner varies n and households adjust K^* in response, this externality produces a welfare effect $\partial \mathcal{W} / \partial K_i = -\Delta_p \cdot \mathbb{P}(i \in \Omega_K) \neq 0$, which after summing over households and multiplying by dK_i^*/dn yields exactly $\mathcal{S}(n)$. In the First-Best, the planner controls the price, so p^* already internalizes the fiscal cost of storage-enabled consumption, the household's storage FOC is therefore socially optimal given p^* , and the indirect welfare effect of storage adjustment vanishes: $\partial \mathcal{W} / \partial K_i|_{K_i^*} = 0$.

$$\underbrace{\sum_{i=1}^N g_i \pi_i(n^{TB}) + \mathcal{S}(n^{TB})}_{\text{Effective Marginal Benefit}} = \underbrace{(c + \lambda) \iota'(n^{TB}) \sum_i \mathbb{E}[q_i]}_{\text{Leakage Cost}} + \underbrace{\Delta'_p \sum_{i=1}^N \mathbb{E} \left[\frac{dq_i}{dn} \Big|_{K_i} \right]}_{\text{Fiscal Wedge}} \quad (2.19)$$

where $\Delta'_p \equiv \Lambda^{TB} - \bar{p}$ and $\mathcal{S}(n)$ is defined analogously to the Second-Best case. The structure is identical to the Second-Best FOC (2.18), but the constrained sets are strictly larger because the absence of quotas leaves high-demand households in the physically-constrained sets rather than capping them at $\bar{q}$. Consequently, the fiscal wedge is strictly larger in the Third-Best, yielding the ordering $n^* > n^{SB} > n^{TB}$.

Proof. See Appendix A.3. □

In the Second-Best, the quota removes the highest-demand households from both the storage- and pipe-constrained sets: their consumption is capped at $\bar{q}^*$ and no longer responds to n . In the Third-Best, these same households remain in the physically-constrained sets, each contributing to the fiscal wedge. Because the wedge enters the FOC as a cost of extending supply, a larger wedge pushes the planner toward a lower n , deepening intermittency relative to the Second-Best.

Summary

For compactness, I collect the primitives, define the decentralized household response to a given policy, and state the planner's problem over that policy.

Definition 7 (Primitives). The economy is characterized by:

$$\mathcal{E} \equiv \left\langle \eta, N, T, S, \phi, c, \nu, \delta, F, \bar{p}, \{a_i, b, \Gamma_i(\cdot), y_i\}_{i=1}^N, \sigma_\zeta^2, \sigma_\xi^2 \right\rangle$$

The components are: the planner's inequality aversion parameter η , which pins down welfare weights $g_i = y_i^{-\eta} / \bar{y}^{-\eta}$; the number of days in a billing cycle T ; the number of households N ; the aggregate pipe capacity S ; the daily network flow rate ϕ ; the marginal extraction cost c ; the scarcity rent ν ; the network decay parameter δ , which pins down the iceberg multiplier $\iota(n; \delta)$ (Assumption 6); the fixed cost of supply F ; the administered price $\bar{p}$ (relevant for the Second- and Third-Best regimes); household-specific preference intercepts a_i , the utility curvature parameter b , and storage cost technologies Γ_i (Assumption 1); household incomes y_i ; and the permanent and transitory components of the demand-space preference shock, σ_ζ^2 and σ_ξ^2 (equivalently, $\sigma_\varepsilon^2 = \sigma_\zeta^2 + \sigma_\xi^2$).

Definition 8 (Competitive Equilibrium). For a fixed policy vector $\Theta = (p, \bar{q}, n)$, a *Competitive Equilibrium* is a collection of investment and consumption plans $\{K_i^*, q_i^*(\omega)\}_{i=1}^N$ such that:

1. **Household Optimality.** Each household solves its two-stage problem (Section 2.4) taking Θ as given: K_i^* maximizes the ex-ante value function $V_i(K_i)$ before ω_i is realized, and $q_i^*(\omega_i)$ maximizes flow utility subject to the pipe, tank, and quota constraints.
2. **Resource Feasibility.** $\iota(n) \cdot \sum_{i=1}^N \mathbb{E}_\omega[q_i^*(\omega)] \leq S$.

The CE defines a mapping $\Theta \mapsto \{K_i^*(\Theta), q_i^*(\Theta, \omega)\}$: for any policy the government announces, households respond optimally. Nothing in the CE requires that Θ is itself optimal from society's perspective.

Definition 9 (Social Optimum). A *Social Optimum* is a policy vector Θ^* that solves the planner's problem:

$$\max_{p, \bar{q}, n} \mathcal{W}(\Theta) = \sum_{i=1}^N g_i V_i(\Theta) + (p - \Lambda(\Theta)) Q(\Theta) - F \quad (2.20)$$

subject to resource feasibility, where $V_i(\Theta)$ and $Q(\Theta) \equiv \sum_i \mathbb{E}[q_i^*(\Theta, \omega)]$ are determined by the Competitive Equilibrium associated with Θ . The planner moves first (Stage 0), anticipating the decentralized response. The Social Optimum is regime-specific: the First-Best optimizes over (p, n) with $\bar{q} = \infty$; the Second-Best optimizes over $(\bar{q}, n)$ at fixed $\bar{p}$; and the Third-Best optimizes over n alone.

Why the Two Diverge. In the Competitive Equilibrium, each household takes the policy Θ as given and maximizes its own surplus. In the Social Optimum, the planner chooses Θ to maximize aggregate weighted welfare $\sum g_i V_i$, where the weights g_i encode redistributive preferences. When $g_i = 1$ for all i and $p = \Lambda$, the private and social first-order conditions coincide and no household decision generates an externality. Two forces break this alignment. First, heterogeneous welfare weights ($g_i \neq 1$) create a motive to distort the price away from Λ as a redistribution device, so the planner's optimal price p^* generally differs from the marginal-cost benchmark Λ that would obtain under uniform weights. Second, under the status quo the administered price $\bar{p} < \Lambda$ creates a subsidy wedge that distorts both the consumption and storage margins. On the consumption margin, each unit of water costs the system Λ but the household pays only $\bar{p}$, so the fiscal gap must be financed externally. The model does not impose a government budget constraint: the deficit $(p - \Lambda)Q$ enters the social welfare function as a direct cost, so the planner tolerates fiscal losses when the consumer surplus gain outweighs them. This is equivalent to assuming that external financing (in practice, concessional donor loans and arrears to the electricity utility, as described in Section 2.2) is available at a unit cost per dollar.⁹ On the storage margin, privately optimal tank investment enables additional subsidized consumption, generating a fiscal cost that the household's FOC for K_i^* ignores. The planner accounts for both channels when choosing n , which is why the adaptation correction $\mathcal{S}(n)$ appears in the Second-Best and Third-Best FOCs (Theorems 2 and 4) but not in the First-Best, where the planner controls p directly and can eliminate the wedge.

⁹If the true marginal cost of public funds exceeds unity, the welfare cost of the subsidy is larger and the gains from pricing reform correspondingly higher.

Aggregate Welfare

I evaluate the welfare cost of the intermittent status quo by comparing it against two counterfactual regimes. In both cases, the departure point is the Status Quo: $\Theta^{SQ} = (\bar{p}, \infty, n^{SQ})$. The three counterfactuals are:

1. **Optimal intermittency** (Status Quo $\rightarrow$ Third-Best): $\Theta^{TB} = (\bar{p}, \infty, n^{TB})$;
2. **Introducing quotas** (Status Quo $\rightarrow$ Second-Best): $\Theta^{SB} = (\bar{p}, \bar{q}^*, n^{SB})$;
3. **Market-clearing pricing** (Status Quo $\rightarrow$ First-Best): $\Theta^* = (p^*, \infty, n^*)$.

Regime-Specific Costs. The three counterfactuals involve different infrastructure configurations. The Third-Best and the degraded First-Best retain the status quo network and fixed cost F_t , isolating the gains from optimizing the instruments on the existing infrastructure. The rehabilitated First-Best incurs the additional amortized capital of pipe replacement ($F_r > F_t$) but operates against the lower leakage schedule $\iota(n; \delta_r)$. The Second-Best retains δ_t but adds the smart-metering cost required to enforce household-level quotas ($F_q > F_t$). The scarcity rent ν is common across regimes, since it reflects the long-run aquifer shadow value independent of the delivery technology. Table 2.3 reports the calibrated values.

Theorem 3 (Welfare Gains from Counterfactual Reforms). *For any counterfactual regime $r \in \{TB, SB, FB\}$ with policy $\Theta^r = (p^r, \bar{q}^r, n^r)$, the welfare gain relative to the status quo decomposes as:*

$$\begin{aligned} \Delta \mathcal{W}^{SQ \rightarrow r} = & \underbrace{\sum_{i=1}^N g_i \mathbb{E}[\Delta \text{CS}_i^r]}_{\text{(I) Change in Weighted Consumer Surplus}} + \underbrace{\left[(p^r Q^r - \bar{p} Q^{SQ}) - (\Lambda^r Q^r - \Lambda^{SQ} Q^{SQ}) \right]}_{\text{(II) Government Fiscal Change}} \\ & - \underbrace{\sum_{i=1}^N g_i [\Gamma_i(K_i^r) - \Gamma_i(K_i^{SQ})]}_{\text{(III) Weighted Storage Cost Change}} - \underbrace{(F^r - F^{SQ})}_{\text{(IV) Infrastructure Cost Change}} \end{aligned} \quad (2.21)$$

where $\Delta \text{CS}_i^r \equiv (\text{GU}_i^r - p^r q_i^r) - (\text{GU}_i^{SQ} - \bar{p} q_i^{SQ})$ is the change in private consumer surplus, $Q^r \equiv \sum_i \mathbb{E}[q_i^r]$ is aggregate consumption, and $\Lambda^r \equiv (c + \lambda) \iota(n^r)$ is the social marginal cost under regime r . For the Third-Best and Second-Best, $p^r = \bar{p}$ so term (II) simplifies to $\bar{p} \Delta Q - (\Lambda^r Q^r - \Lambda^{SQ} Q^{SQ})$. For the First-Best, $p^r = p^*$ and term (II) is the net fiscal change from both the price increase and the cost change.

Proof. See Appendix A.3. □

Three features of the decomposition deserve emphasis. First, term (II) is the key channel through which pricing reforms generate aggregate welfare gains: raising p from $\bar{p}$ toward Λ converts an implicit subsidy into fiscal revenue, which the planner values at a dollar-for-dollar rate in

the baseline (and more under marginal cost of public funds greater than one). Second, term (III) is typically positive across all reforms because $K_i^r < K_i^{SQ}$: any policy that relaxes intermittency induces households to shed storage capital. This storage-cost saving is the mechanical counterpart of the “deadweight capital” result in Section 2.7. Third, term (IV) captures the amortized infrastructure cost differential: positive for the First-Best with rehabilitation ($F^{FB} > F^{SQ}$) and for the Second-Best with smart metering ($F^{SB} > F^{SQ}$), but zero for the Third-Best and the First-Best on degraded pipes, which retain the existing network. Whether $\Delta \mathcal{W}^{SQ \rightarrow r}$ is positive depends on whether terms (I)–(III) collectively outweigh the infrastructure investment in (IV).

2.5 Quasi-Experimental Evidence

Before estimating the model, I begin by presenting reduced-form evidence on the impact of supply duration on water consumption using a geographic regression discontinuity design (RDD). After that, I present the model estimation, and then validate it using the RDD.

Econometrics

Let d_i denote the perpendicular distance from household i ’s water meter to the nearest boundary separating zones with different supply intensities, normalized so that $d_i > 0$ indicates the household is on the higher-supply side and $d_i < 0$ indicates the lower-supply side. The treatment indicator is:

$$T_{j(i)} = \mathbf{1}\{S_{j(i)} > S_{j'(i)}\}$$

which equals one if household i ’s zone j has higher supply duration than the adjacent zone j' . The intuition behind the comparison is that households located just on opposite sides of zone boundaries are plausibly similar in all respects except their assigned supply schedule.

Because supply gaps vary across boundaries, I estimate two complementary specifications. The first pools all boundaries to recover a weighted-average treatment effect:

$$q_{it} = \beta_1 T_{j(i)} + \beta_2 d_i + \beta_3 (d_i \cdot T_{j(i)}) + \text{FE}_{b(i)} + \text{FE}_t + \varepsilon_{it}, \quad |d_i| \leq h \quad (2.22)$$

where β_1 captures the average discontinuity in consumption at the boundary, d_i is distance to the boundary, $\text{FE}_{b(i)}$ are border segment fixed effects (one for each distinct boundary, so that β_1 is identified from within-boundary comparisons across the two sides), and FE_t are quarter-by-year fixed effects. The unit of observation is a household-quarter, with the panel pooled across all quarters.

To allow the treatment effect to vary with the magnitude of the supply gap, I also estimate within subsamples defined by the gap $\Delta S_{b(i)} \equiv S_{j(i)} - S_{j'(i)}$ between adjacent zones at boundary b :

$$q_{it} = \beta_1^g T_{j(i)} + \beta_2^g d_i + \beta_3^g (d_i \cdot T_{j(i)}) + \text{FE}_{b(i)} + \text{FE}_t + \varepsilon_{it}^g, \quad |d_i| \leq h_g, \quad \Delta S_{b(i)} \in \mathcal{G}_g \quad (2.23)$$

where $\mathcal{G}_g$ indexes gap-magnitude bins (e.g., 1 day, 2–3 days, 4+ days) and β_1^g is the gap-specific treatment effect. This specification nests the pooled model and allows the data to reveal whether

the consumption response scales with the dosage of additional supply. Estimation uses triangular kernel weights $K(d_i/h)$ and coverage error rate-optimal bandwidth selection following Calonico et al. (2020). Standard errors are clustered at the water zone level (around 400 clusters), the level at which supply schedules are assigned.

The validity of this design rests on the continuity assumption: potential outcomes $\mathbb{E}[q_{it}(0) \mid d_i = d]$ and $\mathbb{E}[q_{it}(1) \mid d_i = d]$ must be continuous in the running variable d_i at the boundary $d_i = 0$. This requires that no household characteristic that independently affects water consumption changes discretely at the zone boundary. The assumption is plausible because zone boundaries were drawn by the utility based on network topology and historical precedent rather than on household-level demand characteristics, and because households within the bandwidth are located in the same neighborhoods, face similar housing markets, and share local amenities. In particular, water zone boundaries are determined by hydraulic infrastructure and do not coincide with other administrative boundaries (e.g., municipal districts or school zones) that might independently affect household behavior. Border segment fixed effects absorb any level differences across boundaries, so identification comes from within-boundary comparisons of households just on opposite sides of the same border.

That said, the principal threat to continuity is selective sorting: if households relocate across boundaries to obtain more favorable supply schedules, the composition of the population changes discretely at the cutoff and potential outcomes are no longer continuous. I probe this using the Cattaneo et al. (2020) density test on the running variable (Appendix A.4), which does detect excess mass on the high-supply side at very fine spatial scales. I address this by implementing a donut specification that excludes households within 20 meters of the boundary, after which the density imbalance is reduced. I provide more details on the RDD design, manipulation testing, and the donut results in Appendix A.4.

Reduced-Form Results

Before interpreting the coefficients, I rescale all treatment effects relative to the mean consumption on the low-supply side of the border, $\bar{q}^{\text{low}} = 2.60 \text{ m}^3/\text{week}$, so that coefficients can be read as percentage changes in weekly consumption. As discussed in Appendix A.4, the density test detects excess mass within 5 meters of the boundary; a 20-meter donut substantially reduces this imbalance but some of the results become more muted.

I first examine the aggregate impact of relaxing the supply constraint on household water consumption. Figure A.8(a) presents the RDD estimates of β_1 grouped by the magnitude of the supply-day discontinuity at each border. The pooled effect is $+0.13 \text{ m}^3/\text{week}$ (95% CI $[-0.02, +0.27]$), or roughly a 5% increase relative to the low-side mean. Disaggregating by gap size shows no specific pattern: the (0,1] bin gives a significant $+0.18 \text{ m}^3/\text{week}$ (+6.8%, CI $[+0.02, +0.34]$), the (1,2] bin a marginally negative -0.22 (-8.6%, CI $[-0.45, +0.01]$), the (2,3] bin a positive but imprecise $+0.32$ (+12%), and the 3+ bin a negative and insignificant -0.15 . The pooled response is therefore small on average, and the non-monotonicity across gap sizes is difficult to reconcile with a model in which additional supply days translate mechanically into additional consumption. The

heterogeneity analysis by baseline supply level below reveals the pattern that the structural model will rationalize.

Heterogeneity

Baseline Supply Level. A natural diminishing-marginal-utility prediction is that the consumption response to an additional supply day should be larger at low baselines, where households are further from satiation, than at high baselines. I test this by stratifying the RDD by the low-side supply level (the minimum of supply days across the two adjacent zones) in addition to the gap size. This differs from the gap-based heterogeneity in Figure A.8(a), which pools a 1→2 day border with a 5→6 day border into the same “1-day gap” bin, averaging over very different baselines. The level-stratified split in Figure A.8(b) instead asks whether the same-size gap produces different responses depending on where baseline supply starts. I estimate specification (2.23) separately within each of three baseline-level bins (≤ 1 , $(1, 2]$, and > 2 days), and report both the pooled-within-baseline estimate and the gap-specific estimates in each bin.

The pooled-within-baseline estimates are statistically indistinguishable from zero at baselines ≤ 1 day (-0.03 m³/week, or -1.3% of the low-side mean; CI $[-0.22, +0.15]$) and at baselines in $(1, 2]$ ($+0.08$, $+3.0\%$; CI $[-0.10, +0.26]$), but become large and statistically significant at baselines above 2 days ($+0.35$, $+13.4\%$; CI $[+0.13, +0.57]$, $p = 0.002$). This ordering is hard to reconcile with diminishing marginal utility alone, which would predict the largest response at low baselines where households are furthest from satiation. The structural model rationalizes this ordering through composition effects, as detailed in Section 2.5.

Infrastructure Interactions. I also test whether local well infrastructure, which feeds the municipal network, interacts with the central rationing schedule. I classify border pairs based on whether the local well production “reinforces” the schedule (the high-supply side has higher local well production) or “offsets” it (the low-supply side has higher production). Figure A.8(c) shows that the pooled effect is actually larger on *offset* borders ($+0.35$ m³/week, CI $[+0.11, +0.59]$) than on *reinforce* borders ($+0.09$, CI $[-0.06, +0.24]$). At the individual gap level, the strongest single coefficient comes from the $(2, 3]$ bin on *reinforce* borders ($+0.67$, CI $[+0.23, +1.12]$), but the pattern is otherwise noisy and no clear ordering emerges across bins. Overall, the larger pooled response on *offset* borders is consistent with a composition story: where local wells compensate for a tight central schedule, households on the low-supply side have adapted less (smaller tanks, lower latent demand), so an additional central supply day has a larger marginal effect. This heterogeneity is of independent interest but does not overturn the main finding that the aggregate response to supply duration is muted.

Seasonal Demand. Figure A.8(d) stratifies the RDD by season (April–September vs. October–March). The pooled estimates are nearly identical across seasons ($+0.14$ vs. $+0.12$ m³ per week) with overlapping confidence intervals, indicating no detectable seasonal asymmetry in the consumption response to supply duration.

Connection to the Model

The RDD compares households on opposite sides of a boundary, but households on each side have had time to adapt their storage to their respective supply schedules. The RDD therefore identifies the *total* derivative of consumption with respect to n , which by the chain rule includes both the direct effect of additional supply days (holding storage fixed) and the indirect effect operating through endogenous storage adjustment:

$$\hat{\beta}_1 = \mathbb{E} \left[\frac{dq_i}{dn} \right] = \mathbb{E} \left[\underbrace{\frac{\partial q_i}{\partial n} \Big|_{K_i}}_{\text{Direct } (> 0)} + \underbrace{\frac{\partial q_i}{\partial K_i} \cdot \frac{dK_i^*}{dn}}_{\text{Adaptation } (< 0)} \right] \quad (2.24)$$

The direct and indirect terms differ by case. From the household solution and under the parameterized utility (see Assumption 4 and Appendix A.3), the RDD estimate decomposes as:

$$\hat{\beta}_1 = \underbrace{\pi_1^\beta \cdot 0}_{\text{Case 1}} + \underbrace{\pi_2^\beta \cdot \left(\frac{\bar{q}^*}{T} + \frac{dK^*}{dn} \right)}_{\text{Case 2 (direct + indirect)}} + \underbrace{(\pi_3^\beta + \pi_4^\beta) \cdot \phi}_{\text{Case 3}} \quad (2.25)$$

where π_j^β denotes the population share in Case j near the boundary, $\bar{q}^*$ is the mean unconstrained demand among Case 2 households, and $dK^*/dn < 0$ is the mean storage response among Case 2 households. Because $dK_i^*/dn < 0$, the net Case 2 coefficient $(\bar{q}^*/T + dK^*/dn)$ is strictly smaller than the direct effect alone, and could in principle be negative if storage adaptation is sufficiently large. This decomposition rationalizes two features of the reduced-form evidence: mild overall consumption increase and negative withdrawal rate effect.

The pooled $\hat{\beta}_1$ admits two complementary explanations. First, a large Case 1 share (π_1^β large) implies that most households near the boundary are unconstrained and do not respond to additional supply days. Second, even among responsive Case 2 households, the storage adaptation channel ($dK^*/dn < 0$) attenuates the observed consumption response: households with more supply days have disinvested in storage, reducing their off-day consumption capacity. Disentangling these two channels requires the structural estimation.

The level-stratified pattern in Figure A.8(b) can be rationalized through the storage-adaptation channel in Equation (2.25). The Case 2 net coefficient is $\bar{q}^*/T + dK^*/dn$, where the second term is strictly negative. At low baseline supply, households hold substantial storage (because K^* is decreasing in n), and an additional supply day induces them to shed a large fraction of it; the adaptation term is therefore large in magnitude and can offset or dominate the direct effect. At high baseline supply, installed storage is already low, so the same marginal increase in n triggers only a small storage adjustment, and the Case 2 coefficient converges to its direct-effect ceiling $\bar{q}^*/T$. The monotonic increase in $\hat{\beta}_1$ with baseline supply is consistent with this weakening of the adaptation channel.

The same pattern of existence of non-pipe-constrained households can be seen in the withdrawal rate. For Case 2 households, the network withdrawal rate $q_i/n = q_i^*/T + K_i^*/n$; as n

increases, the second term falls both mechanically (larger denominator) and through storage adaptation (K_i^* itself declines), so the withdrawal rate falls even as total consumption rises. For Case 1, $q_i/n = q_i^*/n$, which also falls in n . For Cases 3 and 4, $q_i/n = \phi$, which is constant. A negative population-average withdrawal rate effect in Figure A.8(e) is therefore consistent with Cases 1 and 2 dominating the boundary population, with few pipe-constrained households. This implies that most households are unconstrained, which is something the structural model later confirms.

2.6 Structural Estimation

Figure 2.4 displays the estimation procedure. I first back out demand primitives using maximum likelihood estimation (MLE), then pin down storage cost parameter from the household FOC. After that, I validate the demand side with quasi-experimental variation from the RDD. Finally, to close the model, I calibrate the supply side and planner weights to simulate counterfactuals.

First, to estimate the demand slope, I need to parametrize the utility function:

Assumption 4 (Utility Function). The sub-period utility function takes the quadratic form with stochastic preferences:

$$u_i(q) = (a_i + \omega_i)q - \frac{b}{2}q^2 \quad (2.26)$$

where q denotes the consumption at each t (either q_i^{on} or q_i^{off}), $a_i > 0$ is the household's persistent marginal valuation, $b > 0$ ensures diminishing marginal utility, and $\omega_{it} \sim \mathcal{N}(0, \sigma_\omega^2)$ is a preference shock observed by the household but not by the econometrician. Decompose $\omega_{it} = \zeta_i^\omega + \xi_{it}^\omega$ into a permanent household-specific component $\zeta_i^\omega \sim \mathcal{N}(0, \sigma_{\zeta,\omega}^2)$ and a transitory component $\xi_{it}^\omega \sim \mathcal{N}(0, \sigma_{\xi,\omega}^2)$, so that $\sigma_\omega^2 = \sigma_{\zeta,\omega}^2 + \sigma_{\xi,\omega}^2$.

The quadratic form is chosen for two main reasons. First, it implies a linear demand curve $q_i^* = T(a_i + \omega_i - p)/b$ per sub-period, which allows the demand slope b to be identified from a single IV coefficient (Step 1 below) rather than requiring functional-form restrictions or higher-order moments. Second, the quadratic form delivers closed-form shadow prices in each of the four binding regimes (Appendix A.3, Table A.9), which makes the regime-specific likelihood contributions in Step 2 tractable. The main cost of the assumption is that the linear demand response is constant in the price level, so extrapolating to the counterfactual price $p^* = 6.14 \text{ JD/m}^3$ (about eight times the administered rate) relies on preferences being well-approximated by a local quadratic expansion around observed consumption.

Estimation proceeds in three sequential steps, leveraging two complementary data sources. The billing universe (approximately 910,000 households observed quarterly over five years) contains metered consumption and the corresponding bill but no information on storage capacity, supply hours, or household demographics. The household survey (approximately 3,500 households across four waves) records storage capacity K_i , hours of supply, income, and other demographics needed for regime classification, but lacks the tariff variation required to identify the demand slope.

The key rationale for this sequential approach is that identifying the demand slope b requires substantial exogenous price variation, which is available only in the billing universe through quasi-

experimental tariff schedule changes as the survey lacks sufficient price variation to separately identify b . Conversely, the physical observables needed for regime classification (K_i , ϕ , n) are available only in the survey. By estimating b first from the billing universe and importing it as a fixed constant into the structural MLE on the survey, I combine the strengths of both datasets while separating the sources of identification.

The structural model (Section 2.4) establishes that observed consumption is determined by the household's binding constraint. A key feature of this setting is that the physical constraints are observable: storage capacity K_i is measured directly in the household survey (roof tank volume), and the pipe ceiling $\phi \cdot n_{d(i)}$ is constructed from the reported average hours of network supply (see below).

The goal of estimation is to recover the parameters of latent demand q_{it}^* : specifically, the preference heterogeneity a_i and the demand slope b , which are required for the welfare calculations in Section 2.7. The four-case structure of the household problem (Table A.8) generates three distinct econometric regimes. Crucially, the regime assignments are observed because K_i and $\phi n_{d(i)}$ are observed.

Regime Classification. For household i in billing period t , define:

$$\begin{aligned} \tau_i &\equiv n_{d(i)}/T && \text{(on-fraction for household } i\text{'s zone)} \\ \bar{q}_i^K &\equiv \frac{K_i T}{T - n_{d(i)}} = \frac{K_i}{1 - \tau_i} && \text{(smoothing threshold)} \end{aligned} \quad (2.27)$$

The smoothing threshold $\bar{q}_i^K$ is the maximum total consumption that the tank can support with perfect temporal smoothing. A critical departure from the standard Tobit formulation is that storage-constrained households (Regime $\mathcal{R}_2$) are *not* censored. Their observed consumption $q_{it} = q_{it}^* \tau_i + K_i$ is a known, invertible function of latent demand:

$$q_{it}^* = \frac{q_{it} - K_i}{\tau_i} \quad (2.28)$$

I observe q_{it} , K_i , and τ_i , so q_{it}^* can be recovered exactly. Treating these observations as censored (as Tobit would) discards information and biases the estimates of preference heterogeneity.

Latent Demand Specification. From the household's first-order condition (see Appendix A.3), unconstrained demand is $q_{it}^* = T(a_i + \omega_i - p_{it})/b$. I parameterize the preference intercept a_i as a function of household demographics $\mathbf{X}_i$ (income, assets, time fixed effects, and location binary for Amman/Zarqa) to avoid the incidental parameters problem:

$$q_{it}^* = \alpha_0 + \mathbf{X}_i' \boldsymbol{\beta} - \frac{T}{b} p_{it} + \varepsilon_{it} \quad (2.29)$$

where $\alpha_0 + \mathbf{X}_i' \boldsymbol{\beta}$ estimates $T a_i/b$ (the ratio of the valuation intercept to the demand slope, in quantity units), and $\varepsilon_{it} \equiv T \omega_{it}/b$ is the rescaled preference shock with $\varepsilon_{it} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$.

The estimable coefficients relate to the primitives as:

$$\underbrace{\alpha_0 + \mathbf{X}'_i \boldsymbol{\beta}}_{\text{Demographic intercept (quantity)}} = \frac{T a_i}{b}, \quad \underbrace{-\frac{T}{b}}_{\text{Price coefficient}} \implies \hat{b} = -\frac{T}{(\text{price coefficient})} \quad (2.30)$$

Utility-space vs. demand-space shocks. The composite residual in the latent demand equation (2.29) is $\varepsilon_{it} \equiv (T/b) \omega_{it}$, which inherits the same decomposition in demand units:

$$\varepsilon_{it} = \zeta_i + \xi_{it}, \quad \sigma_\varepsilon = \frac{T}{b} \sigma_\omega, \quad \sigma_\zeta = \frac{T}{b} \sigma_{\zeta,\omega}, \quad \sigma_\xi = \frac{T}{b} \sigma_{\xi,\omega}. \quad (2.31)$$

Throughout, the plain symbols ζ_i , ξ_{it} , σ_ε , σ_ζ , and σ_ξ denote the demand-space (m^3/week) objects that are directly estimated; the utility-space counterparts ζ_i^ω , ξ_{it}^ω appear only in Assumption 4. Step 2 (MLE) identifies the total standard deviation σ_ε from regime shares; Step 3 (MoM) requires only the transitory component σ_ξ , which is identified from within-household variation in the billing-matched panel; the permanent component is recovered as $\sigma_\zeta = \sqrt{\sigma_\varepsilon^2 - \sigma_\xi^2}$.

Step 1: Demand Slope

Under block-rate pricing, $\text{Cov}(p_{it}, \varepsilon_{it}) > 0$ mechanically: households consuming more face higher average prices. To identify b , I estimate the following specification on the billing universe using instrumental variables:

$$\Delta q_{it} = \gamma \Delta p_{it} + \alpha_i + \lambda_{t,d} + \varepsilon_{it} \quad (2.32)$$

where q_{it} is weekly consumption, p_{it} is average price, α_i are household fixed effects, and $\lambda_{t,d}$ are year-quarter $\times$ consumption-decile fixed effects.

Instrument Construction. In the same spirit as Ito (2014), I construct the simulated price instrument

$$Z_{it}^p = \widehat{AP}_t(Q_{i,t-4}) - \widehat{AP}_{t-4}(Q_{i,t-4}) \quad (2.33)$$

where $\widehat{AP}_t(Q)$ is the predicted average price from a cubic polynomial of average price on quantity estimated within each quarter t , evaluated at household i 's consumption four quarters (12 months) prior ($Q_{i,t-4}$). The choice of the linear specification departs from Ito (2014)'s logarithmic functional form because the demand function I derive is linear, making a linear instrument natural. The instrument isolates variation in prices driven by policy changes in the tariff schedule (between $t-4$ and t), evaluated at a consumption level that is predetermined relative to the current demand shock ε_{it} . The twelve-month lag balances relevance against the exclusion restriction: it is recent enough to predict current tariff placement but sufficiently distant to break correlation with transitory demand shocks. I also show robustness to $t-2$ and $t-6$ lags and provide more details in Appendix A.5.

Identification Conditions. The instrument Z_{it}^p must satisfy three conditions. First, relevance: the instrument must predict variation in the endogenous average price p_{it} . This holds by construction because Z_{it}^p captures changes in the tariff schedule between quarter $t - 4$ and quarter t evaluated at a fixed consumption level, so any policy-induced shift in the price-quantity mapping generates first-stage variation. Second, exclusion: the instrument must affect consumption only through its effect on the price, conditional on the fixed effects. The key threat is that households whose lagged consumption $Q_{i,t-4}$ places them at a kink in the tariff schedule may differ systematically in unobserved demand determinants. The year-quarter $\times$ consumption-decile fixed effects absorb any level differences across consumption groups within each period, so the identifying variation comes from differential changes in the tariff schedule across time for households at the same position on the tariff. The twelve-month lag between the evaluation point $Q_{i,t-4}$ and the current shock ε_{it} further weakens any mechanical correlation, since transitory demand shocks are unlikely to persist over two quarters. Third, independence: the instrument must be uncorrelated with unobserved demand shocks ε_{it} . Because tariff schedule changes are set administratively by the regulator and apply uniformly to all households at a given consumption level, they are plausibly orthogonal to household-level transitory shocks.

Results. The levels-difference IV at lag $t - 4$ yields a demand slope of $\hat{\gamma} = -0.127$ (standard error 0.002, clustered at the household level, $N = 8,642,055$), in units of m^3/week per JD/m^3 . At IV sample means ($\bar{p}^{\text{IV}} = 0.74 \text{ JD}/\text{m}^3$, $\bar{q} = 2.70 \text{ m}^3/\text{week}$), this implies a price elasticity of -0.035 , consistent with inelastic demand for an essential good. The first-stage Kleibergen-Paap F -statistic is 3,697, well above conventional weak-instrument thresholds.

Because the quadratic utility (Assumption 4) implies a linear demand curve with slope $\partial q/\partial p = -T/b$, the structural slope parameter b is identified directly from the levels IV coefficient:

$$\hat{b} = -\frac{T}{\hat{\gamma}} = -\frac{7}{-0.127} \approx 55.04 \quad (2.34)$$

This mapping follows directly from the linear demand $q^* = (T/b)(a + \omega - p)$, whose slope in quantity units is identified by the levels IV coefficient $\hat{\gamma}$.

This estimate of $\hat{b}$ is imported as a fixed constant for all subsequent estimation on the survey data. Fixing b means that the structural MLE does not re-estimate the price coefficient; price enters the likelihood only as a fixed number $-(T/b)p_{it}$.

Discussion. Two features of this estimate merit comment. First, the $\hat{b} = 55.04$ from the billing universe is taken as fixed for the MLE, which means uncertainty in b is not propagated into the estimation, potentially resulting in different standard errors if b were obtained from the same estimation step. Second, because the billing universe pools constrained and unconstrained households, the IV estimate of b reflects a population-weighted average across regimes. To the extent that pipe-constrained households (whose observed consumption does not respond to price) are included,

the estimated slope $|\hat{\gamma}|$ is attenuated toward zero, yielding a larger $\hat{b}$ than would obtain for the unconstrained population alone.¹⁰

Step 2: Demand Heterogeneity

With b fixed, the structural parameter vector reduces to $\theta = \{\alpha_0, \beta, \sigma_\varepsilon\}$. Let μ_{it} denote the predicted mean of latent demand:

$$\mu_{it} = \alpha_0 + \mathbf{X}'_i \beta - \frac{T}{b} p_{it} \quad (2.35)$$

The likelihood function reflects the three-regime structure. Each observation $(q_{it}, K_i, n_{d(i)})$ falls into exactly one regime, determined entirely by observable quantities. Thus, to classify households, I need an estimate of the flow rate ϕ . I use a common pipe ceiling (m^3 per week) inferred from the universe of meter-level billing data. Since the pipe constraint in the household problem is $q \leq \phi \cdot n$, and I observe both q and n , I then take ϕ to be $4.10 \text{ m}^3/\text{day}$, the 95th percentile of this distribution.¹¹

Regime $\mathcal{R}_1$: Unconstrained ($q_{it} < \min\{\bar{q}_i^K, \phi n_{d(i)}\}$). The household reveals its true latent demand $q_{it} = q_{it}^*$. The likelihood contribution is the standard normal density:

$$\ell_{it}^{(1)} = \frac{1}{\sigma_\varepsilon} \varphi\left(\frac{q_{it} - \mu_{it}}{\sigma_\varepsilon}\right) \quad (2.36)$$

Regime $\mathcal{R}_2$: Storage-constrained ($q_{it} \geq \bar{q}_i^K$ and $q_{it} < \phi n_{d(i)}$). Observed consumption is $q_{it} = q_{it}^* \tau_i + K_i$. Because this transformation is invertible, latent demand is exactly recovered as $q_{it}^* = (q_{it} - K_i)/\tau_i$. The likelihood contribution accounts for the Jacobian of this mapping:

$$\ell_{it}^{(2)} = \frac{1}{\tau_i \sigma_\varepsilon} \varphi\left(\frac{(q_{it} - K_i)/\tau_i - \mu_{it}}{\sigma_\varepsilon}\right) \quad (2.37)$$

Regime $\mathcal{R}_3$: Pipe-constrained ($q_{it} \geq \phi n_{d(i)}$). When either the pipe binds but the tank either binds or does not, observed consumption equals the pipe ceiling $q_{it} = \phi n_{d(i)}$, so the two cases are observationally indistinguishable from the consumption data alone. I therefore collapse them into a single econometric regime $\mathcal{R}_3$ for estimation: because the likelihood sees only total consumption,

¹⁰To see why, note that pipe-constrained households have $\partial q_i / \partial p = 0$, while unconstrained households respond at the structural slope $-T/b$. Pooling both types, the IV recovers a population-weighted average $\hat{\gamma} \approx \pi_{\text{unc}} \cdot (-T/b) + \pi_{\text{pipe}} \cdot 0$, so $|\hat{\gamma}|$ is attenuated toward zero. Because $\hat{b} = -T/\hat{\gamma}$, this yields a larger $\hat{b}$ (less curvature in structural units) than the true unconstrained population.

¹¹This number is consistent with other intermittent water supply studies in developing countries. For example, Kumpel et al. (2017) report Indian median metered consumption of 12.0–18.5 kL/month across household types with varying storage infrastructure and tap-sharing arrangements, corresponding to roughly 2.8–4.3 m^3/week . Thus, the 95th percentile of the billing-data distribution in my paper is consistent with the upper bound in their data.

all that is needed is the right-censoring threshold at which the pipe constraint begins to bind. This threshold depends on the observed tank capacity K_i :

$$\tilde{q}_{it} \equiv \begin{cases} \phi n_{d(i)} & \text{if } \phi n_{d(i)}(1 - \tau_i) \leq K_i \\ \frac{\phi n_{d(i)} - K_i}{\tau_i} & \text{if } \phi n_{d(i)}(1 - \tau_i) > K_i \end{cases} \quad (2.38)$$

The first branch gives the Case 3 threshold, which applies when the tank is large enough to smooth the full pipe throughput; the second gives the Case 4 threshold, which applies when both constraints bind simultaneously. The likelihood contribution is the probability that latent demand exceeds this threshold:

$$\ell_{it}^{(3)} = 1 - \Phi\left(\frac{\tilde{q}_{it} - \mu_{it}}{\sigma_\varepsilon}\right) \quad (2.39)$$

These cases differ structurally in the daily allocation of consumption across on- and off-periods, but because the survey records only billing-cycle totals, this distinction is not identifiable from the likelihood alone; both cases contribute the same right-censored density. The distinction is preserved in the counterfactual welfare calculations, where gross utility depends on sub-period rates and each household is classified separately based on the computed K_i^* .

Full Log-Likelihood. Let $R_{it} \in \{1, 2, 3\}$ denote the observed regime. The full log-likelihood is:

$$\ln \mathcal{L}(\boldsymbol{\theta}) = \sum_{i=1}^N \sum_{t=1}^{T_i} \left[\mathbb{1}_{R_{it}=1} \ln \ell_{it}^{(1)} + \mathbb{1}_{R_{it}=2} \ln \ell_{it}^{(2)} + \mathbb{1}_{R_{it}=3} \ln \ell_{it}^{(3)} \right] \quad (2.40)$$

where the structural parameter vector is $\boldsymbol{\theta} = \{\alpha_0, \boldsymbol{\beta}, \sigma_\varepsilon\}$ and b enters as a fixed constant. Standard errors are obtained from the inverse of the numerical Hessian of the negative log-likelihood evaluated at the MLE optimum: $\widehat{\text{Var}}(\hat{\boldsymbol{\theta}}) = \left[-\nabla^2 \ln \mathcal{L}(\hat{\boldsymbol{\theta}}) \right]^{-1}$.

Identification. The parameters are identified by two distinct sources of variation. First, preference heterogeneity $(\alpha_0, \boldsymbol{\beta})$ is identified cross-sectionally from unconstrained households (whose latent demand is directly observed) and the exact inversion of demand for storage-constrained households (equation (2.28)). Second, the total standard deviation σ_ε is identified by the mass of households hitting the observable deterministic constraints versus those interior to the choice set: a larger σ_ε predicts more households at the boundaries, and the observed regime shares discipline this parameter. The transitory component σ_ξ is identified separately from within-household variation in the billing-matched panel (Step 3), and the permanent component is recovered residually as $\sigma_\zeta = \sqrt{\sigma_\varepsilon^2 - \sigma_\xi^2}$; the two are pinned down by distinct moments.

With b fixed from the billing universe, the price term enters the MLE fixed at $-(T/b) p_{it}$; the MLE does not re-estimate the price coefficient. Any residual correlation $\text{Cov}(p_{it}, \varepsilon_{it}) > 0$ from block-rate pricing is therefore absorbed into $\hat{\alpha}_0$ and $\hat{\boldsymbol{\beta}}$. However, demand is inelastic (implied

elasticity of -0.035 at IV sample means), so the price term explains a modest share of the total variance in latent demand q^* , limiting the scope for mechanical endogeneity to distort the other coefficients. Furthermore, to the extent that these coefficients are used to create a measure of heterogeneity, a causal interpretation is not required.

Step 3: Long-Run Storage Cost

Step 2 estimates the total variance $\sigma_\varepsilon^2 = \sigma_\zeta^2 + \sigma_\xi^2$, but recovering the storage cost parameter g requires the transitory component σ_ξ on its own. Storage is a long-run investment chosen to buffer against *transitory* demand risk: each household knows its own permanent type ζ_i when selecting tank capacity and integrates only over the distribution of ξ_{it} . Concretely, ζ_i captures stable household characteristics like plumbing quality, garden area, or family composition, while ξ_{it} captures shorter-horizon shocks like guests, illness, or seasonal variation.

I identify σ_ξ from the billing-matched panel (approximately 4,500 observations in the same household survey linked to 8 quarters of billing records). I use the billing-matched panel rather than the full billing universe because the relevant transitory variance is the within-household variation in the latent demand residual conditional on the same covariates $\mathbf{X}_i$ that enter the MLE, not the within-household variation in raw observed consumption, which conflates preference shocks with regime switching and block-rate price changes. A household fixed-effects regression of weekly consumption on quarter indicators yields the within-household residual standard deviation directly:¹²

$$q_{it} = \alpha_i + \delta_t + \xi_{it}$$

With $\hat{\sigma}_\xi$ in hand, the permanent component is then recovered from the variance decomposition: $\hat{\sigma}_\zeta = \sqrt{\hat{\sigma}_\varepsilon^2 - \hat{\sigma}_\xi^2}$. With the short-run latent demand parameters ($\hat{\theta}$) estimated, I exploit the household's long-run, first-stage investment decision to recover the cost of private adaptation. I parameterize the annualized storage cost function as quadratic:

Assumption 5 (Quadratic Storage Cost).

$$\Gamma_i(K_i) = \frac{g}{2} \cdot K_i^2 \quad (2.41)$$

The quadratic form is a parsimonious specification satisfying Assumption 1 with strict convexity, requiring only a single parameter that is identified from the household's first-order condition. Rising marginal cost is consistent with the engineering and spatial constraints that govern residential tank installation: structural reinforcement of roofs and stands, plumbing and pressure-regulation complexity, and the opportunity cost of dedicated roof or yard space all plausibly rise more than linearly in capacity. The form is also used in climate-adaptation literature (e.g., Carleton et al., 2022). The counterfactual welfare numbers depend on $\hat{g}$ only through the storage-cost term in the

¹²In this case, $\hat{\sigma}_\xi = \sqrt{\sum_{i,t} \hat{\xi}_{it}^2 / (N_{\text{obs}} - N - (T - 1))}$ where N is the number of households, T is the number of quarters, $N_{\text{obs}} \leq NT$ is the total number of household-quarter observations in the unbalanced panel, and $\hat{\xi}_{it}$ is the residual from the two-way fixed-effects regression.

welfare decomposition; the qualitative comparative statics in Propositions 2–3 hold for any convex Γ , so the welfare ranking across regimes is robust to functional form and only the magnitude of the storage-cost saving moves with the curvature.

Because the shadow price of storage μ_K is zero when the tank is slack and strictly positive when the tank binds, it is a kinked, convex function. Rewriting the shadow price formulas from Table A.9 in terms of latent demand (using $q_{it}^* = \hat{\mu}_i + \varepsilon_{it}$, where $\hat{\mu}_i$ is the MLE-predicted mean), the shadow price can be expressed as a function of the demand-space transitory shock ξ_{it} with standard deviation $\hat{\sigma}_\xi$. Because the household knows its own permanent type ζ_i when choosing capacity, the expectation integrates only over the transitory component:

$$\mathbb{E}[\hat{\mu}_{K,it}] = \int_{\bar{\xi}_{it}}^{\infty} \hat{\mu}_{K,it}(\xi) \frac{1}{\hat{\sigma}_\xi} \varphi\left(\frac{\xi}{\hat{\sigma}_\xi}\right) d\xi \quad (2.42)$$

where $\bar{\xi}_{it}$ is the household-specific critical shock threshold at which desired consumption exactly equals the physical smoothing limit ($q_{it}^* = \bar{q}_i^K$), and $\hat{\mu}_{K,it}(\cdot)$ is the algebraic shadow price from Table A.9 evaluated at the MLE parameters $\hat{\theta}$.

Household capacity K_i is thus a fixed investment chosen before the realization of transitory shocks, so the household's first-order condition equates the expected marginal benefit of storage to its marginal cost. The theoretical model treats the administered price p and supply duration $n_{d(i)}$ as constants known at the time of investment. Empirically, reported supply frequency and administered prices vary slightly across quarters due to changes in self reports and tariff adjustments. To make full use of the panel, I average the estimated shadow value across all observed household-quarter pairs, yielding a single expected marginal benefit per household:

$$\overline{MB}_i = \frac{1}{T} \sum_{t=1}^T \mathbb{E}_\xi[\hat{\mu}_{K,it}] \quad (2.43)$$

where the inner expectation integrates over the transitory shock ξ at each observed price p_t .

The theoretical model treats p as constant at the time of investment, which is valid if administered prices do not vary over the sample. In the data, the cross-wave coefficient of variation (CV) of administered prices is 2.8%, and the cross-wave CV of $\mathbb{E}_\xi[\mu_{K,it}]$ stripped of seasonal demand shifters (i.e., computed on the wave-FE-residualized series) is 3.0%. Because within-sample variation in the structurally relevant quantity is small, the time-averaging in $\overline{MB}_i$ is not a structural assumption: the moment condition $\overline{MB}_i = gK_i^*$ holds approximately period-by-period.¹³

The single capital cost parameter g is then consistently estimated in a second step using the Method of Moments, matching the sample average expected marginal benefit of storage to the sample average marginal cost across the population:

$$\hat{g} = \frac{\sum_{i=1}^N \overline{MB}_i}{\sum_{i=1}^N K_i} \quad (2.44)$$

¹³The estimator remains valid under the more general forward-looking interpretation in which the household integrates over the stationary distribution of shadow prices over the tank's lifetime.

This sequential estimation strategy separates the identification of two margins: short-run flow utility is identified from high-frequency billing and tariff variation (Step 1) and cross-sectional regime shares (Step 2), while the long-run capital cost parameter is identified from the structural first-order conditions governing physical infrastructure investment (Step 3).

Estimation Results

Table 2.1 reports the structural MLE estimates. The intercept $\hat{\alpha}_0 = 3.99$ is precisely estimated, implying a baseline latent demand of approximately 3.99 m³ per week for the reference group (non-Amman households observed in the reference wave with zero expenditure and assets). Among the demographic covariates, total household expenditure enters positively and significantly ($t = 3.19$, $p < 0.01$): households with higher consumption spending also demand more water. The asset index coefficient is negative but statistically indistinguishable from zero ($t = -0.23$); this is consistent with the Engel curve in Fact #2 being driven primarily by expenditure rather than durable-goods wealth. Conditional on expenditure and assets, the Amman indicator is small and not statistically distinguishable from zero ($\hat{\beta}_{\text{amman}} = 0.04$, $t = 0.23$), and its interaction with expenditure is similarly negligible ($t = -0.01$). Consumption differences between Amman and Zarqa appear to be fully mediated by the expenditure and asset controls.

The estimated total error standard deviation is $\hat{\sigma}_\varepsilon = 3.91$ m³/week. Combined with the transitory component $\hat{\sigma}_\xi = 1.71$ m³/week recovered from the billing-matched panel variance decomposition, the permanent component is $\hat{\sigma}_\zeta = 3.52$ m³/week. The dominance of the permanent component indicates that most unobserved demand heterogeneity is time-invariant (e.g., garden area, plumbing quality, family composition). The transitory component is the economically relevant source of variance for the storage investment decision, since households know their own type ζ_i when choosing tank capacity and integrate only over ξ_{it} .

The regime classification (Figure 2.5) reveals that 77.4% of household-round observations are unconstrained (Regime 1), 14.0% are storage-constrained (Regime 2), and the remaining 8.6% are pipe-constrained (Regime 3). The large unconstrained share is reassuring for identification: the preference parameters are primarily pinned down by direct observation of q_{it}^* rather than by functional form assumptions on the censored tail. The demand slope $b = 55.04$, imported from the billing universe IV (Section 2.6), implies a price elasticity of approximately -0.035 at the IV sample mean. The Method of Moments estimate of the storage cost parameter is $\hat{g} = 0.82$.

Model Fit

Observed vs. Model-Implied Storage. Table A.4 documents that observed storage exceeds the static-model optimum for a large share of households: mean observed capacity is 5.24 m³ versus a model-implied optimum of 2.70 m³, with a cross-sectional correlation of 0.040. Formally, 47.1% of households hold capacity exceeding the current-schedule utilization ceiling $\bar{K}_i \equiv \phi n_{d(i)}(1 - \tau_i)$ (Proposition 5), with mean excess of 2.15 m³. The excess rises monotonically with income, from 42% in Q1 to 55% in Q5.

Two interpretations are consistent with this pattern, and both reinforce the paper’s central mechanism. First, $n_{d(i)}$ is not in fact deterministic: households installing a durable tank face a lifetime distribution of supply days, with substantial cross-sectional variation (the within-wave interquartile range of $n_{d(i)}$ in the estimation sample is roughly 1–2 days per week). Privately optimal capacity is therefore sized against bad realizations of $n_{d(i)}$ rather than its median, and what the static proposition flags as “excess” is the option value of buffering against worse-than-median supply weeks. Second, intermittency has been easing slowly over the sample window: mean supply duration in the survey rose modestly, from 2.08 days/week in 2015 to 2.30 days/week by 2018. Because tanks are lumpy and long-lived (Table A.6), the installed stock partly reflects earlier, more severe conditions. The apparent over-investment is therefore a combination of option-value buffering and a small legacy component, rather than a current optimization failure.

This has a direct implication for the welfare analysis. The counterfactuals in Section 2.7 use the static model-implied K_i^* rather than observed K_i , which understates the stock of storage capital actually bearing on households and therefore produces a *lower-bound* estimate of the welfare gain from reform. If households in fact hold observed K_i for either option-value or legacy reasons, the true deadweight cost of intermittency and the true gain from pricing with continuous supply are larger than the numbers reported below.

The shadow value of storage (Figure A.5a) drops to zero for all households with $K_i > 5 \text{ m}^3$; for 52.1% of the sample, $\mathbb{E}[\mu_K] < 0.01$. Table A.3 decomposes the cross-sectional variation in K_i : housing variables alone explain $R^2 = 0.011$, the model’s structural determinants explain $R^2 = 0.031$, and the combined specification yields $R^2 = 0.032$, with housing variables retaining their magnitudes while the structural variables remain individually insignificant. These findings motivate treating K_i as a predetermined housing endowment in the short run, while the comparative statics $dK_i^*/dn < 0$ and $dK_i^*/dp < 0$ describe how the installed stock evolves in the long run as the government adjusts policy.

Comparison to the RDD. Table 2.2 compares the model-implied response to supply duration against the quasi-experimental RDD estimate. The aggregate structural model predicts a city-wide marginal effect of $dq/dn \approx 0.475 \text{ m}^3/\text{week}$ per additional supply day. The theoretical decomposition reveals that this is driven primarily by pipe-constrained households (contributing +0.351), with the storage-constrained channel contributing 0.124 (direct: +0.178; adaptation: −0.054) and unconstrained households contributing nothing.

The empirical RDD identifies a LATE of $0.081 \text{ m}^3/\text{week}$ per additional supply day, substantially smaller than the aggregate model prediction. The structural decomposition rationalizes this gap: restricting the model to non-pipe-constrained households yields a predicted marginal effect of 0.124, comparable to the RDD estimate. The muted empirical response is therefore consistent with a boundary population that is predominantly unconstrained, so that the local average treatment effect understates the aggregate response that includes pipe-constrained households who mechanically absorb all additional supply. The comparison should be interpreted as a qualitative consistency check rather than a formal test: the structural model can rationalize a small dq/dn through composition effects without requiring a revision of the demand parameters.

2.7 Counterfactual Analysis

Other Parametrizations

To operationalize the aggregate welfare calculations and evaluate counterfactual policies, I must specify a functional form for the welfare weights introduced in Definition 4. Similar to the public finance literature (e.g., Saez, 2002) that parameterizes welfare weights as a function of disposable income, I summarize the redistributive tastes of the planner using a simple parametric form:

$$\tilde{g}(y_i) = \frac{1}{y_i^\eta} \quad (2.45)$$

where $\eta \geq 0$ is a scalar parameter capturing the planner's aversion to inequality.

To evaluate policies that balance the government budget, these weights must satisfy the normalization condition such that the population average is 1. I normalize the raw weights by $\overline{y^{-\eta}} \equiv \frac{1}{N} \sum_{j=1}^N y_j^{-\eta}$. The empirical social marginal welfare weight applied to household i is therefore: $g_i \equiv y_i^{-\eta} / \overline{y^{-\eta}}$. This formulation maps exactly to the core benchmark cases: a utilitarian planner ($\eta = 0$, implying uniform weights $g_i = 1$), a planner with logarithmic redistributive tastes ($\eta = 1$), and a Rawlsian planner ($\eta \rightarrow \infty$). In the baseline counterfactual simulations, I report results for $\eta = 1$ and check robustness to $\eta = 0$ and $\eta = 5$. Finally, the leakage function requires parametrization for counterfactual policy analysis:

Assumption 6 (Iceberg Cost). The leaked water from poor infrastructure and unauthorized use is proportional to supply duration. Specifically, the fraction of gross production lost to leakage is $\delta\tau$, where $\delta > 0$ is the network decay parameter and $\tau = n/T$, subject to $\delta n < T$ (so that gross extraction remains finite). This implies an iceberg cost multiplier:

$$\iota(n) = \frac{T}{T - \delta n} \quad (2.46)$$

The intuition behind this choice is that to deliver a net quantity S to households, the utility must extract $\iota(n) \cdot S$ from the aquifer. Because the water pumped to compensate for leakage also leaks, the multiplier compounds: gross extraction is $G = S + \delta\tau \cdot G$, hence $G = S \cdot T / (T - \delta n) = \iota(n) \cdot S$. The compounding structure of $\iota(n)$ is the appropriate long-run specification for a network where leakage scales with pressurized pipe-time; Section 2.7 reports that the welfare results are qualitatively robust to a linear (non-compounding) leakage function.

The network decay parameter δ is itself a policy object: comprehensive pipe rehabilitation lowers δ by replacing aging mains and deteriorated joints, shifting the entire iceberg schedule $\iota(n; \delta)$ downward. I therefore distinguish two values, δ_t for the degraded status quo network and $\delta_r < \delta_t$ for a rehabilitated network, and evaluate each counterfactual regime under the relevant δ . In the counterfactuals, I consider two versions of the First-Best: one that retains the existing degraded network (δ_t) and isolates the welfare gain from the price instrument alone, and one that combines market-clearing pricing with rehabilitation (δ_r) and adds the gain from reduced physical

losses. This decomposition is important because it separates the allocative efficiency of pricing which requires no new infrastructure from the additional efficiency of reducing leakage, which does.

Supply Calibration

Table 2.3 summarizes all calibrated parameters and below I go through them.

Supply Capacity (S). I calibrate the aggregate gross supply constraint using administrative records from Miyahuna and the Ministry of Water and Irrigation. Miyahuna’s 2022 annual report documents 209.17 MCM of system input to the Amman governorate (Miyahuna, 2023), supplemented by 62.91 MCM to Zarqa and 10.29 MCM to Madaba, giving total system input for Miyahuna’s service area of 282.4 MCM in 2020. Approximately 10% of system input serves non-residential users, which I hold constant across counterfactuals. The residential gross capacity is therefore $S = 282.4 \times 0.90 \approx 254$ MCM/yr, or approximately 5.43 m³/week per household.

Cost Function Calibration. I calibrate the technology parameters by combining the 2022 *Water Utilities Monitoring Report* (Utilities Performance Monitoring Unit, 2023) with the *National Water Strategy 2023–2040* (Ministry of Water and Irrigation, 2023). The Monitoring Report gives Miyahuna’s 2022 unit operating cost as 0.93 JD/m³, of which 43.8% is electricity; taking this variable share yields a marginal cost of $c = 0.93 \times 0.438 \approx 0.41$ JD/m³. The National Water Strategy reports the full economic cost of supply (inclusive of capital service and debt) at 2.16 JD/m³; calibrating the status-quo fixed cost as the residual gives a unit fixed cost of $2.16 - 0.41 = 1.75$ JD/m³, which scaled by Miyahuna’s quarterly net authorized consumption of 42.8 MCM yields $F_t \approx 75$ million JD/quarter. The rehabilitated regime adds amortized capital from the *National Non-Revenue Water Strategy*’s 1.27 billion JD rehabilitation program, net of reduced pipe-fatigue repairs (Christodoulou and Agathokleous, 2012), giving $F_r \approx 100$ million JD/quarter. The quota regime adds Advanced Metering Infrastructure costs of 355 JD per smart meter (Ministry of Water and Irrigation and USAID, 2025), amortized over 15 years at 5% across roughly one million households, yielding $F_{AMI} \approx 8.6$ million JD/quarter and a total $F_q \approx 83.6$ million JD/quarter. All cost objects are denominated in net delivered water, with the iceberg multiplier $\iota(n)$ grossing up to gross extraction.

Physical Leakage (δ). The observed NRW rate of approximately 43% reflects both aging infrastructure leakage and unauthorized use (Utilities Performance Monitoring Unit, 2023). Because only physical leakage scales mechanically with supply duration n (unauthorized connections draw a fixed volume regardless of pressurization hours), I calibrate δ to match only the physical component. Following Al-Rkebat (2026), who attribute roughly 70% of NRW to commercial losses, the physical leakage share is approximately 12.9% of gross production. Under the status quo ($n \approx 2$ days/week), matching this physical leakage rate implies $\delta_t \approx 0.45$, yielding a *physical* multiplier $\iota_{\text{phys}}(2) \approx 1.15$ and an *effective* multiplier $\iota_{\text{eff}}(2) \approx 1.64$ once constant commercial losses are

grossed up. For rehabilitated infrastructure, I use USAID project evaluations documenting that comprehensive rehabilitation reduced NRW from 46% to 11% in a pilot Amman zone (USAID Water Governance Activity, 2022). Adjusting for the commercial loss share, the physical leakage reduction factor is approximately 0.36, giving $\delta_r \approx 0.16$.

Scarcity Rent (ν). I calibrate the scarcity rent using the Hotelling backstop principle: the long-run social marginal cost of groundwater should equal the marginal cost of the cheapest alternative source (Red Sea–Dead Sea desalination, approximately 1.0 JD/m³ at the point of delivery). Setting $\Lambda = (c + \nu) \iota(n^{SQ})$ equal to the backstop cost and solving, using the effective iceberg multiplier that accounts for both physical and commercial losses ($\iota_{\text{eff}}(2) \approx 1.64$), yields $\nu \approx 0.20$ JD/m³.

Administered Price. The Jordanian water tariff is an increasing block schedule in which the marginal price rises with consumption. For the structural model, I follow Ito (2014) who shows that consumers respond to averages prices and collapse this schedule into a single volumetric price $\bar{p} = 0.76$ JD/m³, calibrated from the mean average price in the billing universe over 2020–2024.¹⁴

Numerical Approach

The counterfactual equilibria cannot be solved in closed form because the endogenous storage response $K_i^*(\Theta)$ and the regime classification interact nonlinearly with the policy instruments. I therefore compute each regime numerically using the procedure described briefly below. For more details, see Algorithm in Appendix A.1.

The computation has two layers. The inner loop takes a candidate policy $(p, n, \bar{q})$ as given, solves each household’s optimal storage from the first-order condition, integrates consumption and gross utility over $M = 100$ draws of the preference shock ω_i , averaging across the regime-dependent consumption rules that preclude closed-form evaluation, and assembles the social welfare function. The outer loop searches over the regime-specific instrument set, retaining the feasible combination that maximizes $\mathcal{W}$ subject to the aggregate supply constraint $\iota(n) \sum_i \mathbb{E}[q_i] \leq S$.

For the Third-Best, the search is a simple grid over n in quarter-day increments. For the First-Best, each grid point in n requires finding the welfare-maximizing price p^* conditional on that n ; if the resulting demand exceeds the supply constraint, the price is raised until the constraint binds exactly. For the Second-Best, the analogous step finds the quota $\bar{q}^*$ that reduces demand to the supply constraint at each n . In both cases, the one-dimensional search over p or $\bar{q}$ is conducted numerically via Brent’s method.

¹⁴The increasing block tariff (IBT) structure is the source of the endogenous-price problem that motivates the simulated instrument in Section 2.6: households who consume more mechanically face a higher average price, so naive OLS of q on p does not identify the demand slope. Once b is recovered from tariff-schedule variation (Step 1), the scalar $\bar{p}$ enters the counterfactual arithmetic only as the average revenue the utility collects per m³, which is the fiscally relevant summary for comparing pricing regimes to intermittency. Modeling the full IBT directly would add more complexity, see for example Szabó (2015) for water or Reiss and White (2005) for electricity.

Counterfactual Results

Before getting into welfare, it is useful to see prices under the different regimes along with their components. Table 2.4 decomposes the social marginal cost across regimes and inequality aversion levels. Under the utilitarian benchmark ($\eta = 0$), the optimal price approximately equals the social marginal cost in both the degraded and rehabilitated pricing regimes, consistent with Theorem 1 when $\bar{g}_q = 1$. As inequality aversion rises, the planner prices water strictly above its social marginal cost as an equity instrument: at $\eta = 1$ the optimal price exceeds SMC by 0.82 JD/m³ on degraded pipes and 0.87 JD/m³ on rehabilitated pipes, and at $\eta = 5$ the wedge rises to 2.63 and 2.64 JD/m³ respectively. The planner sets price above SMC because the equity-weighted marginal value of the fiscal surplus exceeds the equity-weighted loss from the price wedge. This follows from the covariance between welfare weights and consumption being near zero (i.e., $\bar{g}_q \approx 1$, consistent with the flat volumetric Engel curve). Across all panels, the status quo price of 0.76 JD/m³ sits well below the social marginal cost of 1.00 JD/m³, confirming the fiscal deficit documented in Section 2.7.

Table 2.5 reports equilibrium outcomes under each regime at $\eta = 1$. I present five columns: the status quo, optimized intermittency (Third-Best), pricing on degraded pipes, pricing on rehabilitated pipes, and quotas (Second-Best).

Status Quo and Optimal Intermittency. The planner inherits degraded pipes ($\delta_t = 0.45$) and a fixed administered price $\bar{p} = 0.76$ JD/m³. At the observed median supply duration of $n = 2$ days per week, households invest in an average of 3.19 m³ of storage capacity and achieve a mean consumption of 3.39 m³/week. Under these conditions, the effective iceberg cost multiplier is $\iota(2) = 1.64$, reflecting both physical leakage ($\approx 13\%$ of gross) and the commercial-loss gross-up. The utility operates at a fiscal deficit: net revenue $(p - \Lambda)Q = -1,893$ JD/week, reflecting the subsidy wedge between the administered price and the full social marginal cost.

When the planner is allowed to re-optimize the supply duration while keeping the administered price fixed (the Third-Best regime), the optimum coincides with the status quo: $n^{TB} = 2$ days per week and welfare gain of essentially zero relative to the status quo. Intuitively, on degraded pipes with a fixed subsidized price, the benefit of extending supply (relaxing tank constraints for a minority of households) is exactly offset by the marginal leakage cost $(c + \lambda)\iota'(n)Q$ and the fiscal wedge from induced subsidized consumption $\Delta_p \cdot dQ/dn$. The status quo intermittency level is not an accident of politics or inertia: within the rationing paradigm, i.e., holding the price at $\bar{p}$ and ruling out quotas, the observed $n = 2$ is approximately what a social planner with standard inequality aversion would also choose.

Pricing on Existing Pipes. Even without infrastructure rehabilitation, pricing on the existing degraded network improves aggregate welfare by 3.1% at the city level relative to the status quo (Figure 2.6a, $\eta = 1$).¹⁵ The optimal price rises to $p^* = 6.14$ JD/m³ (an 8.1-fold increase over the

¹⁵This baseline figure employs two modeling choices. First, the iceberg cost takes the compounding form $\iota(n) = T/(T - \delta n)$, which treats physical leakage as scaling with pressurized pipe-time (water pumped to replace leaks

administered rate) and supply extends from 2.00 to 3.75 days per week as the higher price dampens demand and relaxes the resource constraint. The effective iceberg multiplier rises to $\iota(3.75) = 1.89$, reflecting higher leakage from longer pressurization on aging pipes. Mean consumption falls from 3.39 to 2.88 m³/week as the price increase rations demand. Storage needs drop from 3.19 to 1.98 m³ as the extended supply window reduces the smoothing motive: households need less buffer capacity when water flows 3.75 rather than 2 days per week.

The welfare gain has a stark fiscal composition. Weighted consumer surplus falls from 128,368 to 96,588 JD/week in the sample (a loss of 31,780 JD/week) as the price increase extracts surplus from households. But net fiscal revenue swings from $-1,893$ to $+33,370$ JD/week, a gain of 35,263 JD/week that more than offsets the consumer surplus loss if redistribution is feasible, a point that I return to later in this section.

Pricing with Rehabilitation. When the planner can set the price and invest in rehabilitated pipes ($\delta_r = 0.16$), the optimal policy raises the price to $p^* = 4.11$ JD/m³ and expands supply to $n^* = 6.50$ days per week, approaching continuous service. The rehabilitated network's lower leakage rate ($\iota(6.50) = 1.68$ versus 1.89 on degraded pipes at 3.75 days) allows the planner to extend supply without the punitive leakage costs that constrain the degraded-pipe regime. Mean consumption recovers to 3.22 m³/week, close to the status-quo level of 3.39 m³/week, but now at a price that reflects the true social marginal cost.

The near-elimination of storage is another interesting household-level outcome. Mean optimal tank capacity falls from 3.19 m³ under intermittency to just 0.50 m³ under the First-Best, an 85% reduction. With water flowing 6.50 out of 7 days, the temporal smoothing motive that drives storage investment is nearly absent. This validates the model's central mechanism: the 3.19 m³ of storage capital observed under intermittency is overwhelmingly deadweight, a private cost borne by households that substitutes for the public infrastructure that efficient pricing would provide.

The aggregate welfare gain relative to the status quo is 7,189 JD/week in the estimation sample, which scales to approximately 2.78 million JD/week at the city level, or 6.4% of status-quo welfare. (Figure 2.6a). However, equity-weighted consumer surplus falls from 128,368 to 115,539 JD/week: households are unambiguously worse off from the perspective of their own consumption welfare. The aggregate welfare ranking reverses only because net revenue rises to $+23,093$ JD/week under pricing, compared to $-1,893$ in the status quo. Additional fixed costs of 19,872 JD/week (versus 14,904 in the status quo) reflect the amortized rehabilitation capital.

also leaks). A linear function $\iota_{\text{linear}}(n) = 1 + \delta n/T$ shifts n^* from 3.75 to 5.00 days and raises the degraded-pipes gain from 3.1% to 5.7%; on rehabilitated pipes the shift is smaller (0.25 days; gain from 6.4% to 7.0%). Thus, the qualitative results are robust to the choice of the iceberg cost.

Second, the baseline values a dinar of fiscal deficit at a dinar of welfare, equivalent to a marginal cost of public funds (MCPF) of one. Because Jordan's water subsidy is financed through distortionary channels (concessional loans with eventual repayment, arrears to the electricity utility that crowd out other public spending), the true MCPF likely exceeds unity; benchmark estimates for developing countries place it in the range of 1.1–1.3 (Auriol and Warlters, 2012). Under MCPF = 1.2, the degraded-pipes gain rises from 3.1% to 9.5% and the rehabilitated gain from 6.4% to 11.0%. Intuitively, pricing reform transfers resources from consumers to the utility; under MCPF > 1, these resources are more valuable in public hands than in private hands because they substitute for distortionary taxation elsewhere. The reported baseline estimates that use MCPF = 1 are thus a *lower bound* on the gains from the price reform.

Quotas (Second-Best). Adding a household-level quota while retaining the existing degraded pipes and the administered price does not improve welfare at the baseline inequality aversion. At $\eta = 1$, the welfare-maximizing policy sets the quota at $\bar{q}^* = 4.50 \text{ m}^3/\text{wk}$ (binding) with $n^{SB} = 2.75$ days; total welfare falls by 0.6% relative to the status quo, reflecting the additional smart-metering infrastructure cost ($F_q = 7.15 \text{ JD/week}$ per household, versus $F_t = 6.41$ under the status quo) that more than offsets the modest allocative gain from capping extraction. At $\eta = 5$, the quota stays at $\bar{q}^* = 4.50 \text{ m}^3/\text{wk}$ and $n^{SB} = 2.75$, with welfare 0.1% below the status quo.

Distributional Incidence. Figure 2.7 shows the distribution of welfare changes as a share of weekly expenditure. Without transfers, pricing produces losers across the entire income distribution: 100% of households are worse off, with the median household losing 13.2% of weekly expenditure (Figure 2.7a). The price increase from 0.76 to 6.14 JD/m³ acts as a deeply regressive tax: the poorest quintile loses approximately 25% of weekly expenditure while the wealthiest quintile loses roughly 8% (Figure 2.8a). A flat per-capita lump-sum rebate of the net revenue partially reverses this picture but does not flip it: 49.3% of households gain, and the median household still loses a negligible share of weekly expenditure (−0.02%, Figure 2.7b).

Feasibility of Transfers. The fiscal surplus generated by pricing creates room for redistribution, but in a setting whose central motivation is limited state capacity, the relevant question is whether any transfer rule conditioning only on variables the government observes can make the reform a Pareto improvement. Table 2.6 shows the results. Under pricing on existing pipes, no rule considered brings even a majority of households into the winners' column: flat rebates leave 50.7% worse off, pro-rata expenditure 56.4%, pro-rata storage 60.7%, and progressive targeting proportional to $1/y_i$ leaves 60.2% of households worse off and performing strictly worse than flat rebates. The reason is that the heterogeneity driving winners and losers is not expenditure but a combination of installed storage capacity, latent demand type, and which physical constraint each household faces, none of which the utility observes at the individual level. Progressive targeting on the one dimension the utility does observe (income) mis-targets because of the flat volumetric Engel curve documented in Fact #2: poor and rich households consume similar volumes, so the welfare impact of a price increase is only weakly correlated with income, and conditioning transfers on income therefore redistributes along a dimension that is close to orthogonal to who actually gains or loses from the reform.

This echoes the argument in Ryan and Sudarshan (2022) for Indian agricultural groundwater, where feasible transfers cannot close the gap to a Pareto improvement because much of the relevant heterogeneity (farmer productivity) is unobservable to the state; the general theoretical point is also seen in Sallee (2022). The 3.1% aggregate welfare gain from pricing on existing pipes should therefore be interpreted as an efficiency measure, not as a gain the state can cash in. The gap between this efficiency measure and the distributional outcome under the best feasible transfer rule quantifies the welfare cost of the informational constraint that sustains rationing as the equilibrium policy: the government rations water by time not because the efficiency loss is small but because the redistribution that would compensate losers is outside the state's reach.

The Regressivity of Rationing and Pricing. Figure 2.9 displays the total water-related cost (bill plus annualized storage capital) as a share of weekly expenditure by quintile, together with the welfare change ΔV from each reform. Under the status quo, water consumes 12.5% of weekly expenditure for the poorest quintile versus 3.6% for the wealthiest, with storage capital alone accounting for 7.5 percentage points of that gap for Q1. Panel (a) shows that pricing without transfers amplifies this regressivity: welfare losses scale from -26.1% of weekly expenditure for Q1 to -7.3% for Q5, roughly a factor of four. Panel (b) shows that a flat lump-sum rebate on existing pipes is mildly progressive but far from sufficient: Q1 gains 2.4% while Q5 actually loses 0.7%, so the rebate returns enough of the surplus to help the poor in absolute terms. Panel (c) combines pricing with rehabilitation, collapsing storage costs toward zero and yielding strictly progressive gains ($+9.1\%$ for Q1, $+1.7\%$ for Q5). Two points are worth emphasizing. First, the progressivity in Panel (c) depends as much on the elimination of regressive storage costs as on the fiscal surplus being redistributed. Second, because the private storage capital stock represents a sizeable share of weekly expenditure under the status quo and that share falls steeply with income, intermittency itself functions as a regressive implicit tax, collected outside the fiscal system through forced private capital investment, and borne most heavily by low-income households.

Sensitivity to Inequality Aversion. The welfare ranking is robust across all values of the inequality aversion parameter considered (Figure 6a), and gains in fact *increase* in η under both pricing regimes. At $\eta = 0$ (utilitarian), the First-Best generates a gain of 5.8%; at $\eta = 1$, the gain is 6.4%; at $\eta = 5$, it is 8.1%. Pricing on degraded pipes generates 2.2% at $\eta = 0$, 3.1% at $\eta = 1$, and 5.9% at $\eta = 5$. The gains grow in η because the equity-weighted marginal value of the redistributed fiscal surplus rises faster than the equity-weighted consumer-surplus loss from the higher price: $\bar{g}_q \approx 1$ makes the price wedge roughly distributionally neutral, while the lump-sum rebate is mechanically progressive, so a more redistributive planner values the reform more.

Summary Welfare Weights. I compute the distributional summary statistics under the status quo to assess the redistributive power of each instrument. At $\eta = 1$, $\bar{g}_q = 0.97$ and $\bar{g}_\pi = 0.96$; at $\eta = 5$, $\bar{g}_q = 0.91$ and $\bar{g}_\pi = 0.87$. Both summary weights are slightly below unity and decline modestly with inequality aversion, indicating that higher-income households consume somewhat more water and benefit somewhat more from additional supply days, but that neither covariance is large. The near-neutrality of $\bar{g}_q$ is consistent with the flat volumetric Engel curve documented in Fact #2: because physical water consumption varies little across the income distribution, a uniform price change transfers roughly the same absolute amount to (or from) each household, so the *redistributive gain* from the price instrument is limited.

The *optimal wedge*, however, is not. Because $\bar{g}_q$ is below unity and demand is extremely inelastic ($|\varepsilon_{Q,p}| \approx 0.035$), the inverse-elasticity amplification $(\bar{g}_q - 1)/|\varepsilon_{Q,p}|$ is large, and the sign of the wedge $p^* - \Lambda^*$ flips relative to the status quo. Under the status quo at $n = 2$, water is subsidized by $\Lambda^{SQ} - \bar{p} \approx 0.24 \text{ JD/m}^3$. Under optimal pricing at $\eta = 1$, the wedge reverses: water is implicitly *taxed* by $p^* - \Lambda^* \approx 0.82 \text{ JD/m}^3$ above its social marginal cost, rising to 2.63 JD/m^3 at $\eta = 5$ (Table 2.4). The tax base is attractive precisely because demand is inelastic: the quantity distortion

is small, the Harberger triangle is negligible, and the revenue raised funds the progressive rebate that delivers most of the welfare gain in Table 2.5. This connects to the conclusion's framing: the reform replaces an implicit regressive tax (the deadweight storage capital borne disproportionately by the poor) with an explicit progressive one (a price wedge above SMC, rebated flat per-capita).

2.8 Conclusion

This paper provides a structural evaluation of intermittent water supply: a rationing regime affecting over a billion people worldwide. By formalizing the relationship between supply duration, physical leakage, and household adaptation through private storage investment, I embed the physical realities of water infrastructure directly into the planner's optimal pricing problem.

The central finding is that market-clearing pricing dominates intermittent supply in aggregate welfare. The gain comes through the revenue channel: efficient prices generate a fiscal surplus that, if redistributed, more than compensates for the reduction in consumer surplus. Critically, this result holds even on the existing degraded network, implying that the allocative efficiency of the price instrument does not require costly infrastructure rehabilitation.

The status quo, by contrast, generates modest revenue while forcing households to bear the deadweight cost of private storage investment. Storage capital represents a persistent implicit tax that transfers resources to no one. This is the stock-channel analogue of the classical rent dissipation result: because piped delivery eliminates the queuing channel, dissipation is forced entirely into the investment in durable storage assets.

These results do not imply that the transition to pricing is straightforward. Without transfers, all households are worse off under market-clearing prices. The price increase is large in both absolute and relative terms, and falls disproportionately on households for whom water constitutes a large share of expenditure. The welfare gain from pricing is therefore contingent on the government's ability to redistribute the fiscal surplus, a capacity that is limited in many developing-country settings where intermittent supply is prevalent. This observation reframes the policy problem. Intermittent supply is a costly accommodation of limited state capacity. The pipe's physical constraints substitute for the transfer infrastructure that efficient pricing requires. The welfare cost of this substitution is the paper's central object of interest.

Several elements of the analysis are specific to Amman, and distinguishing them from the general insights is important for applying the framework elsewhere. The quantitative magnitudes depend on calibrated parameters particular to Jordan: the leakage schedule of the municipal network, the scarcity rent anchored to regional desalination costs, and the near-unity of the distributional summary statistics $\bar{g}_q$ and $\bar{g}_\pi$ that follows from the relatively flat volumetric Engel curve. In cities where water consumption rises more steeply with income, $\bar{g}_q$ would exceed one, the equity case for subsidizing water would strengthen, and the efficiency gain from pricing would shrink. Similarly, the finding that no feasible transfer rule achieves a Pareto improvement is a consequence of the weak correlation between income and water consumption in this setting; in contexts where that correlation is stronger, progressive targeting would perform better.

The main results, however, are features of the rationing technology itself and should generalize. The decomposition of intermittency's welfare cost into a quantity channel and a smoothing channel, the inverse response of private storage to both supply duration and price, and the distinction between the consumption-weighted and duration-weighted welfare weights all follow from the structure of the planner's problem. Most centrally, the identification of private storage as a dead-weight capital stock is a feature of any setting in which durable, household-level infrastructure substitutes for public delivery frequency. This mechanism applies to water supply across much of South Asia, the Middle East, and Latin America, and to electricity outages wherever households invest in generators or batteries to compensate for unreliable grids. The general lesson is that physical rationing converts a public allocation problem into a private capital accumulation problem, and the welfare cost is the private capital stock.

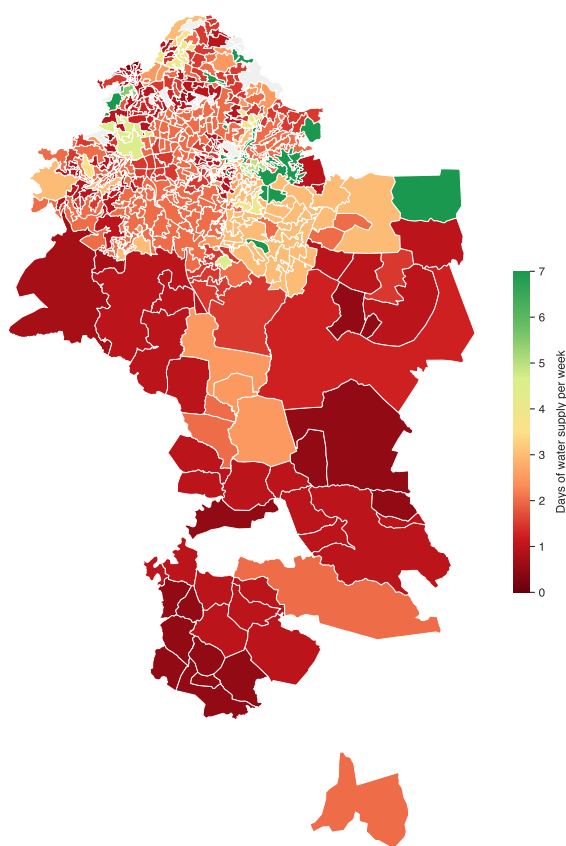
Several caveats apply. The model assumes perfect information about household types. In practice, governments may lack the information to target transfers effectively, which would reduce the feasible welfare gain from pricing. The model also abstracts from political economy constraints that may make large price increases infeasible. Finally, the calibration of the scarcity rent and infrastructure costs involves judgment.

Future work could extend the analysis in two directions. First, introducing private information about household types would allow a formal analysis of the screening properties of intermittent supply, potentially rationalizing the status quo in a mechanism-design framework. Second, modeling the political economy of water pricing including affordability concerns and public opposition would bridge the gap between the normative benchmark and implementable policy.

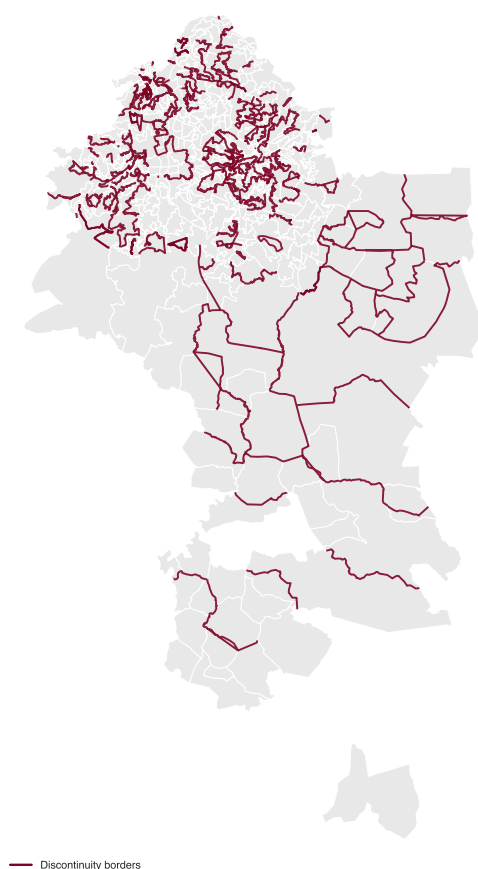
2.9 Figures and Tables

Figure (2.1) Spatial Variation in Rationing

(a) Heatmap of Water Supply Days



(b) Zone Borders with Different Supply



Notes: Panel (a) shows supply days per week across Amman's water zones. Panel (b) highlights boundaries where adjacent zones have different supply schedules, which form the basis of the regression discontinuity design.

Table (2.1) Structural MLE Estimates

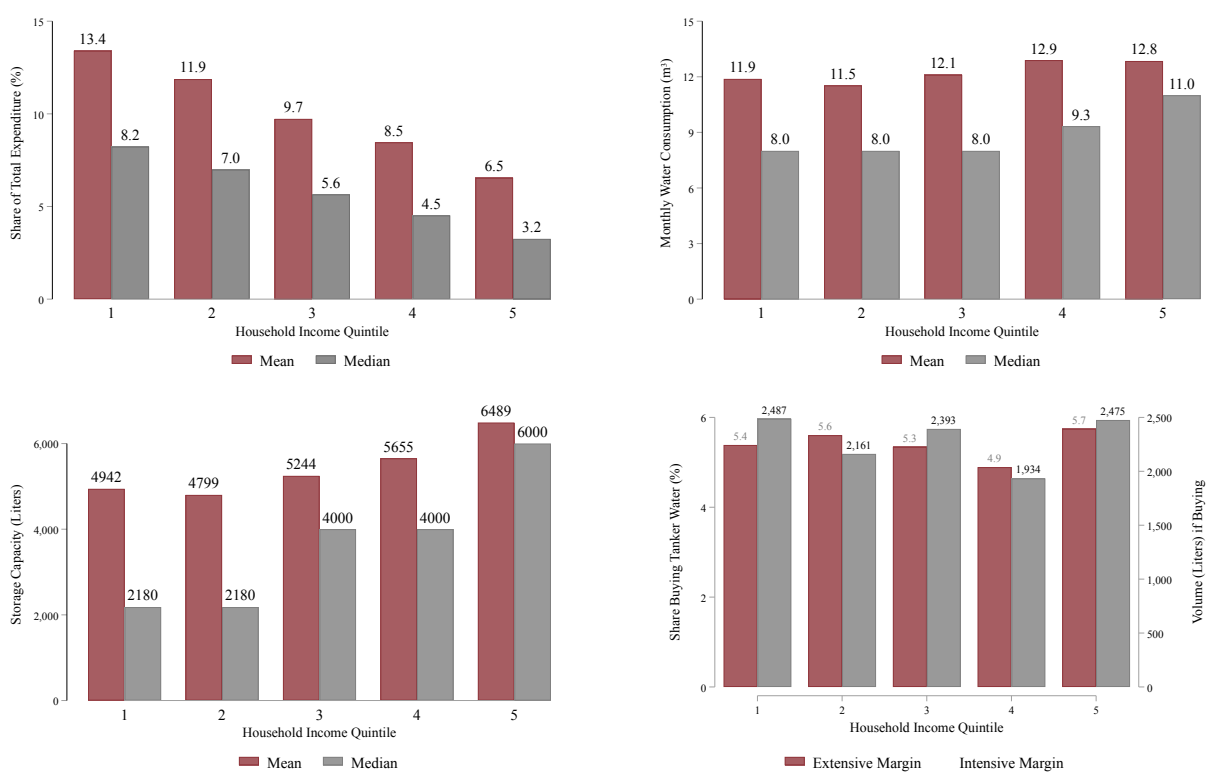
Parameter	Estimate	Std. Error	t-stat
α_0 (Intercept)	3.9904***	(0.3127)	12.76
$\beta_{\text{expenditure}}$: Expenditure (normalized)	0.3682***	(0.1153)	3.19
β_{asset} : Asset index (normalized)	-0.0641	(0.2779)	-0.23
β_{amman} : Amman (=1)	0.0379	(0.1677)	0.23
$\beta_{\text{exp} \times \text{amman}}$: Expenditure $\times$ Amman	-0.0012	(0.1706)	-0.01
β_{wave_2} : Seasonal 2	-0.1209	(0.2173)	-0.56
β_{wave_3} : Endline	-0.8160	(0.5783)	-1.41
$\hat{\sigma}_\varepsilon$	3.9109***	(0.0601)	65.11
Demand slope b (fixed)	55.0352	from billing IV	
$\hat{\sigma}_\xi$	1.7100	from variance decomposition	
Storage cost $\hat{g}$	0.8232	from MoM (Step 3)	
Log-likelihood		-5621.86	
N		2,325	

Notes: Maximum likelihood estimates of $q_{it}^* = \alpha_0 + \mathbf{X}_i' \beta - (T/b) p_{it} + \varepsilon_{it}$, where $b = 55.04$ is fixed from the billing universe IV. Standard errors from the numerical Hessian. $\hat{\sigma}_\xi$ is from the billing-matched panel variance decomposition; $\hat{g}$ is from the method of moments (Step 3). ***, **, *: significance at 1%, 5%, 10%.

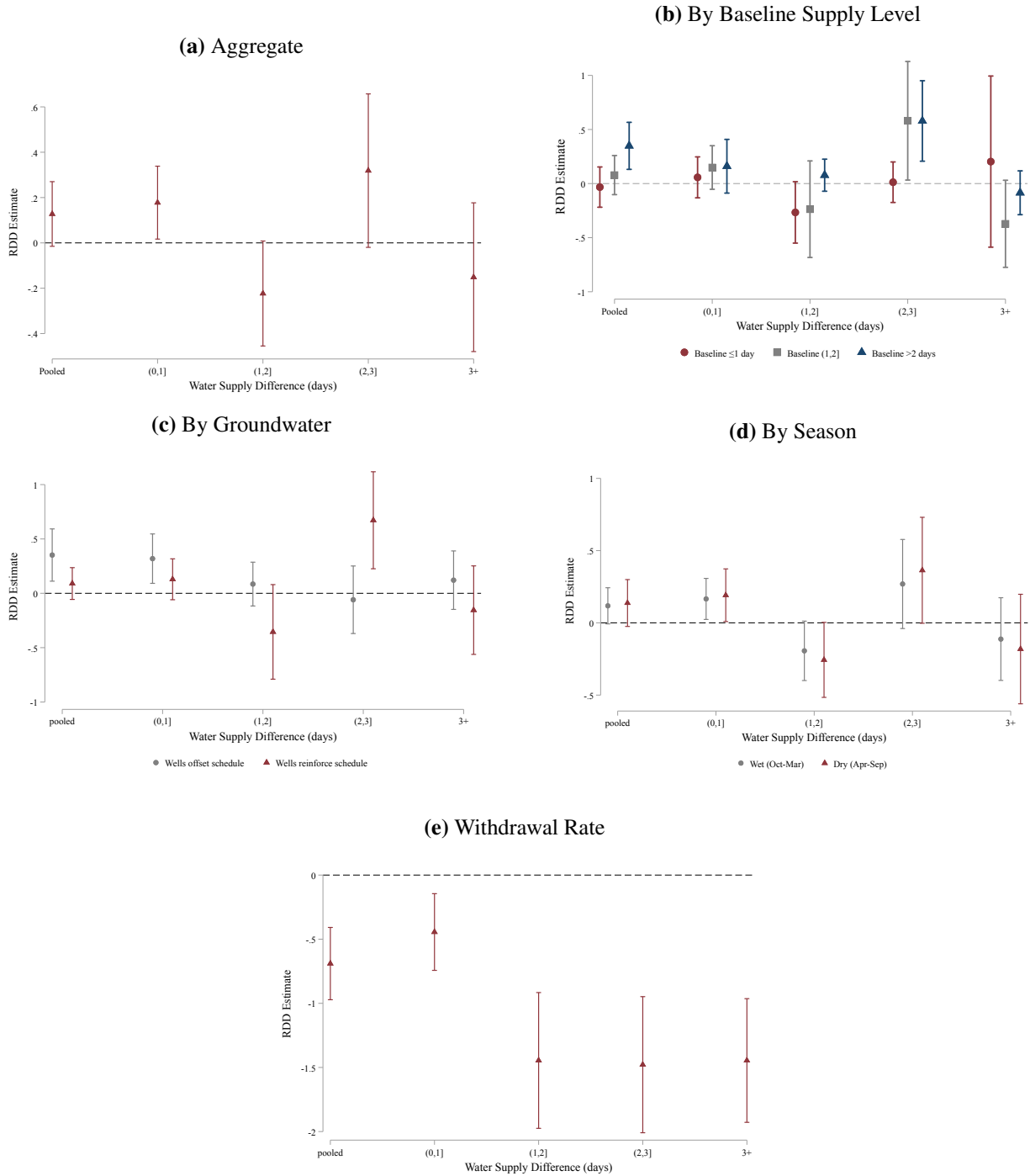
Table (2.2) Model Comparison with RDD Estimate

Metric	m ³ /week/day
Empirical RD Estimate ($\hat{\beta}_1$)	0.081
Model-Implied Decomposition	
$\pi_1 \times 0$ (Unconstrained)	0.000
$\pi_2 \times \bar{q}^*/T$ (Storage: direct)	+0.178
$\pi_2 \times dK^*/dn$ (Storage: adaptation)	-0.054
$\pi_3 \times \bar{\phi}$ (Pipe-constrained)	+0.351
<i>Total model-implied dq/dn</i>	<i>0.475</i>

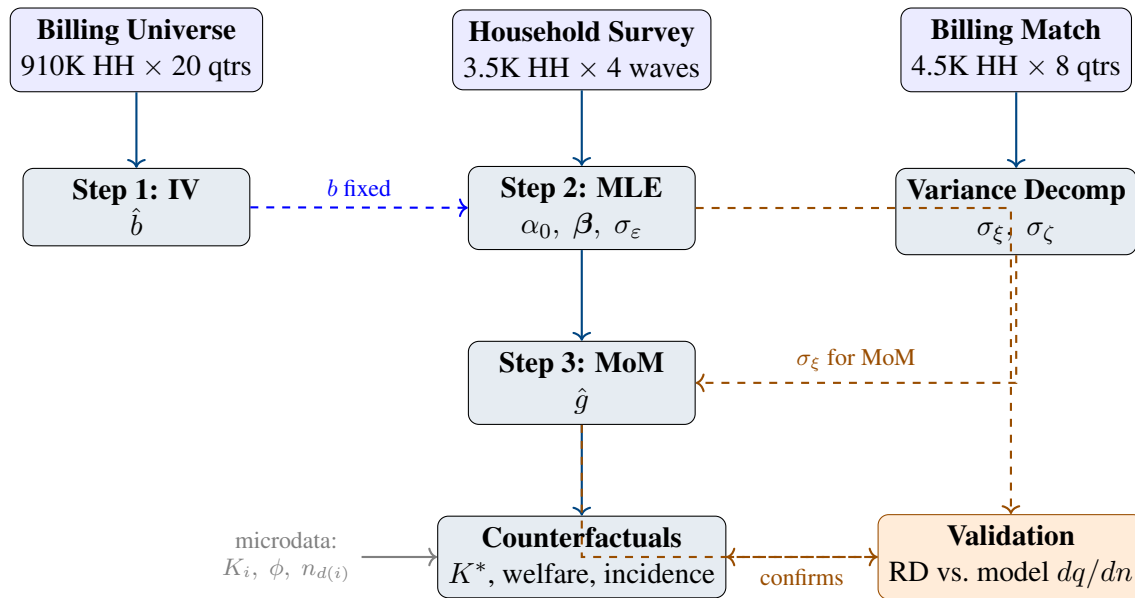
Notes: The empirical RD estimate is from a geographic discontinuity in supply schedules at zone boundaries. The reported figure divides the estimated quarterly LATE by 13 weeks per quarter and by the mean supply-day gap at the boundary (1.47 days/week) to express it as a weekly marginal effect per additional supply day. The model-implied dq/dn is constructed from estimated regime shares and structural parameters without using the RDD variation.

Figure (2.2) Patterns by Income Quintile

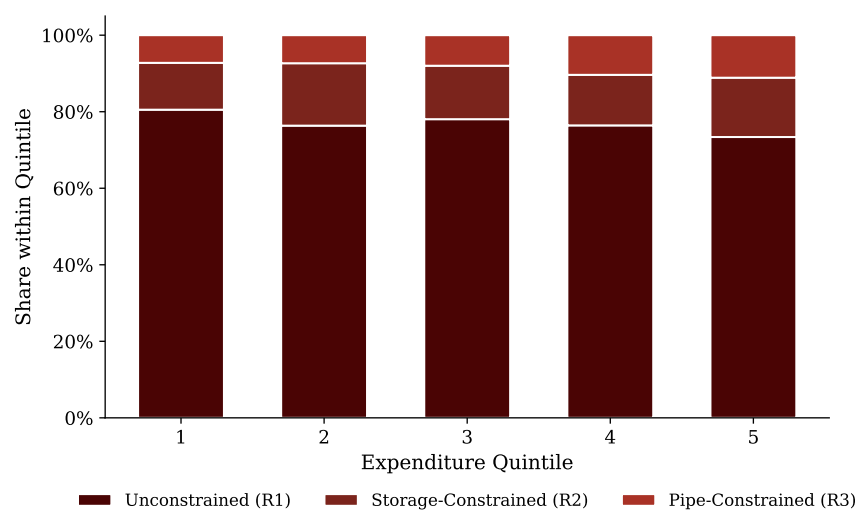
Notes: Top-left: water budget share by income quintile (mean and median). Top-right: piped water consumption levels. Bottom-left: mean storage capacity. Bottom-right: extensive margin (share purchasing, left axis) and intensive margin (volume conditional on purchasing, right axis) of private tanker water.

Figure (2.3) Impact of Supply Intensity on Consumption

Notes: Each point is a separate RDD estimate for water consumption per week with 95% CIs clustered at the water zone level. Panels (a)–(c) group borders by the *gap* in supply days between adjacent zones (e.g., the (1,2] bin pools a 1→2 day border with a 5→6 day border). Panel (d) instead stratifies by the baseline (low-side) supply level, estimating gap-specific effects separately within each baseline bin. Panel (e) shows impacts on withdrawal rate, defined as consumption per supply day. “Dry”: April–September; “Wet”: October–March. “Reinforce”: the high-supply side also has higher local well production; “Offset”: the low-supply side does.

Figure (2.4) Estimation Architecture

Notes: The estimation proceeds in three steps across two data sources. Step 1 estimates the demand slope b via instrumental variables on the billing universe. Step 2 estimates preference heterogeneity via maximum likelihood on the household survey, taking b as fixed. A variance decomposition on the billing-matched panel separates transitory from permanent shocks. Step 3 recovers the storage cost g via the method of moments, using only the transitory component σ_ξ . The RDD estimate provides an external validation of the model-implied consumption response to supply.

Figure (2.5) Regime Shares by Expenditure Quintile

Notes: Share of households in each supply regime by expenditure quintile: unconstrained ($\mathcal{R}_1$), storage-constrained ($\mathcal{R}_2$), and pipe-constrained ($\mathcal{R}_3$).

Table (2.3) Summary of Structural Parameter Calibration

Description	Parameter	Value	Source / Assumptions
Panel A: Supply and Infrastructure Costs			
Gross Supply Capacity	S	254 MCM/yr	MWI 2020 $\times$ residential share (90%)
Marginal Cost (all regimes)	c	0.41 JD/ m^3	Variable share of O&M costs
Fixed Cost (Status quo)	F_t	75 m JD/qtr	Residual from full economic cost
Fixed Cost (Rehabilitated)	F_r	100 m JD/qtr	$F_t + \text{Rehab CapEx}$
Smart Metering Cost	F_{AMI}	8.6 m JD/qtr	Amortized CapEx (1M meters)
Total Fixed Cost (Quotas)	F_q	83.6 m JD/qtr	$F_t + \text{metering}$
Leakage intensity (Status quo)	δ_t	0.45	Physical NRW $\approx 12.9\%$; $\iota_{\text{phys}}(2) \approx 1.15$
Leakage intensity (Upgraded)	δ_r	0.16	USAID rehab ratio $\times \delta_t$
Scarcity Rent	ν	0.20 JD/ m^3	Hotelling backstop (desal. ≈ 1.0 JD/ m^3)
Panel B: Demand Parameters			
Billing Cycle Duration	T	7 days	Weekly sub-cycle
Administered Price	$\bar{p}$	0.76 JD/ m^3	Volumetric rate
Inequality Aversion	η	1.0	Baseline
Demand Slope	b	55.04	Billing universe IV (Step 1)
Transitory Shock Std. Dev.	σ_ξ	1.71 m^3/wk	Billing-matched panel FE
Total Shock Std. Dev.	σ_ε	3.91 m^3/wk	Structural MLE (Step 2)
Daily Pipe Flow Rate	ϕ	4.1 m^3/day	95 th pctl. from billing data
Storage Cost Parameter	g	0.82	Method of Moments (Step 3)

Notes: Panel A: city-wide quarterly estimates for two infrastructure configurations. The marginal cost c is common; regimes differ in δ and F . Physical leakage ($\sim 12\%$ of gross) is separated from commercial losses ($\sim 30\%$); only physical leakage scales with n . Supply duration n is endogenous and generally interior.

Table (2.4) Social Marginal Cost Decomposition

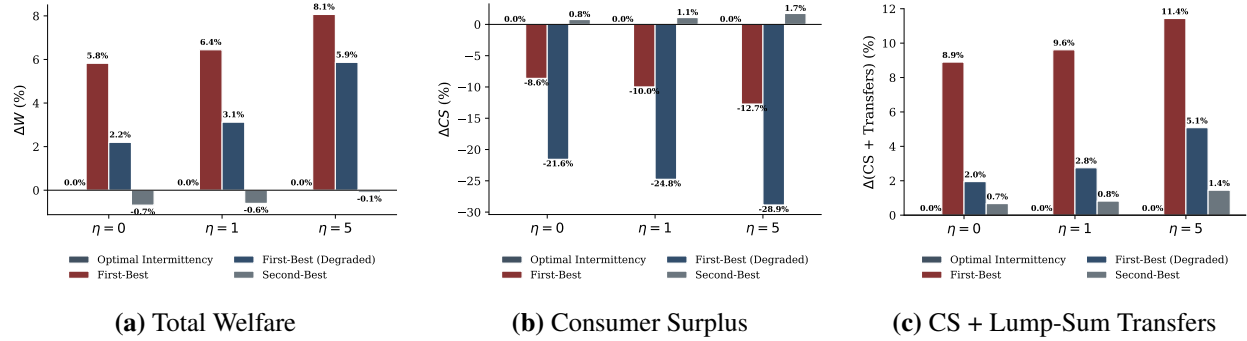
	n	p	$\iota(n)$	$c + \nu + \lambda$	SMC	$p - \text{SMC}$
Panel A: Status Quo						
Status Quo	2.00	0.76	1.64	0.61	1.00	−0.24
Panel B: Utilitarian ($\eta = 0$)						
Pricing (deg.)	3.50	5.50	1.85	2.99	5.52	−0.02
Pricing (reh.)	6.25	3.89	1.67	2.32	3.88	+0.01
Panel C: Log preferences ($\eta = 1$)						
Pricing (deg.)	3.75	6.14	1.89	2.82	5.32	+0.82
Pricing (reh.)	6.50	4.11	1.68	1.92	3.24	+0.87
Panel D: Near-Rawlsian ($\eta = 5$)						
Pricing (deg.)	4.00	6.77	1.93	2.14	4.13	+2.63
Pricing (reh.)	7.00	4.47	1.71	1.07	1.83	+2.64

Notes. $\text{SMC} = (c + \nu + \lambda_S) \iota(n)$, where $c = 0.41$ is the variable cost, $\nu = 0.2$ is the scarcity rent, and $\lambda \geq 0$ is the shadow price of the aggregate capacity constraint. At $\eta = 0$, $p \approx \text{SMC}$ (Theorem 1 with $\bar{g}_q = 1$). At $\eta > 0$, $p > \text{SMC}$: the planner prices water above social marginal cost as an equity instrument (the fiscal surplus is more than offset by the equity-weighted consumer-surplus loss).

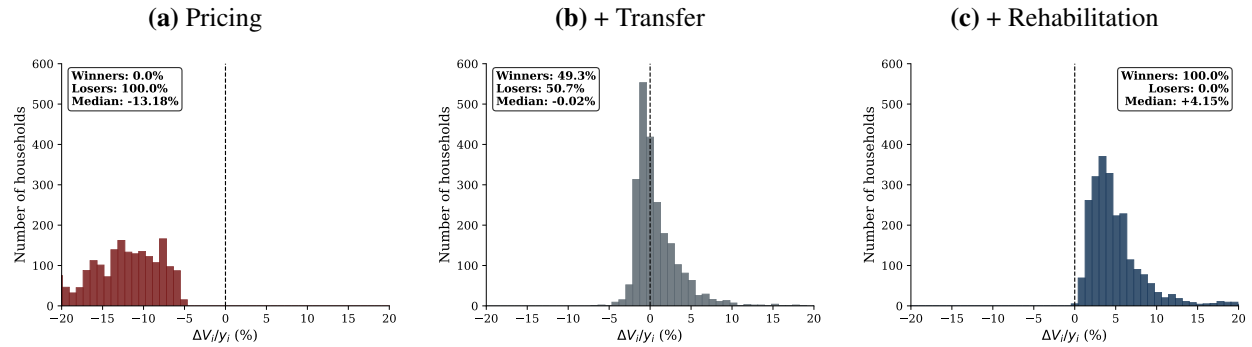
Table (2.5) Counterfactual Equilibrium Outcomes

	INTERMITTENCY		PRICING		QUOTAS
	Status Quo (1)	Optimal (2)	Existing pipes (3)	New pipes (4)	(5)
<i>Policy instruments</i>					
n (days/wk)	2.00	2.00	3.75	6.50	2.75
p (JD/m ³)	0.76	0.76	6.14	4.11	0.76
$\bar{q}$ (m ³ /wk)	∞	∞	∞	∞	4.50
<i>Infrastructure and costs</i>					
δ	0.45	0.45	0.45	0.16	0.45
$\iota(n)$	1.64	1.64	1.89	1.68	1.74
F (JD/wk/HH)	6.41	6.41	6.41	8.55	7.15
<i>Household outcomes</i>					
Mean q (m ³ /wk)	3.39	3.39	2.88	3.22	3.25
Mean K (m ³)	3.19	3.19	1.98	0.50	2.47
<i>Welfare (JD/wk, sample)</i>					
Weighted CS $\sum g_i V_i$	128,368	128,368	96,588	115,539	129,768
Net revenue $(p-\Lambda)Q$	-1,893	-1,893	33,370	23,093	-2,258
Fixed costs $-F \cdot N$	-14,904	-14,904	-14,904	-19,872	-16,613
Total $\mathcal{W}$	111,571	111,571	115,054	118,760	110,897

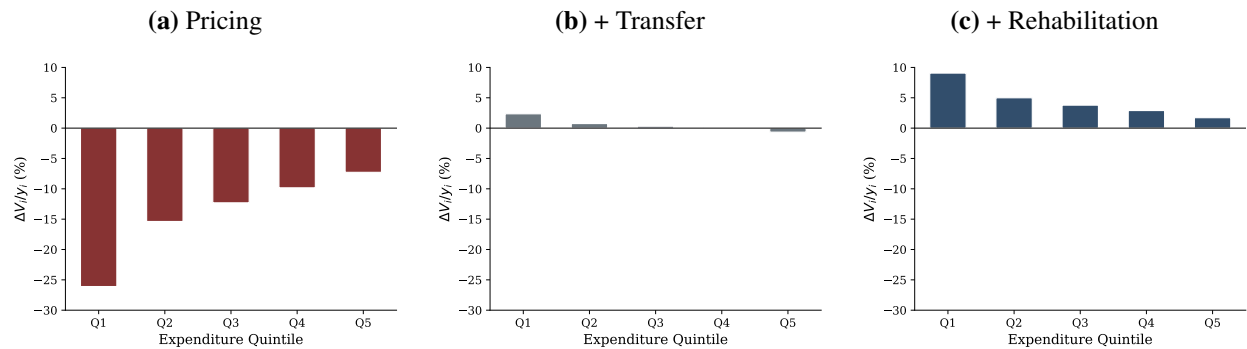
Notes. All regimes face the same resource constraint $\iota(n) \sum_i \mathbb{E}[q_i] \leq S$ with $g_i = y_i^{-\eta} / \bar{y}^{-\eta}$ and $\eta = 1$. Cols. (1)–(2) retain $\bar{p} = 0.76$ JD/m³; cols. (3)–(4) optimize p ; col. (5) retains $\bar{p}$ and adds a uniform quota. Cols. (1)–(3) and (5): degraded pipes ($\delta = 0.45$); col. (4): rehabilitated ($\delta = 0.16$). Col. (5) includes smart metering costs. V_i includes storage costs.

Figure (2.6) Welfare Decomposition: Gains Relative to Status Quo

Notes: Gains relative to the status quo by inequality aversion η . (a) Total weighted welfare ΔW ; (b) weighted consumer surplus ΔCS ; (c) consumer surplus plus lump-sum redistribution of net revenue. All values in millions JD/week.

Figure (2.7) Distribution of Winners and Losers

Notes: Distribution of ΔV_i as a share of weekly expenditure ($\eta = 1$). (c) adds a per-capita lump-sum rebate of net revenue.

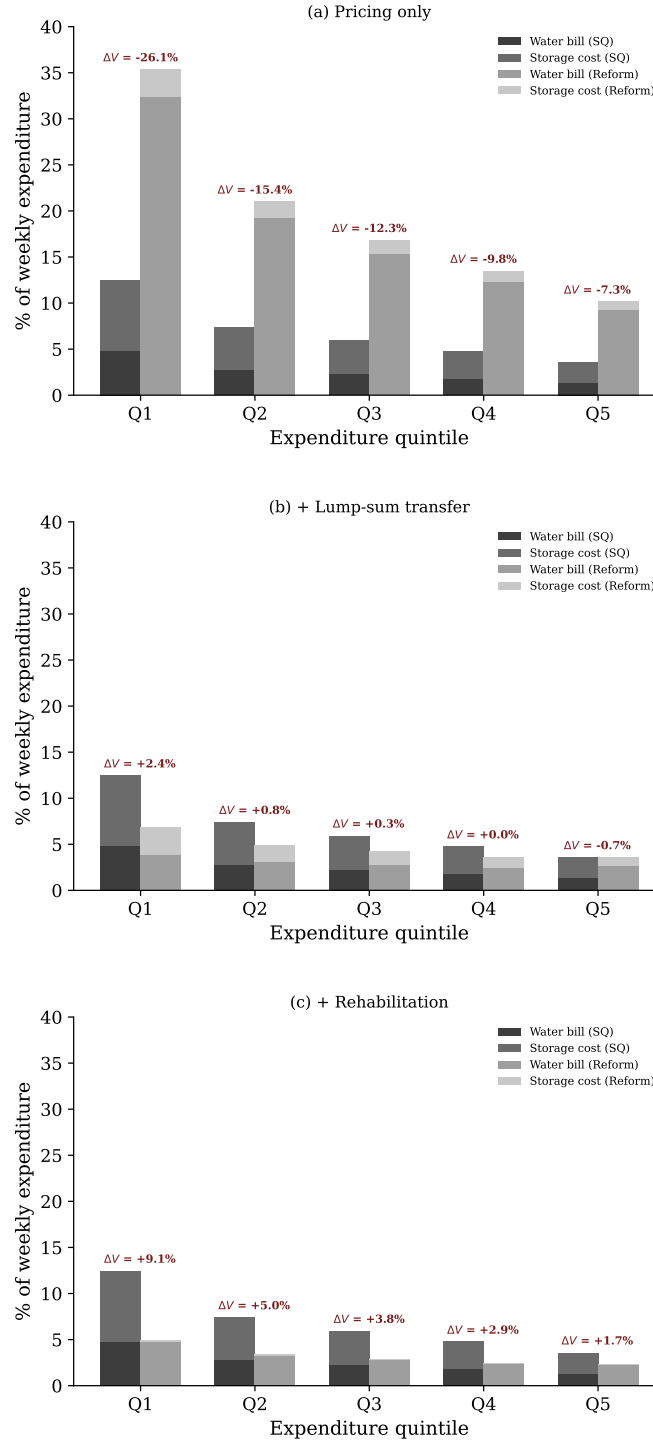
Figure (2.8) Distributional Incidence by Expenditure Quintile

Notes: $\eta = 1$. Change in indirect utility as a share of weekly expenditure by quintile. (c) adds a flat per-capita lump-sum rebate of net revenue, reversing the regressivity in (a).

Table (2.6) Distributional Effects of Pricing Reform

	Share		Mean ΔV_i		Mean y_i	
	Win	Lose	Win	Lose	Win	Lose
No transfer	0.0%	100.0%	–	-14.09	–	499
Flat	49.3%	50.7%	2.02	-1.45	445	552
Pro-rata expenditure	43.6%	56.4%	6.26	-4.37	713	335
Pro-rata storage	39.3%	60.7%	13.29	-8.16	525	483
Progressive ($1/y_i$)	39.8%	60.2%	9.03	-5.54	283	643

Notes. $\Delta V_i \equiv V_i^{\text{alt}} - V_i^{SQ}$ is the change in consumer surplus net of storage costs (JD/wk). y_i is monthly household expenditure (JD). All transfer rules distribute the same total budget $(p^* - \Lambda)Q$. $\eta = 1$ throughout with the existing infrastructure.

Figure (2.9) The Hidden Cost of Intermittency

Notes: Each panel displays two grouped bars per expenditure quintile. The left bar (dark shades) decomposes the status quo water cost into the volumetric bill ($p \cdot q_i$) and the annualized storage capital cost ($(g/2)K_i^{*2}$). The right bar (light shades) shows the same decomposition under market-clearing pricing. ΔV is the mean welfare change as a percentage of weekly expenditure. Panel (a) shows pricing on degraded status quo pipes without transfers: all quintiles lose, with the poorest bearing the largest burden. Panel (b) adds a flat per-capita lump-sum rebate of net revenue on the same degraded network. Panel (c) adds infrastructure rehabilitation, enabling supply to expand to near-continuous service. All values use endogenous K_i^* and $\eta = 1$.

Chapter 3

Housing Subsidies for Forced Migrants

3.1 Introduction

Targeted assistance programs typically direct resources to specific vulnerable groups but may create tensions with non-recipients. This trade-off is of particular concern for refugee assistance: most refugees rely on humanitarian or government aid to meet many of their basic needs, but are also at risk of experiencing social exclusion or even xenophobia from hosting country citizens. This tension could reduce public support for assistance programs, and might set recipients back by eroding important social relationships with neighbors, landlords, and employers. Supporting refugees' needs thus requires understanding both the effectiveness of assistance programs at directly raising refugees' socioeconomic conditions as well as the social costs they may incur. We study this important and under-studied trade-off through a field experiment of a refugee-targeted housing subsidy program.

Housing is a particularly important issue for refugees and other displaced people who often face precarious shelter situations, yet there remains limited evidence on how best to improve housing stability outside of camp settings (Kumar 2021; Agness 2023). This is an issue that is, unfortunately, likely to increase in importance: the total number of forcibly displaced people rose from 38 million in 2000 to approximately 118 million by 2024 (UNHCR 2023c), and is poised to increase further in the coming decades due to the adverse global impacts of conflict and climate change, especially in low- and middle-income countries (LMICs, Burke et al. 2015). Beyond the humanitarian assistance sector, there remains an active debate about how best to provide secure housing for low-income populations, with growing evidence on a variety of policies including housing vouchers, public housing construction, and transfer of home titles (Chetty et al. 2016; Kling et al. 2007; Kumar 2021; Ludwig et al. 2013).

This study examines the impacts of a randomized housing assistance program in Jordan on both the economic outcomes of Syrian refugee recipients and their social cohesion with Jordanian neighbors. Over 80% of Syrian refugees in Jordan live outside camps, making the issue of housing security particularly salient.¹ The intervention, carried out in partnership with one of the largest

¹Statistic calculated directly from data on the universe of Syrian refugees in Jordan registered with UNHCR.

humanitarian aid organizations in Jordan, randomized rental subsidies across 2,870 refugee households residing in Irbid and Mafraq governorates in Jordan. The randomization was implemented over geographic unit clusters and designed with sufficient statistical power to detect moderate impacts on important life outcomes, including living standards measures, labor market performance, and subjective well-being.

The housing subsidy covered, on average, a year of rent, combined with funding for landlords to upgrade the refugees' housing quality. Subsidies were offered for recipients' current housing arrangement, and thus their physical location (and by extension, their neighborhood) was typically held constant. This enabled the study to measure the causal effect of the housing subsidy clear of the potential neighborhood effects that a new housing contract might induce (Chetty et al. 2016). To measure effects on refugee recipients as well as Jordanian neighbors, the study collected both in-person and phone surveys over a total of three and a half years since the start of implementation, with refugee households surveyed three times and neighbors once. The data collection succeeded in attaining relatively high sample tracking rates, overcoming some challenges associated with the COVID-19 pandemic. Although 86% of the refugee sample was surveyed at least once, there was some differential attrition between the treatment and control refugee samples, and therefore we estimate bounded effects following Lee (2009) (and the text highlights the results for which both estimated bounds take the same sign unless otherwise noted). These data are analyzed largely following the econometric models and primary outcomes specified in an AEA pre-registration (AEARCT#0006141) and associated pre-analysis plan, while making note of additional and exploratory results.

Another core contribution of this study is to examine the host community reaction to refugee assistance, utilizing a detailed survey of the attitudes and experiences collected among a representative sample of the Jordanian neighbors (N=2,146) of both treatment and control households. To our knowledge, this is among the first studies to experimentally examine how humanitarian assistance to refugees affects the views of the local communities who do not directly benefit. Specifically, we examine whether refugee-targeted transfers — in this case via rental payments and housing improvements — affect host communities' policy views, altruism, attitudes, and interactions with refugees.² The study relates to a literature on transfers and potential negative spillover effects on non-recipients (Haushofer and Shapiro 2016, Haushofer and Shapiro 2018). But unlike settings where transfer recipients are other members of the local community, this study measures non-recipients' reactions to refugees — who are outsiders — receiving in-kind transfers. The most closely related studies to this one are Baseler et al. (2023) and Beltramo et al. (2024), both of which examine the effects of refugee assistance programs that also directly benefit host communities by, for example, granting them cash transfers. Of course, due to financial constraints, most humanitarian assistance is limited to forcibly displaced populations and not host community members, as in our setting. We also contribute to the literature on how immigration shapes natives' preferences for redistribution (Alesina et al. 2023, Dustmann et al. 2019) and on determinants of immigrant assimilation and social cohesion (Abramitzky and Boustan 2024).

Despite the coinciding COVID-19 pandemic, implementation of the program rolled out as

²We do this by surveying neighbors but excluding landlords, who were often involved in the program.

planned, and consequently the data show reductions in treated households' rental expenditures. However, take-up of the housing assistance was moderate at 33%, and we find that multiple factors contributed to this relatively weak first stage relationship. For one, almost a fifth of homes were deemed ineligible for assistance, typically because the implementing partner viewed the housing shelter as too precarious to support (i.e., temporary structures), and occasionally because it was determined that the residence did not need physical upgrades. There is also qualitative evidence that low take-up was at least in part due to some landlords' reluctance to make signed legal commitments to refugees for the lease, and with the implementing partner for the necessary construction. This is a setting in which most rental contracts are informal and landlords have ample discretion over their terms and a largely free hand to evict tenants. Despite the appeal of guaranteed rent for a year plus funding for housing upgrades, some landlords preferred not to "bind" themselves to the program and the particular recipient refugee household currently residing in their property.³

One of the study's main empirical findings is that we detect no significant positive impacts of the housing assistance program for refugees along a range of pre-specified primary outcomes, with the exception of housing expenditures where there is the expected (and somewhat mechanical) drop in spending. Beyond housing expenditures, the other primary pre-specified outcomes include an aggregate housing quality index, total household consumption, respondent mental health, and child socio-emotional well-being (measured using the standardized Strengths and Difficulties Questionnaire, SDQ). In fact, the analysis unexpectedly shows that the socio-emotional well-being of children in treated households decreases substantially, by 0.34 standard deviations on average; we return to the interpretation of these patterns below. Due to the non-compliance noted above, estimated instrumental variable (IV) treatment on the treated (TOT) effects are relatively imprecise. But the primary analysis, which pools across multiple rounds of follow-up survey data, does allow us to reject the existence of some moderate positive treatment effects.

A second main finding is that the refugee-targeted housing subsidy program led to a meaningful and statistically significant deterioration in community perceptions towards Syrian refugees among the Jordanian neighbors of treatment households. We estimate a 0.33 standard deviation unit decrease in an index of their social attitudes and perceptions of refugees. The analysis also shows that neighbors become better informed about how much assistance Syrian refugees receive. There is no evidence of effects on neighbors' own economic outcomes. We speculate that information about the substantial transfers was observed by Jordanian neighbors and led to tensions in some cases, perhaps especially given that these neighbors also live in the same relatively poor neighborhoods as the refugees and are typically not well-off themselves. The easy visibility of assistance in this case – including construction activity in the treated housing units or the qualification visits from the implementers – may have raised the salience of the intervention relative to some other forms of aid (i.e., mobile money transfers or health care subsidies) that are easier to keep private.

The estimated null and even negative results for most pre-specified primary outcomes are in contrast to the predictions of experts. Building on an increasingly common practice in economics and other social sciences (DellaVigna and Pope 2018; DellaVigna et al. 2020), the research team gathered forecasts from both researchers and policy experts regarding the primary impacts of the

³The data underlying these findings is from the implementing partner's Integrated Assessment Shelter Analysis.

program. Forecasters generally predicted positive but modest program impacts on refugees, and predicted zero effects on average for Jordanian neighbors. They were thus somewhat more optimistic about likely program impacts than the estimated effects, though in most cases we are unable to reject equality between the estimated average impacts and mean forecasts.

Disaggregating the data by survey round assists in understanding the predominantly null impacts on consumption and adult well-being. In the short run the study shows that the program was implemented successfully. Treated households experienced meaningful gains in housing quality and financial stability, and reported increased access to clean water and less reliance on loans. These findings indicate that program investments were able to achieve some of their immediate goals, specifically, in terms of improving the housing quality experienced by recipients, and preventing them from accumulating more debt (much of which is rental related debt in this population). In fact, on average the control group holds 4 to 5 months of rental debt, so potentially reducing this burden could be a meaningful benefit for treated households.

Yet, both food security and self-reported health worsened: specifically, household hunger increased for several measures and more household members exhibited COVID-19 symptoms. We find evidence that these negative effects follow from unanticipated redistribution of aid and household recomposition. In particular, we document a meaningful and statistically significant reduction in food aid that treatment households report receiving (from both formal and informal sources), roughly equivalent to 10% of monthly pre-pandemic income. There is also suggestive evidence that treated households received an inflow of additional household members (especially adolescents), further diluting program benefits in per capita terms. This is in line with findings from Türkiye, where an evaluation of the largest cash transfer program for refugees in the world found a net movement of children into beneficiary households (Özler et al. 2021). Although we are only able to measure the effects of the program on treatment recipients, it is possible that the program might have temporarily benefited these movers.

Data from the second round of surveys (collected shortly after assistance had ended) indicates that treated households continued to report reduced housing expenditures as well as increased savings, but at the same time reported a significant decrease in subjective well-being. In the final survey round, one to two years after assistance had ended, the only significant treatment effect that survives a multiple testing adjustment is the substantial decrease in child socio-emotional well-being noted above.

The effects of the refugee housing assistance program over time suggest a possible link between the negative impacts on child well-being and neighbor attitudes. While all other treatment effects on refugee households dissipate by the final round of surveying (conducted more than a year after assistance had ended), the negative impacts on child outcomes persist, and the negative impacts on neighbors' perceptions towards refugees are collected at a similar time point (about a year after program implementation had ended). For example, two components of the neighbor social attitudes index that are statistically significant capture whether the children of the Jordanian household have Syrian refugee friends, and whether the adult Jordanian neighbor has close Syrian friends. Taken together, the timing of child effects and the data from neighbors offer suggestive evidence that the deterioration in child well-being may be driven by worse social relations with the treated refugee households' Jordanian neighbors.

In all, the empirical findings indicate that this direct housing assistance program offered limited short-run economic improvements that then dissipated following the program, while negative psychological and social cohesion effects proved more lasting. Not only was take-up of the housing assistance quite low (at around one third), but impacts were minimal or even negative for some measures. Given the low take-up, we also predict program compliance and compare individuals in the treatment and control groups with similar predicted likelihoods of treatment to assess whether the findings are potentially driven by violations of the IV exclusion restriction, as discussed in Section 3.7. We find little evidence that such a violation explains the negative effect on children's socio-emotional outcomes. In contrast, the effects on neighbor attitudes do appear to be driven by treatment assignment itself (rather than program take-up alone), suggesting that merely knowing that refugees were offered the program was sufficient to impact attitudes among neighbors; as a result we also present reduced form intention to treat (ITT) estimates.

The negative response from neighbors parallels existing research on targeted transfers, which has documented negative psychological impacts on nearby non-recipients of cash transfers in some cases, although the evidence is notably mixed (Haushofer and Shapiro 2016, Haushofer et al. 2015, Baird et al. 2013, Egger et al. 2022, Lehmann and Masterson 2020). Unlike the above studies, this paper finds that tensions among non-recipients persist more than a year after the transfer implementation period, and – more speculatively – that these negative effects might even undermine the well-being of program recipients and their children via reduced social ties, highlighting the fragility of social cohesion between refugees and host country citizens (Zhou 2019).

If programs like this one are in fact not improving refugees' lives, an urgent question revolves around the feasibility of other policy approaches. A growing body of literature demonstrates that direct cash transfer programs yield meaningful improvements for households, at least while assistance is ongoing (Hidrobo et al. 2014; Özler et al. 2021; Quattrochi et al. 2022; Moussa et al. 2022; Aygün et al. 2024; Haushofer and Shapiro 2016), and cash is one immediate alternative that would likely be less observable to non-recipients. In line with this, and echoing the largely null evaluation results, a large majority (70%) of refugee respondents in this study stated in the endline survey that they would have preferred cash assistance of equivalent value to the rental subsidy. At the same time, such cash assistance would arguably not do much to improve the structural impediments to economic and social integration facing refugees in most settings.

3.2 Background and Context

Syrian Refugees in Jordan

As of 2023, the Syrian crisis was one of the largest displacement crises in the world, resulting in 6.8 million internally displaced Syrians and another 6.5 million Syrian refugees abroad (UNHCR 2023c). The war began in 2011 with pro-democracy demonstrations against President Bashar al-Assad, whose government responded with lethal force against the protesters, leading to escalating violence that drove the country into civil war. Considering Syria's 2010 population was 21 million, more than half of its population was displaced during the study period (UNHCR 2023b).

Jordan hosts roughly 650,000 registered Syrian refugees, accounting for about 6% of its population of 11.1 million people (UNHCR 2023a). While the Jordanian government is not party to the 1951 Refugee Convention nor its 1967 Protocol, which ensure dignity and rights for refugees in signatory countries, it does uphold the principle of *non-refoulement*, which mandates that refugees cannot be forcibly returned to countries where they face persecution.

In response to the influx, the Jordanian government opened multiple refugee camps in coordination with UNHCR, including *Za'atari*, *Azraq*, and *Mrajeeb Al Fhood*. Despite the availability of refugee camps, individuals were granted free mobility and, in fact, over 80% of Syrian refugees currently live in local communities and not camps (UNHCR 2023a).

More broadly, Jordan has accepted thousands of refugees from conflicts in the region over the past century, including refugees from Iraq, Lebanon, Yemen, Sudan and Somalia. Additionally, Jordan has hosted particularly large numbers of Palestinian refugees since 1948, many of whom were granted Jordanian citizenship. In fact, surveys on Jordanian attitudes toward Syrian refugees suggest that the majority feel generally positive about refugees and continuing to host them while also supporting some restrictions, such as camp confinement and work permit quotas (Alrababa'h et al. 2021). Jordan's historical role as a host country for refugees likely contributed to the signing of the 2016 Jordan Compact, which funneled international financial support to Jordan and liberalized work permits for Syrian refugees.

Humanitarian Shelter Program

In Jordan, there is limited secure and affordable housing available for Syrian refugees. The influx of Syrian refugees substantially increased the demand for housing in Jordan immediately after it began. Estimates indicate that the refugee crisis added an annual demand of approximately 86,000 housing units, on top of the pre-existing local demand of 32,000 units (Rozo and Sviatschi, 2021). Despite the increase in housing supply since 2016, the UNHCR Quarterly Analysis of Refugees in Jordan found that in 2022, 25% of Syrian households perceived a current threat of eviction.⁴ Additionally, one in four refugees applying for rental assistance from a major Jordanian humanitarian organization had moved at least three times in the previous year (NRC 2015), demonstrating the widespread precariousness of their housing situation. Refugees often rely on humanitarian assistance to manage rising rental costs, but these funds are insufficient to cover the full expense, and hence 70% of Syrian households living in local communities reported having rental debt (Hagen-Zanker et al. 2018; Lehmann and Masterson 2014). Frequent relocation due to rental affordability issues and a lack of stable rental contracts may also make it challenging for refugees to maintain the necessary documentation to access public services (NRC 2015).

Given this situation (and in addition to the aims that a cash transfer program would typically have), this program focused on formalizing refugees' housing arrangements and increasing their stability. Furthermore the intent was also to use the program to facilitate social cohesion between refugees and hosts by contracting with landlords directly and guaranteeing their rent. The program

⁴Determined using the Vulnerability Assessment Framework, a bi-annual survey of registered refugees in Jordan.

focused on Syrian refugees in the country's northern governorates of Irbid and Mafrqa, which are adjacent to the Syrian border and contain a large share of the country's displaced persons.

The overall Shelter Program is composed of several different assistance modalities, including projects focused on water and sanitation, eviction protection, and "inclusion kits" for individuals with disabilities. This study evaluates the impacts of a housing shelter modality, which we will refer to in what follows as the Housing Subsidy Program (HSP).⁵

Under HSP, the implementing partner directly compensated Jordanian landlords for rent on behalf of their refugee tenants, with the refugee household receiving rent-free housing for 9 to 18 months, depending on negotiations with the landlord. These negotiations were one source of variation in how the subsidies were spent: some landlords agreed to participate in the rental subsidy program for 12 months while others agreed to a shorter or longer period. The program was provided to refugees in their existing rental arrangements, building on established tenant-landlord relationships.

The program offered several secondary benefits in addition to direct rental payment. For instance, the implementing partner paid for physical housing improvements, such as door and lock installation, roof and wall construction, and mold removal.⁶ The program thus benefited not just the Syrian refugee tenants but also their Jordanian landlords, as well as the overall Jordanian housing stock. Together, the rental assistance plus the construction improvements were valued at USD \$2,200 per unit on average, a considerable sum in this context, where per capita income was roughly USD \$4,100 in 2019.⁷ The program also formalized refugees' tenancy agreements: Jordanian landlords and Syrian refugees signed standard tenancy agreements in line with Jordanian law, and the implementing partner and the landlord entered into a contract for the construction work on the property. Finally, the program required landlords to not raise rent for an additional year following the end of assistance. This could have possibly generated further housing savings for treatment households since rental prices rose by an average of 3% per year from 2010 to 2025.⁸

The Covid-19 pandemic occurred during the program implementation period. While Jordan experienced a strict lockdown from mid-March until May 2020, the lockdown lasted approximately 6 weeks and conditions returned to normal soon after (which accounts for less than 10 percent of the average program duration). Despite the unexpected impacts we find, we have no evidence that the program was not delivered as planned among those who took up treatment.

⁵The pre-analysis plan proposed the evaluation of the far less generous energy subsidy program as well. That will be the subject of future research.

⁶The Government of Jordan regulates housing quality through a combination of national building codes, municipal oversight, and public housing initiatives. The primary authority responsible for setting and enforcing these standards is the Ministry of Public Works and Housing. Despite these regulatory frameworks, challenges persist, particularly in informal settlements and refugee camps where enforcement is limited.

⁷This is according to the World Bank's [World Development Indicators](#).

⁸This is according to data from the Jordanian Department of Statistics, at the following link: <https://dosweb.dos.gov.jo/economic/price-indices/table-price-indices/>

3.3 Experimental Design and Data

This study employs a Randomized Control Trial (RCT) with 2,870 Syrian refugee households, and surveys of 2,146 neighboring host-community households, to evaluate the direct impacts of the housing subsidy program on refugees as well as the spillover effects onto host community neighbors. The study additionally assesses the accuracy of forecasts of program impacts from research experts and policymakers gathered via the Social Science Prediction Platform. The research design and primary hypotheses were pre-registered before program roll-out began.⁹ Given the ethical complexities involved in conducting research in humanitarian settings, the research team has written an ethics appendix detailing the decisions made to maintain a high standard of ethical conduct in study design, implementation, and data collection (see [Ethics Appendix](#)).

Study eligibility. Participants were identified by the implementing organization, which disseminated information about the program through their networks, resulting in a pool of refugee applicants that exceeded the sample they had resources to assist. Given this over-subscription, applicants' opportunity to receive HSP assistance was randomized. The implementing partner then conducted a standardized vulnerability assessment. Households found to be in the most vulnerable 10% of the sample were guaranteed access to the program (i.e., were not in the randomization), while the least vulnerable 10% of households were excluded from the program (and the randomization). The remaining 80% of households with more typical levels of vulnerability were then eligible to receive the program if residing in a treatment community, as described below.¹⁰

Treatment Randomization: The housing subsidy program was randomized geographically at the community level.¹¹ In the first step, 158 communities in Irbid and Mafraq were randomized into treatment or control for HSP assistance, stratifying on governorate and district population quartile. One third of communities were randomly assigned to treatment, while the remaining two thirds were randomly assigned to control. All eligible applicants living in the treatment communities were assigned to treatment. In all analysis below, error terms are clustered at the community level. There were several reasons for the cluster randomized design, including the ability to streamline implementation (and thus lower program costs), to reduce tensions among refugee households assigned to treatment versus control, and to improve the analysis of treatment spillovers onto host community members by boosting the local saturation of program assistance, thereby making it more salient to neighbors.

Refugee Sample: The survey sample consists of all refugees assigned to treatment and an equal number of randomly selected households in the control communities who had applied for the program. Figure 3.1 illustrates the location of the treated and control communities in Jordan. These respondents were surveyed even if they moved.

Jordanian Host Community Sample: A key concern with refugee assistance programs is the

⁹See [AER RCT Registration Number AEARCTR-0006141](#).

¹⁰To minimize any potential backlash from individuals assigned to the control group, the survey firm is unrelated to the program implementer, enumerators have no information on household treatment status, and the survey questions make no explicit reference to the program.

¹¹Communities correspond to the Jordanian government's administrative unit of "localities". Communities in Mafraq have 3,866 people on average, while those in Irbid are somewhat larger at 14,626 individuals on average.

potential for backlash from host communities. The HSP was intentionally designed to mitigate such effects among landlords, by disbursing funds directly to landlords for housing improvements, as noted above. The social cohesion component of the design focuses instead on measuring the impacts on attitudes of Jordanian neighbors in the community, who are were not eligible for HSP assistance, rather than on landlords. The Jordanian neighbor sample consists of 1,455 Jordanians living near the treated home location at the time of implementation. The sample was identified in 2022, a year after program aid had ended, through in-person surveying.

Enumerators used a randomized algorithm to select a Jordanian neighbor living close to the original home of the refugee participant (from among the set of nearest neighbors).¹² This selection was based on the refugee participant's home location at the time of study randomization, irrespective of any subsequent moves. The algorithm was successful at identifying and surveying a close neighbor: the median distance between the selected neighbor and the original refugee participant's home is just 63 meters (roughly 200 feet). For households living in apartment buildings, the set of nearest neighbors generally included some other apartments in the same building. For households living in peri-urban or rural areas, the nearest neighbors were often next door, two doors down, etc. The median distance for refugee dwellings in an apartment building is 46 meters, and 70 meters for stand-alone dwellings. It takes approximately one minute to walk 60 meters.

Data on refugee outcomes

The study collected three rounds of surveys tracking participants' outcomes during ("midline"), immediately after ("endline"), and 1.5 years after all assistance was delivered ("follow-up"). The outcomes to be examined were specified in detail in the pre-analysis plan and divided in three main groups including:

1. *Proximate outcomes*: This comprises outcomes that were expected to change directly due to program implementation, and include housing expenditures and a housing quality index. As described above, HSP households received not only full rental subsidies for 9 to 18 months but also physical upgrades to the home environment. These proximate outcome measures can be seen as capturing successful program implementation.

To measure physical improvements to the shelter, we constructed an index of self-reported floor, roof, and wall quality; electricity and water access; and crowding. Housing expenditures were measured in two ways: First, in the midline phone survey, respondents reported their total monthly housing expenditures, including the payments provided by the implementing organization. Second, in the endline in-person survey, once assistance had ended, participants reported their monthly out-of-pocket housing expenditures (excluding the amount paid by the implementer). The latter is our preferred measure of the direct effect of the program on housing expenditures, because it allows us to test whether treated households experienced rental savings. Because the endline measure was collected after assistance

¹²Recall randomization was at the locality level, so there is little potential for contamination, or a Jordanian neighbor living close to both a treatment and a control refugee household.

ended, any observed expenditure reductions would most likely be an underestimate of the benefits experienced during the program. Recall that HSP also required the landlord to not raise rent for an additional year after the assistance ended, which could help account for any persistent effects on housing expenditures in the treatment group.

2. *Primary well-being outcomes:* This category includes three main outcomes that seek to capture refugee household well-being, including household consumption, mental health (measured through the Center for Epidemiological Studies Depression Scale, CES-D), and an index summarizing the 25-item child Strengths and Difficulties Questionnaire (SDQ) that measures emotional and conduct problems, inattention, peer relations, and prosocial behaviors. Both the CES-D and SDQ are validated tools that have been utilized in a variety of international contexts (Park and Yu 2021, Woerner et al. 2004). In each of the three survey rounds, the primary respondent completed the CES-D depression screening. The SDQ was completed by an adult respondent regarding a randomly selected child aged 3 to 8 years old in the endline survey, as well as for the same child in the one and a half year follow-up. Total household consumption was measured only at endline (which was collected in person).
3. *Secondary outcomes among refugees:* These include a broader set of variables that were grouped into 14 broad families in the PAP including: (1) dwelling characteristics and household structure, (2) consumption and expenditure, (3) financial participation, (4) earnings, labor, and occupational choice, (5) migration, (6) physical, mental health, and sleep, (7) marriage and fertility, (8) child outcomes, (9) social capital, (10) political attitudes, (11) time use, (12) education and cognition, (13) behavioral games and preferences, and finally, (14) specific COVID-19 related outcomes. The three primary outcomes noted above are a subset of these measures. Certain outcomes were collected in each of the three survey rounds, for instance, the food security measures (the number of meals eaten and the frequency of going to bed hungry). Respondents also reported information on child school attendance, and completion of learning activities when schools were closed due to the COVID-19 pandemic. In the midline survey collected in 2020 (during the pandemic), the physical health outcomes included additional questions focused specifically on COVID-19 symptoms and treatment but these were dropped in later rounds as the pandemic had eased.

Measures of social cohesion among neighbors

We pre-specified four primary outcomes regarding the attitudes of Jordanian neighbors:

1. *Interpersonal social attitudes and perceptions:* This was assessed using an index derived from questions about social ties between the Jordanian respondents and Syrian refugees, attitudes regarding social proximity, and opinions on refugees' contributions to society.
2. *Economic attitudes and perceptions:* Focused on Jordanians' perceptions of the impact of Syrian refugees on the Jordanian economy, based on survey questions.

3. *Altruism*: Measured through a dictator game, this outcome gauged altruistic behavior towards refugees, with real monetary incentives provided for a random subset of the sample to promote genuine responses.¹³
4. *Policy preferences*: This index summarized the respondents' stances on various policies related to Syrian refugees, including their living arrangements (e.g., a hypothetical requirement to live in camps) and employment rights.

In addition, the neighbor survey collected comparable measures of economic and psychological well-being to those measured among the refugee sample, including housing expenditures, total consumption, and mental health, to assess program spillovers along these dimensions.

Data sources

The study employs four distinct data sources, with the first obtained from partners at the implementing organization, and the other three original survey data sets collected by the research team:

1. *Baseline administrative data*: The implementing organization collected baseline data from program applicants to establish their eligibility for the HSP. This assessment form collected information on household demographics, housing quality, health and disability, employment, and education, which were used to construct a housing vulnerability index employed as a baseline covariate in the empirical analysis.
2. *Surveys collected among refugee households*: The study collected surveys during program implementation (the midline survey), immediately after (endline survey), and 1.5 years after all assistance was delivered (the follow-up survey) (see Figure 4.1). In the midline survey, 1,619 participants were surveyed by phone in 2020, during the first year of the COVID-19 pandemic. The endline survey in 2021 collected in-person data from 1,534 participants shortly after HSP assistance had ended. Finally, in the 1.5 year follow-up collected in late 2022 to early 2023, 1,444 participants were reached for a phone survey. In total, the three rounds of data cover a wide range of outcome families including those listed out above and several others; these are described in detail in the appendix and associated pre-analysis plan.
3. *Neighbor surveys*: A round of in-person data collection of the Jordanian neighbors of the study participants was gathered in 2022 roughly one year after the program had ended. In total, 1,455 Jordanian neighbors of refugee participants were surveyed.¹⁴
4. *Forecast short surveys*: Two rounds of brief 10 minute forecasts were collected to gather the predictions of both academic and non-academic experts on the likely impacts of the program on the primary outcomes. Each round of predictions collected information for approximately

¹³A random one third of respondents were offered financial incentives whereas the others were told it was hypothetical. We cannot reject that the same choices were made in both groups on average.

¹⁴The study also attempted to survey the study participants' landlords, but faced extremely low willingness and therefore had to abandon this approach.

60 individuals, with the first round focused on impacts on study participants, and the second on impacts among neighbors.

Figure 4.1 depicts the data collection timeline and the response rate of each of the surveys. Overall, 86% of the main refugee study sample was surveyed at least once: 80.3% were surveyed at midline, and at least 75% were surveyed in each of the endline and follow-up rounds. Among the treatment group, survey completion was 6 to 8.6 percentage points higher in each survey round, as shown in Table C.2, and this differential attrition is statistically significant. As such, beyond the pre-specified TOT results presented, the analysis was expanded to include Lee bound estimates of the treatment effects (following Tauchmann 2014). We discuss this exercise and its implications in the empirical strategy section below.

Descriptive statistics

We present characteristics of the sample using baseline vulnerability assessment in Table C.1.¹⁵ Several patterns are noteworthy. First, a quarter of respondents likely have a disability, calculated using the Washington Group Short Set on Disability.¹⁶ Sharing a home is common, likely as an economic coping mechanism, with on average of 1.3 families living in the same house. Housing quality is low, with most houses in some state of disrepair or incomplete construction.

Table C.1 also illustrates descriptive statistics for the treatment versus control households to assess balance across groups, where the third column presents a mean difference test. Respondents across the study arms are largely balanced by age (with an average age of 34 years), marital status (84% married), and incidence of a disability (Panel A). There is a slight imbalance in the share of respondents who are female, with the proportion slightly higher in the control group. Moreover, average household size is just above five members, with nearly three children on average, and these are balanced between treatment and control, as is the number of families per residential unit (Panel B). Refugee households in the sample face challenging and precarious housing conditions, but these characteristics are generally balanced across groups (Panel C). For example, only 66% of households have access to piped water, 22% have functional windows, and 44% completed floors. While the majority of refugees plan to stay in the same housing unit (92%), a large share have moved shelter in the last year (with an average of 0.44 and 0.47 moves in the control and treatment groups, respectively).

Across all baseline characteristics presented in the table, the p-value on the hypothesis that all differences are zero is 0.102 (using a Chi-squared test), not quite significant at traditional confidence levels.¹⁷ This finding together with the fact that there are few economically meaningful

¹⁵The table only includes data for 1,619 out of the 2,017 households sampled for the study since the implementing partner only facilitated data on this assessment for the individuals found in the midline sample. The table also presents statistics for gender and age, which are taken from the midline survey.

¹⁶We use the recommended cut-off (level 3) to classify respondents as being disabled or not based on six domains: seeing, hearing, walking, cognition, self-care, and communication.

¹⁷Another way to assess imbalance is by using the baseline covariates to predict each of the primary outcomes, to then examine whether the predicted value of the outcomes varied by treatment. We test and find no significant relationships between any of the predicted primary outcomes and treatment status (Appendix C, Table B.8).

differences across treatment arms in terms of respondent, household and shelter characteristics – of the 18 covariates examined, only two exhibit a significant difference at 95% confidence – indicates that the study randomization generated largely comparable groups of households.

Forecasts of Program Impacts

The forecasters were mainly researchers working on topics in migration and development economics, or those with expertise in humanitarian program implementation, and were asked to predict effects on the program's primary outcomes. For round 1, a total of 61 individuals, comprising 37 researchers and 24 non-researchers, responded to surveys about the program's effects on refugee well-being. For round 2, 63 respondents, including 53 researchers and 10 non-researchers, provided predictions on the effects on neighbors' social cohesion responses. Gathering forecast data allows us to assess whether estimated impacts were in line with the prior beliefs of research and policy experts. Table 3.2 presents the mean predicted effects, as well as the interval from the 10th to 90th percentiles of predictions. Generally, forecasters anticipated modest improvements in refugee well-being, particularly regarding housing outcomes, while predicting no average impact on the social cohesion measures collected among neighbors.

3.4 Empirical Strategy

First stage compliance with program assignment

Table C.2, Panel B presents the first stage analysis and the HSP take-up rate of 33%. While statistically significant, the compliance rate is lower than expected, especially given the substantial funding offered by the program. This is likely attributable to several factors, including titling issues, uncooperative or reluctant landlords (who may have feared being locked into long-term contracts with refugee tenants, as noted above), and the implementing partner's policies that led some homes to be excluded due to their condition. Low take-up raises the possibility of an exclusion restriction violation, if the people assigned to treatment but who did not end up getting the program somehow drove the results; we evaluate this possibility in Section 3.7.

Estimating program impacts on refugee well-being

The primary approach to estimate program treatment effects on the treated (TOT) is instrumental variables (IV) estimation, as represented by the following two equations:

$$T_{ic} = \alpha_0 + \alpha_1 Z_c + X'_c \Lambda_1 + W'_{ic} \Gamma_1 + \mu_{1,ict} + \eta_{ic} \quad (3.1)$$

$$y_{ict} = \beta_0 + \beta_1 \hat{T}_{ic} + X'_c \Lambda_2 + W'_{ic} \Gamma_2 + \mu_{2,ict} + \epsilon_{ict} \quad (3.2)$$

Equation (3.1) estimates the first-stage effect of assignment to treatment on the household's take-up status (with results presented above), and equation (3.2) estimates the TOT effects. To enhance statistical power, data from the midline, endline, and follow-up surveys are combined in

the main analysis when possible, as pre-specified. The data includes multiple rounds t of survey data for each household, and the data are stacked. T_{ic} denotes the treatment take-up for household i in community c . Z_c is an indicator variable signifying whether community c was randomly assigned to HSP treatment, where all eligible individuals were then assigned to the treatment. The outcome variable of interest is y_{ict} , and the predicted treatment take-up, $\hat{T}_{ic}$, is derived from equation (3.1).

The vector X_c encompasses community stratification variables used in the randomization, including indicators for whether a community is in Mafrq or Irbid and the community's district's population quartile¹⁸ The vector W_{ic} includes individual baseline covariates, accounting for household-level demographic variables from the integrated assessment; in particular the vulnerability quartile, month of assessment, household size, number of children, and respondent gender and age. $\mu_{1,ict}$ and $\mu_{2,ict}$ are vectors of enumerator-round fixed effects. η_{ic} and ϵ_{ict} are error terms and are clustered at the community level, which is the level of randomization.

To assess dynamics, the analysis also estimates TOT effects separately with the midline, end-line, and follow-up data, while the regression specification remains effectively the same as previously described, with enumerator fixed effects (rather than enumerator-round fixed effects). The round-by-round results discussed below tend to focus on estimates that survive a multiple testing adjustment within each domain of outcomes, namely the false discovery rate (FDR) adjustment (Anderson 2008), with a q-value of 0.10 or less.¹⁹

Bounding estimated effects in the presence of non-random attrition

Although an overall high tracking rate was achieved, with 86% of the refugee sample surveyed in at least one survey round, there was differential attrition between the treatment and control refugee samples. As detailed in Table C.2, the gaps in attrition are statistically significant, and therefore we estimate bounds following Lee (2009) and report these in the appendix. These bounds estimate the treatment effect under two of the most extreme assumptions of attriters' characteristics – that attriters are either the most affected or least affected participants in the sample. As such, they present some of the most conservative corrections of the treatment estimate. The text highlights results for which both estimated bounds take the same sign; when the bounds are on opposite sides of zero, this is noted in the text and the corresponding results are discussed as suggestive.

¹⁸As noted before, communities are defined by the Jordanian administrative unit “locality”. “District” is a more aggregated, administrative geographic unit used in Jordan.

¹⁹In a few instances, a result with a q-value greater than 0.10 is discussed to contextualize a related result and these cases are noted in the text. Throughout the paper, when two similar outcomes have significant q-values, we tend to report only one of them in the main paper to streamline exposition; refer to the online appendix for estimated effects for all pre-specified outcomes.

Estimating impacts of the program on neighbors' attitudes

The estimation of HSP effects on neighbors is similar to the IV approach carried out in the refugee sample, where the treatment status of the refugee household is applied to their surveyed neighbor:²⁰

$$T_{ic} = \delta_0 + \delta_1 Z_c + X'_c \Phi_1 + W'_{ic} \Psi_1 + \mu_{1,ic} + \rho_{ic} \quad (3.3)$$

$$y_{ic} = \theta_0 + \theta_1 \hat{T}_{ic} + X'_c \Phi_2 + W'_{ic} \Psi_2 + \mu_{2,ic} + \epsilon_{ic} \quad (3.4)$$

T_{ic} equals one if the neighbor's associated refugee household received treatment and zero otherwise, and Z_c equals one if the community was randomized to treatment. The X_c and W_{ic} terms are as above, $\mu_{1,ic}$ and $\mu_{2,ic}$ are enumerator fixed effects, and standard errors are again clustered at the community level. While we focus on this pre-specified analysis of neighbor impacts, the PAP contained several additional secondary or more exploratory analyses, including the estimation of ITT effects, the evaluation of the effects of the program using a continuous treatment variable (measured in either months of treatment and the total cash value of the transfer²¹), the assessment of treatment effects on refugees' economic convergence with their neighbors, and estimation of heterogeneous effects (based on various demographics); the full set of pre-specified results is in the Supplemental Appendix.

Estimating the accuracy of forecasts

To determine whether mean forecasts differed significantly from the estimated program impact, the analysis takes into account both the estimation error in the estimated program impact ($\hat{\beta}_1$ and $\hat{\theta}_1$ in the second-stage estimation equations above) as well as the observed range of forecasts, using 1000 simulated draws of each. For each draw, the difference between the two values is taken, yielding the distribution of the differences. Statistical significance is determined by inspecting whether zero lies within the 95% confidence interval of the distribution of this difference; if zero lies outside the interval, we reject the hypothesis that the mean forecast and estimated treatment effect are equal.

3.5 Housing Subsidy Program Impacts on Refugee Life Outcomes

We first estimate program impacts on refugees pooling survey rounds, and then explore dynamic effects by rounds. As noted, we generally focus on results that both survive a multiple hypothesis testing adjustment (i.e., FDR q-values < 0.10) and do not change sign when bounded to account for differential attrition.

²⁰While it is conceptually possible (as outlined in the pre-analysis plan) that the program might generate economic and labor market spillovers to untreated refugee households and neighbors, the relatively low saturation level of the experiment (at a share of around 0.5% of all refugees treated) makes this unlikely.

²¹Since the implementing organization negotiated the length of the contract and the amount of reduced rent with each landlord, there is some non-random variation in both across treatment households.

Average program impacts on refugees pooling survey rounds

Table 3.4 presents the estimated treatment effects on refugees pooled over as many rounds as are available for the pre-specified primary outcome measures, as well as statistically significant effects on other outcomes. By pooling, these results can be interpreted as the total average program effect.

In terms of primary outcomes, the estimated treatment effect on the housing quality index across all rounds is positive and moderate in size, at 0.31 standard deviation units, though not statistically significant at traditional confidence levels. (Note that we do see some positive treatment effects on housing quality outcomes at midline, described below.) Regarding rental payments, there is a significant reduction in households' out-of-pocket housing expenditures of 84.71 USD (SE 33.10) and the estimate survives the False Discovery Rate (FDR) multiple testing adjustment (q -value=0.07). This measure was reliably collected at the endline survey when the subsidy itself was ending. The reduction observed in housing expenditure is likely the combined result of the subsidy and the contractually agreed stable rent during and after the treatment period. Taken together, the two findings indicate that the program at least partly achieved some of its main proximate objectives of improving housing quality and reducing the rental burden for refugee households, although the housing quality estimate is suggestive. As we are unable to directly measure the reduction in out-of-pocket housing expenditures during the midline survey (as noted above), the endline estimate likely understates the true program benefit to treatment households in terms of reduced rental payments.

In a central finding of the study, we next show that the average estimated program effect on consumption across all survey rounds is small and not statistically significant (Table 3.4). Specifically, across all rounds for which data are available (which varies across the measures, see notes in Table 3.4), there are no significant impacts on household food consumption nor log total consumption expenditures. Likewise there is no significant effect on respondent depression as assessed using the widely used CES-D scale. It appears that a substantial year-long housing subsidy plus shelter upgrade was not sufficient to meaningfully improve these key living standards measures, although it is worth noting that, as with the housing quality effects, the relatively weak first stage increases standard errors on these IV estimates. Despite this, these estimated null effects are still fairly precise: the associated standard errors suggest that an effect of 0.16 log points on consumption expenditures would be statistically significant at 95% confidence, and analogously a 0.26 standard deviation effect on the CES-D depression scale would be statistically significant.

Unexpectedly, the one large and statistically significant result suggests that the program led to worse outcomes: the housing subsidy program leads to a decrease in the socio-emotional well-being of children (measured using the SDQ scale) of 0.34 standard deviation units, with a multiple testing adjusted q -value of 0.07.²² A 0.34 standard deviation on the SDQ scale constitutes a moderate effect and is comparable in absolute size to impacts observed in cash transfer studies.²³

²²The components with the largest negative treatment effects are child conduct, emotional, and hyperactivity. There was no significantly negative impact on the peer subcomponent.

²³For example, Baird et al. (2013) report psychological well-being gains of 0.25–0.3 SD from cash transfers to adolescent girls in Malawi, while Haushofer and Shapiro (2016) find mental health improvements of 0.3–0.4 SD from unconditional cash transfers in Kenya. Such effects are particularly meaningful among vulnerable groups, where even

Beyond the primary pre-specified outcomes, all other outcomes with statistically significant estimated program effects pooled across rounds that survive a multiple hypothesis testing adjustment are reported in Table 3.4, Panel B. One striking finding is the large negative impact on treatment households' food security. In particular, adult respondents in the treatment group were 19 percentage points more likely to go to sleep hungry in the past week (50% more likely) than those in the control group (FDR q -value < 0.01). Children and other adults in treatment households were also 16 and 12 percentage points more likely to go hungry in the past week (q -values 0.01 and 0.04), respectively, though these estimates are less robust to using Lee bounds to account for non-random attrition (Panel B, Table B.1). Although the midline did not document impacts on household food expenditures because these variable were not measured in that round, the midline results discussed below highlight important reductions in food aid received. Household structure, and COVID-19 incidence also appear to have contributed to this finding.

Yet at the same time – and perhaps due to the fact that the program reduced housing expenditures – the program had positive impacts on household savings: pooled over the endline and follow-up survey rounds, treatment households were 8 percentage points more likely to have 30 JDs (roughly 95 USD PPP) in savings, on a base of just 10 percent in the control group (and the q -value of the difference is 0.05, although the finding is less robust to Lee bounds). Though the magnitude of savings is not large, the evidence that households both increased savings and reported more hunger suggests that households may have valued future savings above and beyond current consumption, perhaps as a precaution against anticipated future negative shocks. There is also an insignificant but suggestive reduction in rental debt in the endline, equivalent to approximately a quarter of average rental debt in the control group (where the average household owes the equivalent of four months of rent). There is also a smaller but marginally significant reduction in rental debt in the follow-up 1.5 years after implementation. Additionally, we find some indication that monthly rent paid to landlords (including from the implementing organization) increases in the treatment group, consistent with more reliable rental payments to the landlord through the program (see Supplemental Appendix).

There were no statistically significant impacts on other outcome measures that were collected across multiple rounds, including for household composition, migration, and employment. However, there are some meaningful effects in particular survey rounds and we turn to these next.

Impacts in the midline round

Recall that the midline phone survey was conducted while the HSP was ongoing and many households were still receiving assistance. The midline survey data indicates that HSP improved several dimensions of housing quality in the short run (Panel A, Table 3.5): households experienced a 21 percentage point increase in access to clean water (SE 0.06, control mean 17%) and a 0.51 room increase in dwelling size (SE 0.17), likely the result of door installation and other construction. Yet, as in the pooled results, there is not a statistically significant change in the overall housing quality index.

moderate socio-emotional improvements can have meaningful implications for development and resilience.

Food security outcomes at midline echo those in the pooled sample and also shed light on the possible mechanisms (Panel B, Table 3.5). As in the pooled results, respondents were more likely to report that they and others went to sleep hungry in the past 7 days, with a robust 23 percentage point (q -value = 0.01) increase in respondents' hunger, and other adults and children in the household being 17 and 24 percentage points more likely, respectively, to go to sleep hungry (SE 0.06 and SE 0.07). Most importantly for understanding why, treatment households report substantial decreases in food aid received, consuming 53 USD PPP less in food from assistance (q -value = 0.02), which represents a 24% decrease in food assistance relative to control. This is a nontrivial proportion of households' budgets: it is equivalent to over 10% of households' total monthly labor income pre-pandemic (as control households reported 218 USD PPP in average monthly food assistance and pre-pandemic monthly labor income of 503 USD PPP). It is worth noting that this measure of assistance includes both formal assistance, such as from the UN High Commissioner for Refugees and the World Food Program, as well as informal assistance from local organizations and community members. Regardless of the channel, this substantial decrease in the household's food budget seems likely to have contributed to worse household food security. This is in line with conversations that the research team had with several humanitarian aid organizations during the pandemic, in which some implementers sought to allocate emergency assistance to households that were not already benefiting from large forms of assistance, such as the HSP. In humanitarian settings, it is common for multiple actors to provide various forms of aid, each with their own targeting criterion, which highlights the potential relevance of this result to other contexts with forced migrants.

The midline survey data also indicate that there were several other unexpected short-run changes in household outcomes. For instance, it appears that changes in household composition may have further contributed to household hunger (Panel C, Table 3.5). Treated households at midline report significantly more children than control households (point estimate 0.3 children, q -value=0.03), and this effect is driven by an increase of 0.16 more adolescent boys (q -value=0.33). This movement between households can be understood as a form of redistribution within the Syrian refugee community, by which treated households shared some of the program benefits of the HSP but at the potential cost of food security to the full household. Despite the possible negative effect on household food security, the study cannot rule out that recomposition might have improved the well-being of the new household members who moved in (and the households they came from), if those adolescents came from relatively worse-off households. (Similarly, the targeting of food assistance away from HSP treatment households noted above could have benefited other refugee households that received additional assistance.) Nonetheless, these individuals' joining the treated households appears to have likely come at a cost for other treatment household members. The effects of the program on household recomposition, however, are not fully robust when correcting for non-random attrition (Panel C, Table B.2), and as such should be viewed as suggestive.

Treated households also report experiencing increased COVID-19 exposure during the mid-2020 period when the survey was collected, which coincided with a peak period of the pandemic (Panel D, Table 3.5). Treated participants displayed an average of 0.84 additional COVID-19 symptoms compared to a control group mean of 0.47 (q -value <0.01), and treated households had 0.46 more symptomatic adults and 0.17 more symptomatic children (q -value<0.01 and q -

value=0.01, respectively). These households were also 16 percentage points more likely to report non-adherence to social distancing guidelines (q -value = 0.08). Increased household size, as noted above, may have contributed to increased COVID-19 exposure, especially if individuals from other households were moving in and out of the home. Another more speculative explanation is that the HSP may have allowed household members to travel outside the home more often, for instance, by improving the security of the home through better window, door and lock installations. As noted above, analysis also shows an increase in women's physical mobility (Table B.5). Perhaps surprisingly, treated households were significantly less likely to have visited healthcare institutions at midline, with 0.56 fewer visits on average compared to 1.14 in control (q -values = 0.04). A definitive explanation for the overall pattern of health and COVID-19 results remains elusive.

Despite these negative effects, there were also several indications of improved household outcomes due to the HSP. Consistent with the subsidy program relaxing the household's budget constraint, midline survey data indicate that there were reductions in credit usage among treated participants (Panel E, Table 3.5): treated participants reported taking fewer loans before and during Jordan's COVID-19 lockdown by 20 and 34 percentage points, respectively. We largely view this outcome in a positive light considering the high debt burden in the sample: control households held average total rental debt of 806 USD PPP at endline. Additionally, weekly adult income decreased by 20 USD PPP (SE 7.19, q -value = 0.02) during Jordan's stringent first COVID-19 lockdown when many businesses closed, suggesting individuals could afford to remain out of work during the lockdown. This reduction in income, however, did not persist after lockdowns were lifted (see Supplemental Appendix).

Finally, there are mixed midline estimates on child outcomes. Children in treated households were reported to engage in 0.19 more learning activities (q -value = 0.08) but attend 0.43 fewer days of school on average, at a time when most in-person schooling was suspended due to the COVID-19 pandemic. Children in treated households also experienced a decrease in reported alertness of 0.73 standard deviations (q -value < 0.01). Child SDQ was not measured at midline due to time constraints. Beyond child outcomes, other significant effects at midline include the probability that the respondent is living with a disability that complicates self-care (0.38 SD higher difficulties with self-care, q -value < 0.08). Yet, only the child alertness reductions and days of school attendance are robust to accounting for non-random attrition (Table B.2).

Impacts in the endline survey round

The endline survey was carried out in-person in 2021 as pandemic conditions eased, shortly after program assistance had ended for most households. As noted above, we find that the program significantly reduced housing expenditures by 84 USD PPP at endline relative to the control group (q -value=0.07). Effects on measured housing quality, on the other hand, largely dissipate by this point, with no significant differences between the treatment and control groups in any dimension of the housing quality index. Some of this could have been due to catch-up investments in the control group, or the possibility that housing improvements like window repairs and mold removal depreciate relatively quickly. Likewise the endline survey data reveals no statistically significant effects on any of the other pre-specified primary outcomes: program impacts on total household

consumption, respondent depression levels, and child well-being are all negative, in fact, although relatively small in magnitude and none are significant at traditional confidence levels.

By the endline survey, most impacts observed at midline – including on food security, household composition, child learning outcomes, and COVID-19 – were no longer statistically significant. Four significant effects survive multiple hypothesis testing adjustments (Table 3.6). First, there was a notable reduction in the amount of time respondents devoted to chores and childcare, with a decrease of 6.91 hours in these activities over the past week and corresponding 10.62-hour decrease in total time spend on labor and chores ($q\text{-value} = 0.01$ and $q\text{-value} = 0.01$, respectively), which are robust to accounting for non-random attrition. These are somewhat puzzling given the lack of meaningful changes on labor supply outside the home (which if anything appear to decline). One explanation for the change in childcare hours is that the HSP increased women’s mobility (see Table B.5), perhaps by allowing them to leave children at home due to improved home physical security (as windows and doors were repaired, for instance). Motivated by this hypothesis, the study tested for impacts on women’s mobility in the follow-up round, which shows, as discussed below, that women in treated households in fact did experience increased mobility (Table B.5).

Second, there was a decrease in self-reported respondent happiness, which fell by 0.22 points on a scale of 1-3, from an average of 1.76 ($q\text{-value} = 0.02$). Third, and on a more positive note, there was a significant increase in the proportion of households that reported savings of 30 Jordanian Dinars, likely due to the financial benefits of the HSP in reducing household rent payments. This appears related to the midline finding that treatment households had a lower debt burden and did not take out as many loans. Fourth, program recipients received 128 USD PPP more in assistance on average from the implementing partner in the past 12 months ($q\text{-value} = 0.10$), which likely reflects the small proportion of recipients still some receiving HSP assistance.²⁴ Other results suggest that the program may have had negative impacts on labor supply and earnings but the estimated effects are less precise and not as robust to changes in measurement or regression specifications (see Table B.7).

Impacts in the follow-up survey round

The follow-up survey round was collected in 2022, roughly 1.5 years after all HSP direct rental assistance had ended. Given the length of the rolling implementation window, the follow-up captures outcomes as much as 2.5 years after some households received assistance. At that point the only statistically significant and robust estimated effect of the program is related to child socio-emotional well-being: the standardized Strengths and Difficulties Questionnaire (SDQ) scale decreased by 0.56 standard deviation units in treated households ($q\text{-value}=0.01$). This is a large magnitude and indicates that there was a meaningful decline in child socio-emotional well-being (as reported by parents), an unintended adverse consequence. There are multiple potential explanations for the worsening of child socio-emotional well-being, including short-run (midline) food insecurity, greater COVID-19 exposure, household composition impacts, and changes to respon-

²⁴While the estimated reductions in time spent on chores and in happiness are robust to accounting for non-random attrition, the other two effects noted in this paragraph are not, and are therefore more suggestive.

dent time use, though the persistence of the negative effect on child well-being after those effects have dissipated is perhaps surprising. A further channel relates to the nature of social interactions with neighbors, discussed in the next section. Besides that effect, across the wide range of household outcome measures gathered in the follow-up survey round (including all the pre-specified primary outcomes) there were no other significant results that survived a multiple hypothesis testing adjustment.

3.6 Impacts on Neighbor Attitudes

The 2022 survey of a representative sample of Jordanian neighbors was designed to capture the spillover impacts of the refugee-targeted assistance program on Jordanians' attitudes towards refugees at least one year after program implementation ended (detailed in Figure 4.1). Social cohesion results were also measured for refugees, but were not significant.

We focus first on the pre-specified primary outcomes (Table 3.7, Panel A). Perhaps unexpectedly, neighbors who live near treatment households experience a significant decline in the social attitude and perceptions index (effect -0.33 standard deviation units, $q\text{-value} < 0.08$). A closer examination reveals that this negative effect is primarily due to diminished social ties between Syrian refugees and their Jordanian neighbors (Table 3.7, Panel B): there is a notable decrease in the number of Syrian refugees from whom Jordanian adults seek advice (out of their three closest contacts), and an overall reduction in whether Jordanian children have Syrian friends.²⁵ Additionally, the point estimate is negative for 6 of 8 measures of Jordanians' social attitudes towards Syrians, such as willingness to have a Syrian marry into the family. This pattern suggests that the observed effects are primarily driven by a deterioration in Jordanian attitudes, rather than by voluntary social distancing on the part of Syrians.

Moreover, a heterogeneity analysis using the Marlowe-Crowne Social Desirability Scale (Crowne and Marlowe, 1960) suggests that effects are even stronger among respondents who exhibit less social desirability bias, suggesting that truly honest responses would generate even larger overall treatment effects.²⁶ This latter result could be a partial explanation for the reduced well-being measured among treated Syrian children. A leading explanation for this pattern is that there was a backlash effect among Jordanians living in the working-class neighborhoods that Syrian refugees tend to inhabit, driven perhaps by resentment at the assistance received through HSP.

In line with these results, we find suggestive evidence that refugees also report reduced interactions between Jordanian and refugee children, that they have fewer close Jordanian friends, and

²⁵Given the way that the questions were structured, note that changes in the overall size of a social network alone could not generate these effects.

²⁶Additional analysis that regresses each component of the social attitudes index on treatment interacted with the Marlowe-Crowne Social Desirability Scale reveals two other patterns. First, respondents who score below median on the Marlowe-Crowne Scale (i.e., show less tendency to give socially desirable answers) have more negative treatment effects than the above-median respondents. Second, the coefficient on this term (treatment interacted with an indicator for those with below median scores) is negative for 7 out of 8 index components, including attitude measures (Supplemental Appendix, Table G.20). This lends support to the view that the result reflects negative social attitudes and not only reduced contact.

are less likely to become friends with their Jordanian neighbors. However, these estimates are not statistically significant (Supplemental Appendix, Table F.10).

Regarding the other pre-specified primary outcomes, there are positive but small estimated program impacts on neighbors' policy preferences and altruism towards Syrian refugees but these are not statistically significant. We also find no evidence that the program increased Jordanians' housing costs or had other adverse economic effects on neighbors. There are no significant impacts on Jordanian neighbors' total consumption or housing expenditures (see Supplemental Appendix).

There are several other notable patterns among neighbors (Table 3.7, Panel C). First, neighbors of treatment-assigned refugee households report significantly worse subjective assessments of their own well-being, specifically, a 0.30 standard deviation unit decrease in their self-assessed health (q -value=0.06), and a 0.32 standard deviation unit decrease in life satisfaction (q -value=0.06). There are several potential drivers of these impacts, including the possibility that observing refugee neighbors receive generous assistance led them to feel worse about their own living situation and quality of life, which is perhaps related to the resentment and backlash effects hypothesized above.

Second, neighbors of treated households state that they believe refugees typically receive significantly *less* aid than that stated by control group neighbors (effect -347 USD PPP, q -value=0.04). Interestingly, the neighbors of treated households have more accurate beliefs about aid levels, with fewer of them believing that refugees receive extremely high levels of aid that are rarely observed. There thus appears to have been some learning of objectively true information regarding the reality of refugees' assistance levels among their Jordanian neighbors. This is consistent with greater information exchange and learning as a result of the highly observable HSP.

Third, neighbors of treated refugee households claim to have -1.08 fewer days of media consumption per week (q -value<0.08), relative to a control mean of (4.78). It is difficult to explain this effect on media consumption but one hypothesis is it reflects a decrease in neighborly socializing, which sometimes resolves around watching or listening to the news on the television or radio; but we are unable to offer more decisive survey-based evidence on any underlying mechanisms.

There was no differential attrition when surveying neighbors of treatment and control communities, and so most of the significant results discussed are robust to bounding for non-random attrition; the only exceptions are the effects on life satisfaction and subjective health, which should therefore be viewed as more suggestive.

Heterogeneity analysis lends insight into potential mechanisms underlying these effects (Table B.6). First, variation in physical distance to the treated refugee household allows us to bolster the premise that the worsening in social attitudes is related to direct exposure to HSP and its recipients: the decrease in the social attitudes index is concentrated among those neighbors who live physically closer to the refugee households (based on GPS-measured distance). This suggests that more direct and frequent observation of treated refugees' circumstances led to more negative effects.

A second informative dimension of heterogeneity is whether the respondent has non-Jordanian grandparents. In this setting, 17 percent of the neighbor sample has a grandparent (or spouse's grandparent) born outside of Jordan, with 72 percent of those born in Palestine. Notably, the large negative program impact on the neighbor social attitudes and perceptions index is driven by those with only Jordanian-born grandparents, while those with a non-Jordanian born grandparent have an effect close to zero (Table B.6). Similar patterns exist for the life satisfaction results:

neighbors with exclusively Jordanian grandparents have the largest negative treatment effects on life satisfaction (Supplemental Appendix, Table G.21). While the results are noisy due to the relatively small share of neighbors with non-Jordanian born grandparents, this finding suggests that a family history of immigration or displacement — which is common among those of Palestinian descent, for instance — may enhance openness to Syrian refugees.²⁷

The pre-analysis plan additionally specified heterogeneity analysis by neighbor gender, age, education and socioeconomic status but there are no meaningful differences along these dimensions (see Supplemental Appendix).

3.7 Addressing Low Take-up Rates

Additional analysis was conducted to infer whether the relatively low take-up rate of the housing subsidy program influences the interpretation of treatment effects. As noted above, qualitative feedback from the implementing partner identified several reasons for low participation, including titling issues, housing units deemed either too high-quality or too structurally precarious, and landlords' reluctance to enter into formal rental agreements with refugee tenants. These barriers give rise to two relevant points. First, understanding selection into treatment – driven by differences between compliers and non-compliers – will clarify to whom the estimated treatment effects generalize. Second, the offer of the program itself may have influenced outcomes, raising the possibility of a violation of the instrumental variable exclusion restriction.

Regarding the first point, it is well-known that selection into treatment does not threaten the internal validity of the causal estimates but rather defines the relevant population for the IV Local Average Treatment Effect (LATE, Angrist et al., 1996). This means that the analysis recovers unbiased treatment effects for the subset of individuals whose take-up status was influenced by randomized assignment. Among the treatment group, we examined baseline differences in covariates across program compliers versus non-compliers and found evidence that take-up appeared to be influenced by both housing instability and rental debt owed to the landlord (Appendix D, Table D1). This suggests that the estimated effects are most likely to generalize households experiencing residential instability, as they are most likely to take-up the program.

Regarding the second point, the exclusion restriction may be violated if treatment assignment *per se* influences outcomes through channels other than actual program take-up. To assess the plausibility of this assumption, we conduct analysis that compares individuals in treatment and control groups who have similar predicted probabilities of taking-up the program, based on their observed covariates. Specifically, if the exclusion restriction holds (and effects are working through program receipt), we would expect no meaningful differences in outcomes among household with very low predicted program take-up rates, as few would have complied with treatment even if assigned. Conversely, any estimated treatment effects should be concentrated among those household with high predicted take-up probabilities.

²⁷This finding is in contrast to the correlational evidence in Ghosn et al. (2019), who show that a history of displacement did not improve openness towards refugees.

Calculating predicted take-up rates. Using the treatment group, we first train LASSO to predict take-up based on a wide range baseline household covariates. Then, using that model, we predict what take-up would have been (if assigned to treatment) for the full sample, including the control group. We then divide the full sample into terciles based on the predicted take-up propensity. Finally, we estimate the effect of assignment to treatment for each tercile as follows:

$$y_{ict} = \theta_0 + \theta_1 Z_c + \sum_j \theta_{j2} \text{Tercile}_{jic} + \sum_j \theta_{j3} (Z_c \times \text{Tercile}_{jic}) + X'_c \Phi_1 + W_{ic} \Psi_1 + \mu_{2,ict} + \epsilon_{ict}$$

where all variables are defined as before and Tercile_{jic} is an indicator for the household's baseline take-up propensity being in tercile $j \in \{2, 3\}$.

The results indicate that LASSO performs moderately well in predicting take-up (Appendix D, Figure B.1a): in the tercile with the lowest predicted take-up, only 5% of the treatment group actually took up, while in the third tercile take-up was 63%.

Results for refugee outcomes. The estimated coefficients for the equation above indicate that for the primary outcomes for refugees, only one of the six outcomes exhibits a statistically significant treatment effect among individuals in the lowest tercile of predicted take-up, i.e., the coefficient estimate on θ_1 (Appendix D, Table B.11). This provides suggestive evidence that the exclusion restriction generally holds, as assignment to treatment alone appears to have no effect for households unlikely to be program participants.²⁸

Results for neighbor outcomes. In contrast, the same analysis using neighbor outcomes points to several significant effects of assignment to treatment across predicted take-up terciles, including for the lowest tercile (Appendix D, Table B.12). This suggests that some impacts on social cohesion occurred regardless of whether the assigned household ultimately received the housing subsidy. In other words, even among neighbors of non-compliers, their simply having been offered the program appeared to trigger responses, perhaps because treated households may have inadvertently revealed their program status through implementation visits, landlord negotiations, or informal conversations. Indeed, the data from the follow-up round phone survey reveals that treated refugee households reported that the majority of their neighbors were aware that they had been offered the program. The fact that the program was implemented with a cluster randomized design also means that Jordanian neighbors may have been aware that other refugees nearby had also been offered the program, possibly contributing to effects even among neighbors of non-compliers. Given the possible violation of the IV exclusion restriction, the ITT estimates are arguably appropriate; see Tables B.16 for these results, which show similar patterns.

3.8 Comparing Forecasts to Estimated Program Impacts

Comparing experts' forecasts to estimated the program effects allows us to highlight areas where the study results advanced learning relative to prior expectations. The forecasts focus on the mid-line and endline measures of the primary outcomes, namely: i) the well-being measures among refugees, and ii) the attitudes and perceptions of Jordanian neighbors towards refugees.

²⁸The Intention to treat (ITT) reduced form effect are presented in Table B.13.

In terms of impacts among refugees, the findings reveal that the actual impacts of the intervention on refugee well-being were generally smaller than the modest positive impacts that experts had anticipated, with the exception of housing quality (Figure 3.3, Panel A). Notably, statistically significant discrepancies (at over 95% confidence) between the estimated and predicted effects were observed in two areas: housing expenditures and child socio-emotional health. Housing expenditures in the endline survey were notably lower for treatment households compared to control, while forecasts anticipated no such effect. One limitation of the predictions we collected, however, was the fact that we failed to inform the forecasters that the program had mandated that landlords had to maintain stable rent levels for one year following the end of subsidy distribution, and this appears to be a likely driver of the non-effect on forecasted housing spending. Thus, some caution is needed in interpreting these effects.

Experts also forecasted a marginally significant improvement in child socio-emotional well-being at endline, yet if anything there are reductions in this outcome at endline though they are not significant; as noted above, there are significant reductions in the child well-being measure at the 1.5-year follow-up but forecasts were not collected for that round. There is thus a meaningful and significant (at 95% confidence) difference between the predicted and actual effects on the child outcome, indicating that the study generated new insights that diverge from the experts' priors.

There is also a meaningful and significant gap between experts' forecasts and the observed effects of the program on neighbors' social attitudes (Figure 3.3, Panel B). Experts had anticipated an average null effect of the program on Jordanian neighbors across all three primary pre-specified outcomes (namely, policy support, economic perceptions, and social attitudes), and for the first two the predictions and actual estimates are all close to zero. The program's negative impact on neighbors' social attitudes is significantly different from the predicted null effect at 95% confidence, once again a meaningful update relative to priors.

3.9 Conclusion

This paper investigates the short- and medium-term effects of a substantial housing subsidy program on Syrian refugee households and their Jordanian neighbors. The program provided one year of free rent on average, along with improvements in shelter quality and continued rent stability even after the subsidies ended. The study examines the impact on the living standards and well-being of Syrian refugees, as well as the social attitudes and interactions of their Jordanian neighbors. The study's novel design, which restricts the use of subsidies to existing rental relationships, largely prevents the migration responses typically associated with rental subsidies.

The program ultimately was executed largely as planned despite the COVID-19 pandemic, albeit with relatively low take-up, but had unintended consequences on the ground. The main analysis documents largely null effects on the primary pre-specified living standards measures, including household consumption and respondent well-being. This is perhaps surprising given that both the humanitarian implementing organization and expert forecasters including the research team, believed that the program would generate moderate positive impacts.

Two striking and unexpected results are the significant deterioration in measured child socio-emotional well-being following the program and the increased strain in relations between neighbors and refugees. In terms of the child outcomes, there are multiple potential channels that could explain the results, including the finding that food security worsened among treated households (as they appear to have been partially cut off from formal or informal food aid), that treated households experienced higher incidence of COVID-19, and treated households' changes in living arrangements (at least in the short run, such as the increase in adolescents).

The negative child well-being and social cohesion outcomes may in fact be interrelated: there are decreased interactions between treated refugee children and Jordanian children. Importantly, we identify these negative social and attitudinal effects toward treatment recipients in Jordan, a country that has a long history of accepting refugees from various conflicts, and close cultural ties to Syrian refugees, which might signal potentially even worse outcomes in places where refugees might not be as welcome. This deterioration in refugee-host community relations may have also jeopardized the informal support in terms of food or other forms of assistance that some refugees received from their Jordanian neighbors. We also show that the negative responses among neighbors are not driven by increases in their housing costs in treated areas, ruling out a potential alternative explanation.

Taken together, the results indicate that a meaningful housing subsidy did not lead to transformative positive changes for recipient households. Moreover they also induced a temporary re-composition of treated households, something observed also in other humanitarian program evaluations (Özler et al., 2021). Even if these movers benefited, their stay in the treated households was only temporary, while the negative child development and social cohesion effects on treated households persisted to the post-program follow-up survey round. The findings thus offer a word of caution when designing assistance programs in settings with strong social ties and the potential for both formal and informal redistribution of assistance. More broadly, the lack of sustained impacts for treatment households may reflect the numerous other constraints that refugees face in gaining access to livelihood opportunities, credit, and quality housing. The deterioration in social cohesion offers novel evidence that assistance targeted exclusively to refugees can prompt host community backlash.

This study discusses the possible alternative explanations for adverse program effects. First, an analysis of take-up suggests that the significant negative effects on children are not driven by a violation of the exclusion restriction. Yet there is evidence that just being offered the program may have contributed to the negative effects on social attitudes observed among Jordanian neighbors. Second, program implementation was rolled out largely according to plan, as indicated by the temporary improvements in housing and rental expenditures and qualitative reports from the implementer. Third, although the COVID-19 lockdown occurred during the period of implementation, it only covered roughly six weeks of the program's total duration (amounting to approximately 10% of the period) and we actually observe that the housing subsidy was able to help households temporarily improve their income resilience in response to the shock. Fourth, it seems unlikely that measurement problems and reporting bias explain the results given that the enumerators never explicitly referenced the housing program and did not know households' treatment status. Fifth, given that the majority of effects are null, it seems unlikely *prima facie* that respondents manipu-

lated their reports to receive more aid. Additionally, we can largely rule out that the reductions in household debt enabled by the program resulted in large future benefits by allowing households to better smooth consumption, since we find predominantly null effects on consumption levels across three years of data, and no significant effects on the variance of consumption.²⁹

We view these as unexpectedly discouraging results of the housing subsidy program, and they are less optimistic than predicted by experts (or by the research team). Although different results might hold in populations other than Syrian refugees in Jordan, on some level this setting would seem to be almost a global “best-case” scenario for the potential success of such a program, given the high degree of linguistic, cultural and religious similarity between the Syrian refugees and their Jordanian hosts, and the fact that groups with closer cultural ties have been shown to exhibit better social cohesion (Alrababa’h et al., 2021; Barron et al., 2023). Furthermore, the fact that the program was launched roughly a decade after displacement had begun from the Syrian Civil War allowed ample time for construction of additional housing units and improved social integration.

The results also speak to the active ongoing policy debate on how best to support refugees and host states at the same time (Ash and Huang 2018; Baseler et al. 2023). When study participants were asked directly about what form of assistance they preferred (during the 1.5 year follow-up survey), respondents overwhelmingly stated that they preferred direct cash transfers to landlord subsidies (at 70%). These results might reflect the unintended negative effects of the program. Cash transfers to refugees in other contexts have been shown to provide only short-run benefits rather sustained gains (Hidrobo et al. 2014; Özler et al. 2021; Quattrochi et al. 2022; Moussa et al. 2022; Aygün et al. 2024), but the findings of this study suggest that delivering benefits more discreetly (such as via cash or mobile money) may at a minimum reduce the risk of host community backlash. An alternative approach, supported by Baseler et al. (2023), would explicitly pair refugee assistance with enhanced host community assistance, which in this case might have led to some housing investments among host community neighbors as well. There is promising evidence from Colombia that labor market reforms may be an effective way to improve refugee living standards without leading to crowd-out of host community jobs (Bahar et al. 2021; Rozo et al. 2023; Ibáñez et al. 2024). These remain critical areas for future research and policy innovation, and present important opportunities to improve program design and, ultimately, refugees’ lives.

²⁹For instance, the std. dev. of log consumption is 0.64 and 0.55 in the treatment and control groups, respectively.

3.10 Figures and Tables

Figure (3.1) Geographic Location of Treated and Control Communities

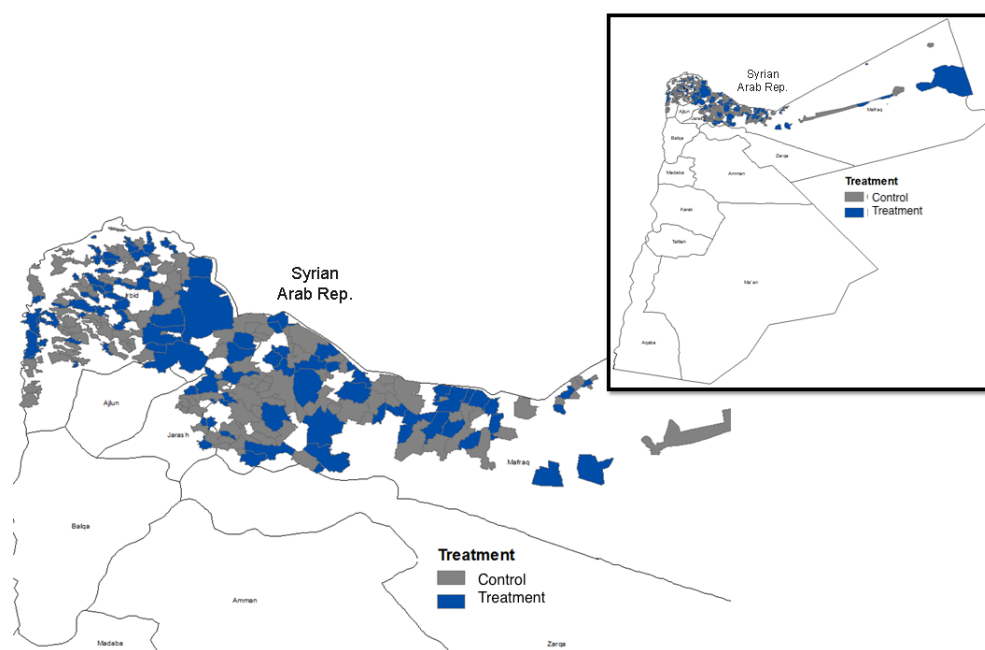


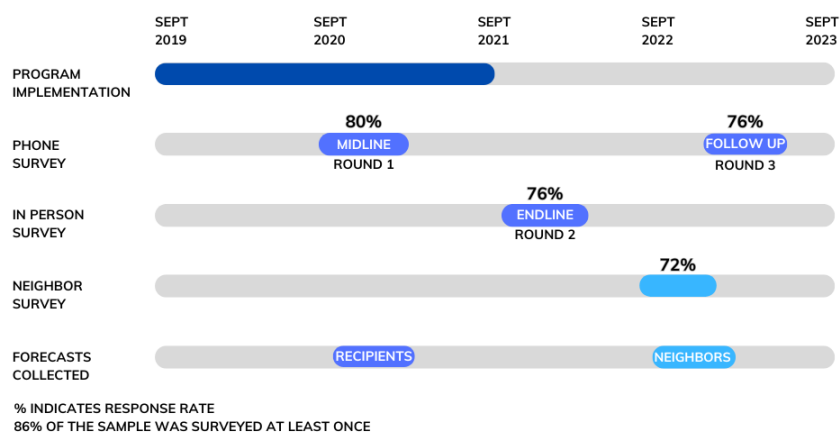
Figure (3.2) RCT, Data Collection, and Forecasts Timeline

Table (3.1) Average Respondent, Household, and Shelter Characteristics

	Treatment	Control	Difference	(se)	N
Panel A: Respondent characteristics					
(1) Female respondent (=1)	0.46	0.52	-0.05**	(0.03)	1,619
(2) Age (categorical)	33.99	34.19	-0.21	(0.61)	1,619
(3) Married (=1)	0.84	0.84	0.00	(0.02)	1,616
(4) Disabled (Washington Group, =1)	0.26	0.25	0.02	(0.03)	1,616
Panel B: Household characteristics					
(5) Dependency Ratio	1.31	1.27	0.04	(0.07)	1,616
(6) Household size	5.20	5.14	0.06	(0.15)	1,616
(7) Number of families in the same house	1.30	1.30	-0.00	(0.04)	1,616
(8) Number of children	2.96	2.89	0.07	(0.12)	1,619
Panel C: Shelter characteristics					
(9) Access to piped water (=1)	0.66	0.67	-0.01	(0.07)	1,616
(10) Fully constructed roof (=1)	0.12	0.08	0.04**	(0.02)	1,616
(11) Functional windows (=1)	0.22	0.22	-0.01	(0.03)	1,616
(12) Completed floor (=1)	0.44	0.40	0.04	(0.06)	1,616
(13) Toilet (=1)	0.93	0.92	0.01	(0.04)	1,616
(14) Plan to stay in shelter (=1)	0.92	0.92	0.00	(0.02)	1,537
(15) Monthly rent (USD PPP)	328.02	335.28	-7.26	(17.57)	1,431
(16) Lease contract (=1)	0.77	0.78	-0.01	(0.04)	1,431
(17) Number of times moved shelter	0.47	0.44	0.03	(0.08)	1,616
(18) Permanent shelter (=1)	0.91	0.86	0.05	(0.06)	1,616
<i>Joint significance p-val: 0.102</i>					

Notes: This table reports treatment balance results for the 1,619 household interviewed at midline from the 2,017 households sampled for the study (see Table C.2). We have admin data from the implementing partner's integrated assessment but only for the 1,619 households that we were able to reach at midline, due to data management by the implementing partner. The first two columns report averages by group, while the third column reports results of estimating $y_i = \beta_0 + \beta_1 T_i + \epsilon_i$. Robust standard errors in parentheses in the forth column and are clustered at the locality level. The last column reports the total sample size. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. All outcomes are taken from the implementing organization's data, except for gender and age, which are taken from the phone survey. "USD" denotes United States Dollars and "PPP" stands for Purchasing Power Parity, which is used to compare the absolute purchasing power of countries' currencies. Rent is winsorized at the top 1% of values, in order to limit the influence of outliers. Disability is calculated using the Washington Group Short Set on Disability and uses the recommended cut-off (level 3) to classify respondents as being disabled or not based on six domains: seeing, hearing, walking, cognition, self-care, and communication. For household size, we drop one observation in which the number of individuals at baseline is reported to be 300. For times changed shelter, we drop one observation in which the respondent is reported to have changed shelter 48 times.

Table (3.2) Forecasts of Treatment Effects

Predicted Treatment Impact (SD)	Mean	SE of Mean	p10	p90	N
Panel A: Refugee Impacts					
Housing Expenditures (Midline)	-0.28	0.07	-0.50	0.30	61
Housing Expenditures (Endline)	-0.01	0.03	-0.20	0.25	61
Housing Quality (Midline)	0.28	0.02	0.10	0.50	61
Housing Quality (Endline)	0.19	0.02	0.00	0.40	61
Household Consumption (Endline)	0.16	0.02	0.00	0.40	61
Adult Mental Health (Midline)	0.22	0.02	0.00	0.40	61
Adult Mental Health (Endline)	0.14	0.02	0.00	0.30	61
Child Socio-Emotional (Endline)	0.16	0.03	0.00	0.30	61
Panel B: Neighbor Impacts					
Social Attitudes	0.01	0.03	-0.29	0.30	63
Economic Perception	0.01	0.03	-0.25	0.30	63
Policy Support	0.02	0.03	-0.22	0.32	63

Notes: This table reports summary statistics of the distribution of predictions of the experimental treatment effects. Treatment effect predictions were elicited in terms of standard deviations. p10 and p90 refer to the 10th and 90th percentile of the distribution of predictions. We report here only observations who made predictions for all outcomes.

Table (3.3) Panel Retention and Compliance

	Ever found	Midline	Endline	Follow-up
Panel A: Refugee sample retention				
All	86.4%	80.3%	76.1%	75.7%
Treatment	90.5%	83.3%	80.4%	80.1%
Control	82.4%	77.3%	71.9%	71.5%
p-value (T-C)	0.000	0.014	0.000	0.000
N	2,017	2,017	2,017	2,017
Panel B: Treatment take-up among the surveyed				
All	17.8%	16.8%	17.6%	17.9%
Treatment	32.8%	33.0%	33.5%	34.0%
Control	0.0%	0.0%	0.0%	0.0%
p-value (T-C)	0.000	0.000	0.000	0.000
N	1,729	1,619	1,466	1,460
Panel C: Neighbors sample retention				
All	72.1%	—	—	—
Treatment	74.7%	—	—	—
Control	69.5%	—	—	—
p-value (T-C)	0.421	—	—	—
N	2,017	—	—	—

Notes: This table shows panel retention and compliance with treatment assignment. The p-values are obtained by estimating $y_{ic} = \beta_0 + \beta_1 \text{Treatment}_c + \varepsilon_{ic}$. Standard errors are clustered at the locality-level. Retention is defined as survey retention, meaning that the participant was located and surveyed in that round. These numbers only include people who were successfully surveyed. Panel A shows retention of the 2,017 households that were sampled for the study. Panel B shows treatment take-up for the sub-sample of households that were interviewed at midline (1,619), endline (1,466), and follow-up (1,460), and Panel C shows retention for the neighbors of the 2,017 refugees sampled. In this case, retention is an indicator for whether a Jordanian neighbor was located near the pre-randomization home location of the refugee household. The neighbor was selected using a randomized algorithm starting from the original home of the refugee participant, with a median distance between the selected neighbor and the original refugee participant's home of 63 meters. Cases of non-retention could result from an inability to identify the refugees' home location based on midline data, or the inability to locate Jordanian neighbors during field visits to the location.

Table (3.4) Pooled Treatment Effects on Primary Outcomes

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N	Rounds
Panel A: Primary Outcomes							
Overall Housing Quality (Z-Score)	0.31	(0.20)	[0.19]	-0.01	(1.00)	4,218	1, 2, 3
Total Monthly Housing Expenditures (\$PPP)	-84.71***	(33.10)	[0.07]	228.80	(208.80)	1,421	2
Food Consumption (Log \$PPP)	-0.02	(0.07)	[0.63]	4.41	(0.72)	2,753	2, 3
Log Total Consumption (Log \$PPP)	-0.03	(0.08)	[0.63]	8.73	(0.55)	1,422	2
CESD Score (Higher: Less Depression)	-0.10	(0.13)	[0.49]	0.01	(1.00)	4,207	1, 2, 3
SDQ Score (Higher: Better Wellbeing)	-0.34**	(0.15)	[0.07]	-0.01	(1.00)	1,782	2, 3
Panel B: Other Outcomes							
Respondent Hunger Last Week (=1)	0.19***	(0.05)	[0.00]	0.38	(0.49)	4,168	1, 2, 3
Adult Hunger Last Week (=1)	0.12***	(0.05)	[0.04]	0.36	(0.48)	4,099	1, 2, 3
Child Hunger Last Week (=1)	0.16***	(0.05)	[0.01]	0.25	(0.43)	3,741	1, 2, 3
At least 30 JD (95 USD PPP) in savings (=1)	0.08**	(0.04)	[0.04]	0.10	(0.30)	2,770	2, 3

Notes: This table reports 5 primary outcomes plus per capita food consumption (not pre-specified), pooled across as many rounds as were collected for each outcome, for refugees only. The table shows that treatment had no impact on housing quality on average over all 3 rounds, reduced housing expenditures at endline, had no impact on food consumption over endline and followup, had no impact on depression over all 3 rounds, and reduced child socio-emotional wellbeing over rounds 2 and 3. Overall Housing Quality is defined as a normalized housing quality index that includes indicators for quality floors, roofs, and walls, indicators for access to grid electricity and piped water, and the number of people per room. Total monthly housing expenditures is only reported for endline in the pooled table due to variations in measurement of housing expenditures across rounds and due to when treatment ended. Housing expenditure was measured including assistance payments at midline, and excluding assistance payments at endline since most recipients had stopped receiving payments but were still subject to rent freezes at part of the program. Results on housing expenditures in each round are reported in the appendix. Total consumption, including food and non-food expenses, was only measured at endline due to the length of the module. Reported in this table is food consumption, which was measured at endline and followup. Total food consumption is the log of the sum of food consumed in last seven days (Cereals and cereal products, Live animals, meat, and other parts of slaughtered land animals; Fish and other seafood; Milk, other dairy products, and eggs; Oils and fats; Fruits and nuts; Vegetables, tubers, pulses; Sugar and desserts; Ready-made food and other food products (baby food, spices)), plus home-produced foods produced in the last 7 days, plus assistance received in the last 30 days divided by 4.33. CES-D was measured in all rounds with exactly the same questionnaire. It includes 10 questions on depressive symptoms. Individuals who skipped more than 2 questions were marked as missing. The continuous scores were standardized with respect to the control group. The Child Strengths and Difficulties questionnaire (SDQ) was administered regarding a randomly selected child aged 3-8 at endline. At follow-up, the SDQ was administered regarding the same child. If not surveyed previously or if that child is no longer in the household, the survey was administered with respect to a randomly selected child age 4-9. The continuous scores were randomized with respect to the control group. At least 30 JD (95 USD PPP) in savings (=1) is an indicator equal to one if the respondent answered yes to “Do you currently have at least 30 JD (95 USD PPP) in personal savings you can draw from in an emergency? Whether or not it is in a bank?”. Respondent, adult, and child hunger last week are indicators equal to 1 if the respondent, other adults in the household or children in the household (respectively) went to bed hungry at least once in the last 7 days, zero otherwise. In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to various families in the pre-analysis plan. Because not all outcomes are available in multiple rounds, some families include fewer outcomes in the pooled estimates than the round-by-round estimates.

Table (3.5) Statistically Significant Impacts of the Program at Midline

Outcome	TOT	(se)	FDR q-values	Control mean	(sd)	N
Panel A: Housing Improvements						
Quality Roof (=1)	0.24***	(0.09)	[0.08]	0.71	(0.45)	1,610
Clean water (=1)	0.21***	(0.06)	[0.01]	0.17	(0.38)	1,610
Occupied Rooms	0.51***	(0.17)	[0.04]	2.85	(1.05)	1,610
Panel B: Food Aid and Insecurity						
Food consumption (aid) USD PPP	-53.05**	(22.40)	[0.02]	218.51	(201.68)	1,569
Food consumption (aid) USD PPP (<i>Per capita</i>)	-9.73***	(3.44)	[0.01]	35.88	(29.19)	1,569
Respondent hunger last week (=1)	0.23***	(0.07)	[0.01]	0.35	(0.48)	1,604
Adults hunger last week (=1)	0.17***	(0.06)	[0.01]	0.38	(0.49)	1,610
Child hunger last week (=1)	0.24***	(0.07)	[0.00]	0.23	(0.42)	1,469
Panel C: Household Recomposition						
Number of children under 18 years	0.30**	(0.15)	[0.03]	3.27	(1.97)	1,610
Number of boys aged 13-17 years in household	0.16**	(0.07)	[0.33]	0.40	(0.66)	1,610
Panel D: COVID-19 Outcomes						
Weekly adult income (during first lockdown)	-20.07***	(7.19)	[0.02]	30.22	(66.86)	1,609
Total COVID-19 symptoms	0.84***	(0.19)	[0.00]	0.47	(1.22)	1,610
Number of people who are symptomatic	0.46***	(0.09)	[0.00]	0.29	(0.61)	1,581
Number of children who are symptomatic	0.17***	(0.05)	[0.01]	0.06	(0.32)	1,472
Number of visits to healthcare institutions	-0.56***	(0.23)	[0.04]	1.14	(2.03)	1,608
Did not keep distance (=1)	0.16***	(0.06)	[0.08]	0.51	(0.50)	1,609
Panel E: Loans						
Loans taken pre-first-lockdown (=1)	-0.20***	(0.07)	[0.01]	0.61	(0.49)	1,608
Loans taken during first lockdown (=1)	-0.34***	(0.07)	[0.00]	0.78	(0.41)	1,608
Loans taken after first lockdown (=1)	-0.33***	(0.07)	[0.00]	0.42	(0.49)	1,607
Panel F: Other effects						
Total number of school days attended per child	-0.43*	(0.24)	[0.04]	2.33	(2.49)	1,031
Child alertness (s.d. units)	-0.73***	(0.20)	[0.00]	0.00	(1.00)	975
Difficulty with self-care (z-score)	0.38***	(0.14)	[0.08]	-0.00	(1.00)	1,610
Number of learning activities	0.19**	(0.09)	[0.08]	0.22	(0.56)	1,385

Notes: The table shows the regression results on statistically significant results from the midline survey. Each row is its own dependent variable. **Panel A** outcomes are Clean Water, defined as an indicator for households having treated drinking water (such as by a filter). Rooms and rooms occupied by family members are reported by the respondents. **Panel B** outcomes are food consumption, Respondent (and adult) food insecurity equals 1 if the respondent went to bed hungry on at least one day in the last week and 0 otherwise. **Panel C** outcomes are COVID symptoms, which is the sum of indicators that increase the likelihood of COVID contraction by the respondent including leaving the house, attending social gatherings, not keeping distance from others, going to mosque or other religious institutions, going to grocery store or market, and leaving village/neighborhood. The other outcomes are as explained in the labels. **Panel D** The periods of the loans referred to are before lockdown: (January 15 - March 15, 2020); during lockdown: (March 15 - May 15, 2020); and after lockdown: (May 15 - interview date). **Panel E** outcomes are self-explanatory. **Panel F** Difficulty with self-care with higher values indicating more disability. This is an item of the The Washington Group Short Set on Functioning; learning activities are done in the last 24 hours and include homework, e-learning, educational programs and videos, and reading. Monetary values are in USD PPP and are winsorized at the top 1% of values in order to limit the influence of outliers. Health visits is also winsorized at the top 1% due to outliers in that measure. The regressions also have month-by-year fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to various families in the pre-analysis plan.

Table (3.6) Statistically Significant Impacts of the Program at Endline and Follow-up

Outcome	TOT	(se)	FDR q-values	Control mean	(sd)	N
Panel A: Endline						
Respondent Childcare & Chores (Hours)	-6.91***	(2.02)	[0.01]	17.76	(18.39)	1,418
Total Labor and Chores Hours Last Week	-10.62***	(3.24)	[0.01]	25.54	(23.39)	1,422
General Happiness (Scale 1-3)	-0.22***	(0.08)	[0.02]	1.76	(0.61)	1,422
At least 30 JD (95 USD PPP) in savings (=1)	0.14***	(0.05)	[0.02]	0.12	(0.32)	1,422
Applied but Did Not Take Loan (=1)	0.06***	(0.02)	[0.00]	0.01	(0.10)	1,422
IP Assistance in Last 12 Months (=1)	0.08**	(0.03)	[0.14]	0.04	(0.19)	1,422
IP Assistance in Last 12 Months (USD PPP)	128.25***	(46.50)	[0.10]	28.74	(198.87)	1,422
Panel B: Follow-up						
Child SDQ score, std	-0.56***	(0.17)	[0.01]	-0.01	(1.00)	859

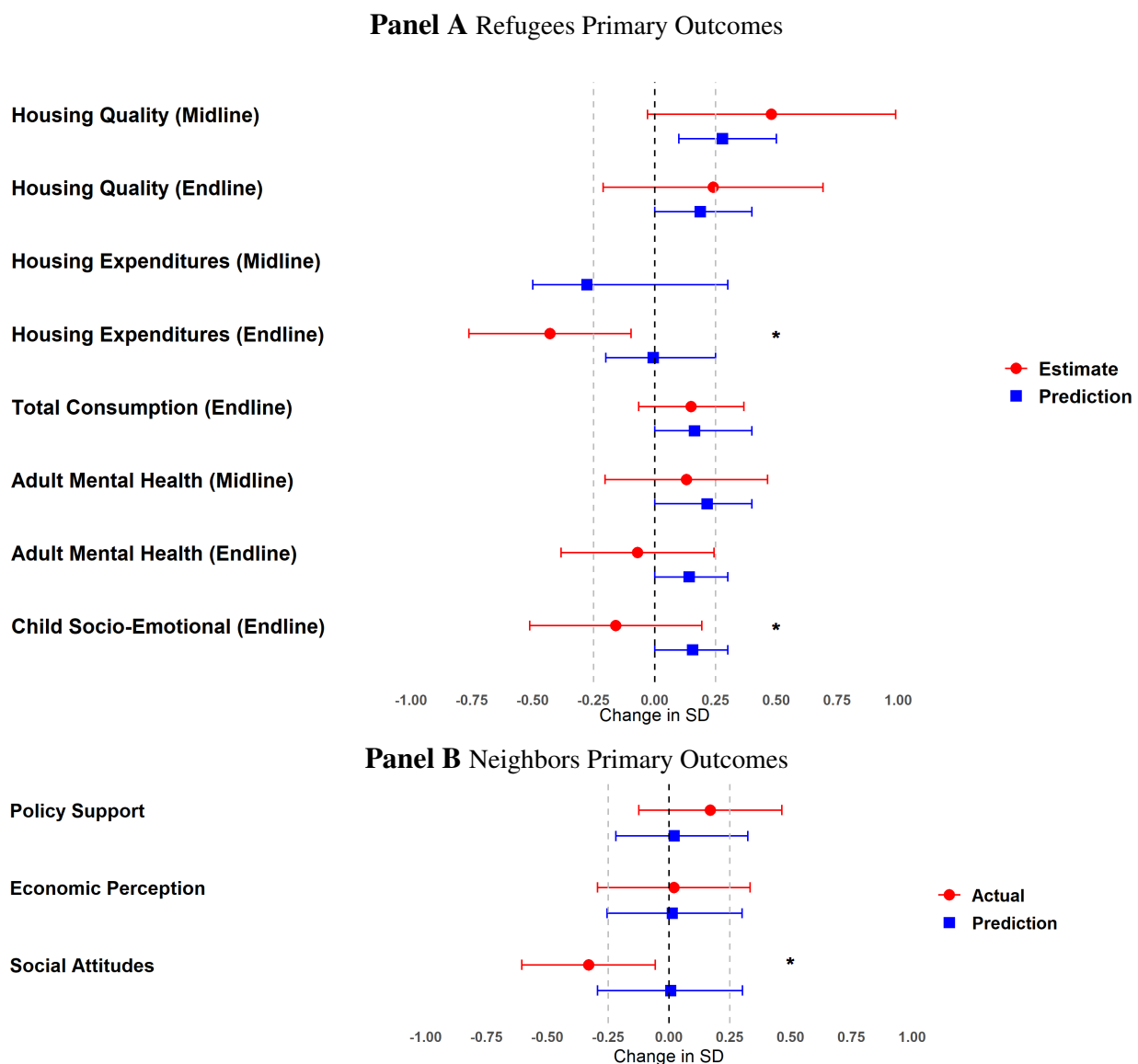
Notes: The table shows the regression results on statistically significant results from the endline survey. Each row is its own dependent variable. IP refers to the implementing partner. Monetary values are in USD PPP and are winsorized at the top 1% of values in order to limit the influence of outliers. The regressions also have month-by-year fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to various families in the pre-analysis plan. Focus Respondent Childcare & Chores (Hours) was pre-specified in two outcome families; the table reports the lower of the two q-values; the higher q-value is 0.02.

Table (3.7) Impacts of the Program on Neighbor's Attitudes Toward Refugees

Outcome	TOT	(se)	FDR q-values	Control mean	(sd)	N
Panel A: Primary Outcomes						
Social attitudes & perceptions (SD)	-0.33**	(0.14)	[0.08]	-0.00	(1.00)	1,102
Economic attitudes & perceptions (SD)	0.02	(0.16)	[0.56]	0.02	(1.02)	1,102
Policy preferences (SD)	0.17	(0.15)	[0.37]	-0.02	(1.00)	1,102
Altruism to Syrians	0.21	(0.18)	[0.37]	0.88	(1.18)	1,102
Panel B: Selected Social Attitudes Index Components						
Of the 3 people you exchange advice with, how many are Syrian refugees?	-0.24**	(0.11)	[0.32]	0.34	(0.72)	1,102
Do the children in this household have any Syrian refugee friends?	-0.19*	(0.11)	[0.37]	0.40	(0.49)	704
Panel C: Other Related Outcomes						
Days of media consumption (last week)	-1.08***	(0.40)	[0.08]	4.78	(2.78)	1,102
Neighbor perceptions of average refugee aid receipt (PPP)	-347.17***	(109.69)	[0.04]	630.84	(657.41)	685
Life Satisfaction (SD)	-0.32**	(0.15)	[0.06]	0.01	(0.98)	1,097
Subjective Health (SD)	-0.30*	(0.16)	[0.06]	0.02	(0.99)	1,102

Notes: This table reports impacts of treatment on Jordanian neighbors of Syrian refugees in the experimental sample. **Panel A** reports pre-specified treatment effects on primary neighbor outcomes. All indices are standardized so that a positive point estimate reflects more “pro-refugee” sentiments. The only statistically significant outcome is a negative treatment effect on social attitudes and perceptions. This outcome is a standardized index of the following outcomes: out of 3 closest friends, how many are Syrian refugees; out of 3 people you exchange advice with, how many are Syrian refugees; do the children in this household have Syrian refugee friends (=1); do the children in this household share recreational spaces with Syrian refugee children (=1); how comfortable would you be accepting the marriage of a friend or loved one to a Syrian refugee (1-5); how comfortable would you be accepting a Syrian refugee as a neighbor (1-5); what is the net effect of Syrian refugees on Jordanian society (1-3); are Syrian refugees hardworking or lazy (1-7). Economic attitudes and perceptions is a standardized index of the following outcomes: an indicator =1 if the respondent listed “hosting Syrian refugees” as one of the most important challenges facing Jordan (options not read); an outcome with higher values for responses associated with beliefs that Syrian refugees pay more in taxes than Jordanians; an outcome with higher values for responses associated with refugees having a positive effect on the economy. Policy preferences in a standardized index of the following outcomes: less belief that Syrian refugees should be forced to live in camps; more belief that refugees should have the right to work outside camps; more belief that refugees should be able to attain Jordanian citizenship; more support for unrestricted work permits for refugees; more support for integrated classrooms with Syrians and Jordanians; more support for refugee right to enter/exit camps freely; support for housing assistance for Syrian refugees; belief that the international community should spend more money supporting refugees. Altruism to Syrians reports how many JD out of 5 the respondent allocated to a charity supporting Syrian refugees. The other two options were allocating to a charity supporting low-income Jordanians or keeping for self. **Panel B** reports the two significant subcomponents of the social attitudes and perceptions index. **Panel C** reports other statistically significant treatment effects on neighbors. Subjective health is the response to “Would you describe your general health as good, fair, poor, or very poor?” Life satisfaction is the response to “All things considered, how satisfied are you with your life as a whole these days on a scale of 1 to 10?” Days of media consumption is the response to “In the past 7 days, how many days did you read or listen to the news from any source, including newspapers, online, WhatsApp, etc.?” Neighborhood perceptions of refugee aid receipt is the response to a question asking “Of the refugee households in your neighborhood who receive assistance, what do you think is the average value in Dinar of the assistance (in cash or in kind) that they receive from organizations in a typical month?” The sample size for this question is smaller due to the large number of “don’t know” responses. There is no treatment effect on the probability of reporting “don’t know” to this question.

In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values in panel A and C are calculated per Anderson (2008) using the outcome families in the pre-analysis plan. Panel B was not prespecified and is reported to aid in interpretation of the negative treatment effect on social attitudes and perceptions. The q-values are estimated using an outcome family comprised of all the components of the social attitudes and perceptions index.

Figure (3.3) Program Impacts vs. Expert Predictions

Notes for Panel A: 10-90th Percentile of Predictions ($n = 61$), 95% CI of Actual Outcomes, 4-12 months ($n = 1610$), 12-18 months ($n \sim 1395$), Housing Expenditures (Midline) was not measured in a comparable fashion (see text for details) and is not shown here, * Denotes Estimate = Prediction rejected at $p < 0.05$.

Notes for Panel B: 10-90th Percentile of Predictions ($n = 63$), 95% CI of Actual Outcomes ($n = 1102$), * Denotes Estimate = Prediction rejected at $p < 0.05$.

Chapter 4

Digital Payments Adoption

4.1 Introduction

Public service delivery in developing countries imposes large transaction costs on the citizens it is meant to serve. Whether through the physical tax of long queues and travel to distant payment locations or the financial strain of severe administrative bottlenecks and delayed wages (Muralidharan et al., 2016; Banerjee et al., 2020), these systemic frictions fall disproportionately on those who can least afford to bear them. Water provision provides services that are fundamental to household welfare, yielding significant returns in terms of physical health (Galiani et al., 2005; Kremer et al., 2011) and time savings and subjective well-being (Devoto et al., 2012). However, sustaining and expanding these critical services requires reliable revenue collection. For these utilities, bill payment is among the most frequent points of contact with urban households. The cumulative cost of these routine, in-person interactions for both the household and the state is large enough that even modest improvements to the payment interface can generate first-order fiscal and welfare effects.

Governments across low- and middle-income countries have been investing in digital infrastructure for public service delivery. Such digital payments have been shown to have positive impacts, such as decreased time costs and bribes for taxes (Okunogbe and Pouliquen, 2022), reduced administrative costs and time for social transfers (Aker et al., 2016), curbed diversion to black markets (Barnwal, 2024), lower leakage in welfare disbursement (Muralidharan et al., 2016) and workfare programs (Banerjee et al., 2020), and increased savings among low-income households (Bachas et al., 2021). The common thread across these supply-side reforms is how digital infrastructure structurally alters market incentives: by directly linking benefits to verified individuals, these systems eliminate the dual-pricing and informational asymmetries that intermediaries historically exploited for arbitrage.

Yet adoption is often slow: About 620 million adults pay their utility bills in cash, even though they have bank accounts (Demirgüç-Kunt et al., 2022). Thus, even when mobile money penetration is high and digital payment for government services is technically available, large shares of urban households continue to settle bills in person. This gap between availability and use is particularly costly in the utility sector: it directly constrains the operational efficiency gains that motivated the

digital investment, and it sustains the queuing burden on households that the new infrastructure was designed to eliminate. The Addis Ababa Water and Sewerage Authority (AAWSA), the public utility serving Ethiopia's capital, illustrates the problem starkly: it serves roughly 659,000 customer accounts, has had digital payment channels available, and yet a share of residential customers continue to pay exclusively at bank branches month after month, generating substantial costs for the utility and queuing costs for households.

This paper studies what it takes to close that gap. In partnership with AAWSA, we conduct a randomized controlled trial on the adoption of electronic payment for water bills in Addis Ababa. Our sample comprises almost 600 residential customers identified from administrative records as having paid exclusively at bank branches for at least six consecutive months. We randomly assign these households to one of three groups: a control group receiving no intervention; an information group receiving step-by-step instructions on how to pay digitally; and an incentive group receiving a 50% discount on electronic payment (or roughly \$1.97). Both interventions were conducted for two consecutive billing cycles (December 2025 and January 2026). Primary outcomes are measured directly from AAWSA's administrative billing records over three billing cycles, avoiding reliance on self-reports.

Both interventions increased digital adoption, with the incentive roughly twice as effective as information. After the first offer, 7% of control-group households paid their December bill digitally; the information treatment raised this by 6.8 percentage points (pp) ($p < 0.05$) and the incentive by 11.6 pp ($p < 0.01$). After both interventions were re-offered, the control group's January digital payment rate was 8.9%, and the information and incentive treatments increased it by 9.3 ($p < 0.01$) and 16.7 pp ($p < 0.01$) respectively. The difference between the two treatments is marginally significant after the second dose ($p = 0.075$). Critically, neither treatment reduced overall bill payment compliance: the interventions shifted the *mode* of payment, not whether households paid at all.

Adoption effects did not fade after the incentive was withdrawn. In the first post-treatment billing cycle (February 2026), the incentive arm's digital adoption rate exceeded the control group by 22.5 pp ($p < 0.01$), and the information treatment effect grew to 13.3 pp ($p < 0.01$). Persistence is consistent with the broader administrative record: 96.6% of customers who paid digitally in March 2025 continued to do so ten months later. Together these patterns indicate that a temporary incentive delivers a durable behavioral shift, with the effect compounding rather than decaying after the incentive is withdrawn.

The channel decomposition reinforces the cost-of-transacting interpretation. After the first offer, new digital adopters split roughly evenly between telebirr (Ethiopia's public mobile money platform, comparable to Kenya's private M-Pesa) and the Commercial Bank of Ethiopia (CBE) smartphone app, each rising by about 6 pp ($p < 0.05$). By the second billing cycle, telebirr had become the dominant channel, with a 13 pp treatment effect ($p < 0.001$) from a control-group base of zero, while the CBE app effect became insignificant. Simultaneously, the incentive reduced in-person payments at CBE bank branches by 9.4 pp ($p = 0.038$) and at Awash Bank front desks by a similar margin ($p = 0.069$). Households gravitated to the lowest-cost channel: one accessible via

a USSD code on any handset, with no smartphone or app required.¹

To contextualize these findings against prior expectations, we elicited predictions from researchers via the Social Science Prediction Platform. Experts accurately anticipated the magnitude of the information treatment but systematically misjudged the dynamics of the financial incentive. Forecasters expected an immediate response that would subsequently decay once the offer was withdrawn, a trajectory consistent with standard temporary price effects. Instead, actual adoption was initially more gradual but compounded over time, growing even after the incentive ended. This divergence highlights that experts' priors focused on households simply responding to a temporary price variation instead of overcoming a behavioral fixed cost.

Cost-effectiveness calculations imply roughly \$33.6 per durable adopter for the incentive versus \$93.1 for information, before accounting for downstream administrative savings to the utility. Given the near-absorbing nature of digital payment in the administrative universe, even a one-time temporary incentive can be expected to deliver lasting reductions in the in-person transaction load.

This paper makes three contributions. First, we contribute to the literature on digitization in low- and middle-income countries. Existing work has focused predominantly on the supply side: how digital infrastructure can reduce leakage (Muralidharan et al., 2016; Banerjee et al., 2020; Barnwal, 2024) and lower costs and curb corruption (Aker et al., 2016; Okunogbe and Pouliquen, 2022). We provide complementary evidence on the demand side, asking what it takes to move habitual in-person users onto a digital channel that the state has already built. The closest precedents are Aggarwal et al. (2020), who provide mobile money accounts and training to micro-entrepreneurs, and Riley et al. (2025), who test financial incentives and peer endorsement for mobile banking adoption. Our study is distinct in targeting a recurring household-level transaction with the state, among a population specifically selected for persistent non-adoption, and in measuring outcomes from administrative records spanning the universe of customers in a single utility.

Second, we contribute to the literature on public service delivery. Existing experimental work in this area has focused on water access and infrastructure investment (Galiani et al., 2005; Devoto et al., 2012; Guiteras et al., 2015), on water quality and downstream health outcomes (Kremer et al., 2011; Madajewicz et al., 2007), and on conservation through social norms (Ferraro and Price, 2013) and bill compliance through education (Szabó and Ujhelyi, 2015). Our study addresses a different margin: not whether households pay, but *how* they pay. The distinction matters for utility operations because shifting payment mode from in-person to digital reduces transaction costs for both sides, while the near-universal payment compliance in our sample (94.5% in the control group) means there is little room to improve on the extensive margin. To our knowledge, this is the first experimental evidence on the payment-channel margin in a developing-country residential water utility.

Third, we contribute to the literature on fixed costs of adoption and switching frictions in recurring transactions (Handel, 2013; Ericson, 2014). A related literature documents that temporary interventions which move households across this fixed-cost threshold can generate persistent be-

¹USSD (Unstructured Supplementary Service Data) is a menu-based protocol accessed by dialing a short code on any mobile phone, including basic handsets without internet.

havioral change after the intervention is withdrawn (Dupas, 2014; Royer et al., 2015; Schaner, 2018; Celhay et al., 2019). We refine the standard fixed-cost-of-adoption interpretation in two ways. First, we show that cognitive and financial components of the bundled cost can both bind in the same setting: information alone produces roughly half the effect of the incentive, indicating that knowledge is itself a binding constraint, while the incentive's larger effect indicates that the financial component is more binding. This contrasts with Cole et al. (2011), who find that financial literacy has little effect on bank account adoption while price subsidies dominate. Second, we document that the response to a temporary financial incentive can be *growing* rather than merely persistent, driven by delayed crossing of the threshold among households moved closer to it during the offer.

The remainder of the paper is organized as follows. Section 4.2 describes the institutional context. Section 4.3 details the experimental design. Section 4.4 describes the data sources and sample characteristics. Section 4.5 presents the empirical strategy. Section 4.6 reports the main results, Section 4.7 examines mechanisms, Section 4.9 reports cost-effectiveness, and Section 4.10 concludes.

4.2 Institutional Setting

The Addis Ababa Water and Sewerage Authority (AAWSA) is the public utility serving Ethiopia's capital, a city of roughly 6.7 million residents (United Nations, 2025). AAWSA reports about 659,000 active customer accounts, including roughly 553,000 residential accounts, 104,000 non-residential accounts, and 2,600 shared fountain connections, organized across ten branch offices.

According to AAWSA estimates, total production is approximately 630,000 m³ of water per day against estimated daily demand of about 1.2 million m³, implying a nominal coverage rate near 65%. In practice, this supply-demand gap means many households receive piped water only two to three days per week. AAWSA operates under an increasing-block-tariff structure intended to cross-subsidize low-volume users. The utility also faces two reinforcing fiscal pressures. First, the utility and city sources describe tariffs as recovering less than half of service cost. Second, non-revenue water (pumped but not billed) from leakage, aging infrastructure, and unauthorized connections is reported at 35–38% of production as per AAWSA estimates. Together, these facts imply that AAWSA loses money both on a large share of the water it delivers and on a large share of the water it never successfully bills.

Bill payment has also been a persistent friction point for customers. From 2013 to 2019, utility payments in Addis Ababa were consolidated through Lehulu, a one-stop payment platform operated by Kifiya Financial Technology under a public-private partnership. Public reporting on the platform indicates that, at its peak, Lehulu processed payments for roughly one million customers per month across 32 branches in Addis Ababa. When Kifiya's contract expired in August 2019, payment responsibilities reverted to individual utility offices, reportedly generating long queues and substantial customer frustration. Today, customers can pay water bills at commercial bank front desks, via CBE's Birr mobile banking, telebirr, or through Awash Bank's payment channels.

AAWSA serves roughly 659,000 customer accounts in a city of over six million, and it has an acute institutional interest in moving customers online: every in-person transaction requires manual reconciliation and imposes queuing costs on customers, while digital payments reduce administrative overhead and generate real-time payment records. Digital payment channels have been available for nearly a decade, yet as of 2025, many residential customers had never used them. If digital adoption proves sticky and administrative data from AAWSA's full customer base show that 96.6% of customers who paid digitally in March 2025 continued to do so in January 2026, then even a temporary intervention that induces a first digital transaction could generate lasting behavioral change and meaningful reductions in administrative costs for the utility.

In April 2025, AAWSA signed an additional agreement with Awash Bank to enable payment through its banking channels. Yet in-person payment at branch offices remains common.² A plausible explanation is that digital payment still requires customers to know their customer key number, trust electronic confirmation for an essential service, and navigate mobile-wallet interfaces that are not equally familiar across households. These reasons motivate the experimental design.

4.3 Experimental Design

Sample and Randomization

Our sampling frame is drawn from the residential segment of AAWSA's administrative universe, which comprises approximately 553,000 accounts. Using monthly payment channel records, we identified customers who paid exclusively through in-person channels. We defined in-person channels as Awash Bank front desk (Teller) or Commercial Bank of Ethiopia branch (CBEER), in every month from March through August 2025, the six months for which channel data were available at the time of sample selection. Customers who used any digital channel in any of these six months, or who failed to make a payment in any month, were excluded. This strict filter yielded a frame of 2,423 persistently in-person-paying households.

We then stratified this frame along two dimensions: geographic location (AAWSA branch, of which there are ten) and water consumption level (above or below the median of average monthly consumption across the six observed months), producing 20 strata. Within each stratum, we randomly selected 30 households (10 per treatment arm, more below) for the main sample, plus an equally sized replacement pool where stratum size permitted.³ This procedure yielded a main experimental sample of almost 600 households.

1. **Control group** ($n = 197$): Received no intervention.

²We are not aware of a published official series on the share of AAWSA payments made via mobile channels. Establishing that baseline is one motivation for the present study.

³Two strata with fewer than 60 eligible households received adjusted allocations: one stratum contributed 27 households (9 per arm, no replacements) and another contributed 47 (10 per arm for the main sample with a reduced replacement pool).

2. **Information treatment** ($n = 202$): Received detailed information about electronic payment options, including step-by-step instructions on how to pay water bills through mobile banking, internet banking, and agent networks. Information was delivered through in-person visits and printed materials for two consecutive billing cycles (December 2025 and January 2026).
3. **Incentive treatment** ($n = 198$): Received a 50% discount on electronic water bill payment for the same two consecutive billing cycles, with average transfer of around \$1.97 per household. The incentive was re-offered in the second month regardless of whether the household had adopted in the first cycle.

Treatment Implementation

Both treatments were delivered by trained field enumerators at the end of the baseline interview in December 2025, ensuring that baseline covariates were measured before any intervention exposure. Figure 4.1 shows the timeline. Enumerators provided households with step-by-step instructions on how to pay water bills electronically through mobile banking, telebirr, and agent networks. Printed instructional materials were left with the household; delivery was confirmed for 98.5% of targeted households. Households in the incentive arm received an offer of a 50% reduction on their water bill conditional on making the payment through a digital channel. The incentive was initially offered for the December 2025 bill and re-offered for the January 2026 bill, regardless of whether the household adopted in the first cycle. Reimbursement was made by the research project upon verification of an electronic payment in the AAWSA administrative records.

The endline survey is administered approximately in February–March 2026 and provides the post-treatment measurement window for secondary outcomes. The primary outcomes are observed in administrative data across three billing cycles following treatment delivery.

Outcomes

Our *primary outcomes* are binary indicators equal to one if the household paid its December 2025 or January–February 2026 water bill electronically (through any digital channel) and zero otherwise. The December outcome captures the effect of the first incentive offer; the January outcome captures the cumulative effect after both offers; and the February outcome captures the first post-treatment billing cycle and provides a direct test of persistence. We also have data on monthly bill payments and consumption from the same admin data.

Secondary outcomes are drawn from the endline survey, administered two to three months post-treatment (February–March 2026). These include: stated plans to adopt electronic payment, willingness to pay digitally, self-reported knowledge of how to pay electronically, awareness of electronic payment options, mobile money account ownership, and transaction costs (transportation costs in USD and travel time in minutes to pay the bill).

Administrative Context

To situate the experiment within the broader landscape of digital payment at AAWSA, we examine macro-trends in the full residential customer universe. Figure 4.2 decomposes the payment behavior of the full AAWSA residential customer universe across available data months. The digital payment rate is broadly stable at 50-60%, while in-person payment at bank front desks (Teller and CBEBR channels) holds steady at approximately 5% of the universe. The remaining 35-45% of customers either did not pay in a given month or cannot be cleanly classified. The stability of the in-person share underscores that a persistent segment of customers uses physical payment channels despite the availability of digital alternatives, precisely the behavioral inertia our experiment targets. The universe-wide in-person rate is substantially lower than the rate in our experimental sample, which was drawn from customers identified as paying exclusively at bank front desks across all six months with available channel data.

To contextualize the external validity of our experiment, Figure C.1 in the appendix plots the distributions of January water consumption (Panel a) and bill amounts (Panel b) for our experimental sample against the full AAWSA residential administrative universe. The overlaid densities demonstrate that our experimental sample sits comfortably within the primary support of the population distribution in terms of consumption and bill amounts, with near-complete overlap across both variables.

4.4 Data, Descriptives, and Internal Validity

Data Sources

We draw on three data sources: a baseline household survey, an endline follow-up survey, and monthly administrative records from AAWSA. Baseline and endline survey responses are linked by a unique customer key, which means that survey records are merged with the administrative data using the same identifier. The administrative data provide the primary outcome along with continuous measures of consumption and billing; the survey data provide baseline covariates and secondary attitudinal outcomes.

Baseline survey. The baseline survey was administered in December 2025 by trained enumerators who visited sampled households in person. A total of 597 households consented to participate. The instrument collected household demographics (age, gender, education, household size), asset ownership, water use patterns and billing history, digital literacy and mobile phone access, current payment behavior and barriers to electronic payment, and measures of trust in digital systems and willingness to pay electronically. Treatment was delivered after the baseline interview, ensuring that pre-treatment covariates were measured before any intervention exposure.

Endline survey. The endline survey was conducted approximately two to three months later, in February–March 2026. Of the 597 original households, 547 completed the endline. The endline

re-measured attitudinal outcomes such as stated plans to adopt electronic payment, willingness to pay digitally, trust in electronic systems and collected information on mobile money account status, transport costs incurred for bill payment, and awareness of digital payment channels.

Universe of billing data. We obtained monthly payment records from AAWSA covering March–August 2025 and December–February 2026 for Addis Ababa’s full administrative universe of approximately 659,000 customer accounts. These records contain the payment channel used (digital or in-person), monthly water consumption in cubic meters, and the billed amount (converted to USD). Our primary outcomes (whether a household paid its bills electronically) are constructed from these administrative records rather than self-report, avoiding potential measurement error from recall.

Sample Characteristics and Balance

Before examining balance across experimental arms, we briefly characterize the dynamics of digital payment adoption in the broader administrative universe. This context is useful for interpreting both the level of the control-group adoption rate in our experiment and the potential for short-term interventions to generate lasting behavioral change.

Table 4.1 presents a transition matrix of payment modes between March 2025 (the first month in the dataset) and January 2026 among the approximately 246,000 residential customers with a cleanly classified payment channel in both months.⁴ Two patterns stand out. First, digital payment is nearly an absorbing state: 96.6% of customers who paid digitally in March continued to do so in January, with only 3.4% reverting to in-person channels. This persistence suggests strong habit formation once the initial adoption barrier is crossed, and implies a high return on investment for even short-lived interventions that induce a first digital transaction. Second, in-person payment is far less sticky: 68.6% of customers who paid at a bank front desk in March had switched to a digital channel by January, while only 31.4% remained in person. Some of this transition reflects broader time trends and channel rollouts during the ten-month window rather than purely individual responsiveness to a fixed choice environment. Even so, the magnitude indicates that habitual in-person payment is not a permanent state for most customers, which is precisely the margin our information and incentive treatments target.

Table C.1 in the appendix reports means, standard deviations, and pairwise tests of equality for thirteen baseline covariates across the three experimental arms. All regressions control for strata fixed effects and use robust standard errors. The sample is predominantly female (60–66%), with an average age of approximately 50 years and a household size between 4.3 and 4.8 members. Nearly all respondents hold a bank account, roughly two-thirds report mobile-money access, and about one-third have internet at home. The mean asset index ranges from 2.3 to 2.5 (out of a possible 6), and average monthly water bills range from roughly \$2.28 to \$2.77, though with substantial dispersion (standard deviations of \$3.64 to \$5.15).

⁴Customers with missing or unrecognized channel codes and those who did not pay in either month are excluded from this table.

Across the thirteen covariates, only one pairwise difference is statistically significant at conventional levels: household size is larger in the incentive arm than in the information arm ($p = 0.01$). Given that we test 39 pairwise comparisons (13 variables $\times$ 3 pairs), one rejection at the 5% level is well within the range expected by chance under the null of no differences. Crucially, the joint F -test of orthogonality from a regression of each treatment indicator on all baseline covariates simultaneously fails to reject the null in every case: $p = 1.000$ for control vs. information, $p = 1.000$ for control vs. incentive, and $p = 0.997$ for information vs. incentive. We conclude that randomization achieved balance and hence we do not include further controls beyond stratification indicators in the estimating equations.

Attrition

Table C.2 in the appendix reports attrition rates by treatment arm. Of the initial 590 sample, 540 (91.5%) completed the endline survey.⁵ Attrition is balanced across arms: 7.9% in control ($n = 190$), 7.9% in information ($n = 202$), and 9.6% in incentive ($n = 198$). None of the pairwise differences are statistically significant. To get a sense of how attriters differ from non-attriters, Table C.3 compares baseline characteristics of attriters and stayers. Several differences are noteworthy: all attriters hold bank accounts compared to 95.0% of stayers ($p < 0.001$), and attriters have a lower asset index ($p = 0.030$). Other differences (gender, water bills, shared meter, late fees) are not statistically significant.

Finally, the joint F -tests for differential attrition (testing whether treatment-covariate interactions predict attrition) yield $p = 0.846$ for control vs. information, $p = 0.604$ for control vs. incentive, and $p = 0.830$ for information vs. incentive, providing evidence that attrition is uncorrelated with both treatment assignment and baseline characteristics. We therefore proceed with the available sample, noting that the balanced attrition pattern does not suggest a threat to identification.

4.5 Empirical Strategy

Our primary estimating equation is:

$$Y_i = \alpha + \beta_1 \text{Info}_i + \beta_2 \text{Incentive}_i + \gamma_{s(i)} + \varepsilon_i \quad (4.1)$$

where Y_i is the outcome for household i , Info_i and Incentive_i are binary indicators for assignment to the information and incentive treatments, respectively, γ_s denotes strata fixed effects (to account for the stratified randomization design), and ε_i is an error term. The coefficients β_1 and β_2 are the intent-to-treat (ITT) effects of the information and incentive interventions, respectively, relative to the control group. Note that for our primary outcome, we have three rounds of post-treatment data, and we run cross-sectional regressions on each separately.

⁵Recall that our initial sample is 597 households. However, seven control-group households were found to have received bank-assisted digital payment training during the study period (at bank offices, while paying in person); we exclude these from the analysis as contaminated controls, leaving an analysis sample of 590.

We use the linear probability model (LPM) rather than logit or probit because the coefficients are directly interpretable as percentage-point changes in adoption probability and for consistency with other continuous outcomes. We report heteroskedasticity-robust standard errors throughout. To examine heterogeneous effects, we run separate regressions for baseline characteristics (gender, education, asset holdings, and age) and estimate all models with strata fixed effects.

4.6 Results

We organize results in three parts. First, we examine treatment effects on primary outcomes measured from AAWSA administrative records: likelihood of digital payment adoption, likelihood of any bill payment, water consumption, and bill amounts. Second, we report effects on secondary outcomes from the endline survey: stated plans, willingness, trust, mobile money access, and transport costs. Third, we explore heterogeneity across different subgroups. Note that these dimensions were not pre-specified and should be interpreted as exploratory.

For the main treatment effects, we also break down results by payment channel. AAWSA's administrative records classify each payment into one of several channels. We group these into five categories: *telebirr* (mobile money platform), *CBE app* (Commercial Bank of Ethiopia mobile banking, including CBE Birr, CBE Internet, and CBE standing instruction), *CBE branch* (in-person payment at a CBE front desk), *Teller* (in-person payment at an Awash Bank front desk), and *AwashPay* (Awash Bank's mobile banking channel).⁶ Of these, *telebirr*, *CBE app*, and *AwashPay* are digital; *CBE branch* and *Teller* are in person.

Main Treatment Effects

Table 4.2 presents intent-to-treat estimates of the two interventions on primary outcomes measured from administrative records. All regressions include strata fixed effects and report heteroskedasticity-robust standard errors.

Electronic payment adoption. Table 4.2 reports treatment effects for the three billing cycles. After the first incentive offer, the control group's December digital payment rate was approximately 7%. The information treatment increased this by 6.8 pp ($p < 0.05$) and the incentive by 11.6 pp ($p < 0.01$). After the incentive was re-offered, adoption grew: the control group's January rate was 8.9%, and the information and incentive treatments increased it by 9.3 ($p < 0.01$) and 16.7 pp ($p < 0.01$), respectively. The incentive nearly tripled the January adoption rate relative to the control mean. The incremental effect of the second offer was approximately 5 pp for the incentive and 2.5 for information, suggesting that the reminder amplified but did not create the treatment effect. The difference between information and incentive effects is marginally significant in January ($p = 0.075$).

⁶The administrative data record the channel used for each billing cycle. Of the 590 sample households, 26 did not pay the December bill and 33 did not pay the January bill; these are excluded from the channel regressions.

The February billing cycle which is the first after the incentive period ended provides a direct test of persistence. Rather than reverting toward baseline, treatment effects continued to grow: the information treatment increased February digital adoption by 13.3 pp ($p = 0.001$) and the incentive by 22.5 pp ($p < 0.001$), both substantially larger than the corresponding January effects. The February control-group adoption rate was 11.9%, implying that the incentive arm's adoption rate exceeded 34%.

The continued strengthening of treatment effects after incentive withdrawal is consistent with two complementary mechanisms: habit formation among first-cycle adopters who continued paying digitally, and delayed adoption among households that received the intervention but did not act until the second or third billing cycle. Section 4.7 later shows that habit formation is real (digital payment is sticky once tried) but that the differential between the treatment arms in February operates almost entirely through the delayed-adoption channel rather than through differential stickiness.

Figure 4.3 plots the trajectory of digital adoption rates across the three billing cycles by treatment arm. Two patterns are worth noting. First, adoption rises across the treatments and between the December and January billing cycles, consistent with both the repeat intervention delivered in late January and early February and with experimentation following initial exposure. Second, the incentive arm separates sharply from the other two arms by the January bill and maintains its advantage through the February bill. Since this bill is the first post-treatment billing cycle, the pattern suggests that adoption induced by the financial incentive persisted after the incentive was removed. The information arm likewise sustains its gains through February.

Channel decomposition. Figure 4.4 decomposes the treatment effect by payment channel across both billing cycles. The pattern shifted notably between the two offers. After the first incentive (Panel a), new digital adoption was split between telebirr (+5.5 pp, $p = 0.006$) and the CBE smartphone app (+6.0 pp, $p = 0.038$), with a marginally significant reduction in Awash Bank front desk (Teller) payments (−8.5 pp, $p = 0.096$). After the second offer (Panel b), telebirr surged to +13.0 pp ($p < 0.001$) while the CBE app effect became insignificant (+4.2 pp, $p = 0.203$). Simultaneously, the incentive now significantly reduced CBE branch payments (−9.4 pp, $p = 0.038$), an effect absent in the first cycle, and the Teller reduction persisted (−9.4 pp, $p = 0.069$). The information treatment followed a similar pattern but had a slightly more attenuated impact. The evolution suggests that the first incentive prompted experimentation across multiple digital channels, while the second offer consolidated adoption onto telebirr, which is the lowest-barrier platform, accessible via USSD codes on any handset.

Bill compliance, consumption, and payment. Payment compliance is uniformly high across both billing cycles as our sampling frame covers paying households only: 94.7% in December and 95.3% in January for the control group (Table 4.2, panel B). Neither treatment significantly affects whether the bill is paid at all in either cycle, suggesting that the interventions shifted the *mode* of payment rather than payment behavior per se. This is a reassuring null result: the treatments did not crowd out payment compliance. Furthermore, neither treatment significantly affects water

consumption or bill amounts in either billing cycle (Table 4.2, panel C). Point estimates are small and imprecisely estimated relative to control means.

Secondary Outcomes

Table 4.3 reports treatment effects on secondary outcomes measured in the endline survey.

Plans to adopt e-payment. Both treatments generate large and highly significant increases in the probability of planning to pay electronically next month. The information treatment raises stated plans by 19.3 pp ($p < 0.01$), while the incentive treatment raises plans by 31.4 pp ($p < 0.01$), relative to a control-group mean of 36.0%. The incentive effect is significantly larger than the information effect ($p = 0.017$), indicating that the financial incentive generates substantially more behavioral intent than information. By contrast, hypothetical willingness to pay electronically “if made easy” is unaffected by either treatment, with a control-group mean of 80.5%. The contrast between null effects on willingness and large effects on concrete plans suggests that the constraint is not attitudinal resistance but the gap between hypothetical openness and concrete behavioral intention.

Trust in e-payment. The two treatments move trust in opposite directions: the incentive increases trust by 0.165 points on the 5-point scale while information decreases it by 0.185 points. Neither point estimate is individually significant against the control group, but the difference between them is ($p = 0.011$). One interpretation is that the incentive arm’s hands-on experience with a successful transaction built confidence in the system, while the information arm which described the process without prompting a trial may have made salient the risks and complexity of digital payment without providing the reassurance of a completed transaction. The incentive group’s continued adoption in the post-treatment period suggests that the trust that was built through experience proved durable.

Perceived and actual time savings. The incentive significantly increases agreement that e-payment saves time (4.7% from control, $p < 0.05$), while information has no effect. This perception is consistent with the incentive’s effect on actual payment time: incentivized households report spending roughly 5 fewer minutes paying their bill (18% from control, $p < 0.05$). This mild response suggests that households may have been bundling their bill payment with other visits they needed to make. The information treatment shows no significant effect on either measure.

Mobile money accounts and Spillovers. Neither treatment significantly increases mobile-money account ownership, though the incentive coefficient (7.4 pp) is larger than the information coefficient (1.0 pp). The control group already has a 53.7% mobile-money penetration rate, suggesting that account access may not be the binding constraint for most households.

Cross-arm communication appears minimal. Only 4.6% of control households report hearing about e-payment training and 2.9% about incentives, with no significant treatment effects on ei-

ther indicator ($p > 0.5$). This suggests that the interventions did not diffuse meaningfully across experimental arms.

Heterogeneous Treatment Effects

We estimate treatment effects separately by four baseline characteristics: respondent gender, education (above vs. at or below secondary), asset holdings (above vs. below median), and respondent age (above vs. below median). These cuts are motivated by the expectation that digital payment adoption may vary with digital familiarity (proxied by education and assets), gender, and age, given documented differences in technology adoption across some of these dimensions in developing countries (e.g., Lee et al. (2022)). These analyses are exploratory: the study was not powered for subgroup comparisons, and we do not adjust for multiple testing.

The intervention had a two-dose structure. The first treatment cycle (December) corresponds to the initial offer, the second cycle (January) to a renewed offer, and February is a post-treatment cycle in which the incentive was no longer available. Tracking outcomes across all three cycles allows us to distinguish three distinct margins: (i) the immediate response to a first exposure, (ii) the marginal response to a second exposure, and (iii) persistence of behavior once the incentive is withdrawn. Figures 4.5, 4.6, 4.7, and 4.8 display the results. We organize the discussion by heterogeneity dimension, reporting digital adoption at all three doses before moving to secondary outcomes.

Gender. At the first dose (December) neither treatment is significant for males ($n = 219$), and the female ($n = 371$) response is driven entirely by the incentive (12.9 pp, $p = 0.003$). At the second dose (January), the male response sharpens dramatically: information generates an 18.3 pp increase in adoption ($p < 0.001$) and the incentive 22.2 pp ($p < 0.001$). Among females, the second-dose information effect remains smaller and insignificant (4.5 pp, $p = 0.33$), while the incentive effect is essentially unchanged from December (15.1 pp, $p = 0.002$). By February, both treatments are significant for both genders, with the female incentive effect catching up to the male level (22.9 pp vs. 22.2 pp).

Stated plans to pay electronically tell a complementary story. Among females, the incentive generates a 37.0 pp increase in plans ($p < 0.01$), the largest point estimate across all subgroups, compared to 23.5 pp among males ($p < 0.01$). The gap between February stated plans and February realized adoption is therefore wider for women than for men, suggesting that female respondents' intentions to switch were not fully converted into behavior even by the post-treatment cycle, possibly reflecting remaining practical barriers such as phone access or household decision-making dynamics.

The male subsample also shows a significant incentive effect on January water consumption (3.7 m³ or 42% of the control mean, $p = 0.012$) and January bill amounts (roughly \$1.92 or 102% of the control mean, $p = 0.009$), effects that are absent in the pooled sample and among women. These are second-dose effects; the corresponding December estimates are smaller and insignificant. One interpretation is that the reduced friction of digital payment, once habituated, led

male respondents to higher consumption; another is compositional, if the incentive differentially retained higher-consuming male households in paid status.

Education. At the first dose, the incentive coefficient is 14.1 pp for above-secondary respondents ($p = 0.015$) and a smaller 8.7 pp for at-or-below-secondary ($p = 0.032$). At the second dose, the two subgroups converge (16.1 pp and 15.3 pp respectively, both significant). By February the above-secondary group has pulled ahead (27.8 pp, $p < 0.001$ vs. 20.0 pp, $p < 0.001$), a large post-treatment incentive effect.

Among at-or-below-secondary respondents, the incentive is associated with a 6.5 pp reduction in February bill payment compliance ($p = 0.014$), which corresponds to roughly a 6% decline against the high baseline compliance rate. We note that this coefficient appears only in this single subgroup-cycle combination, that compliance remains very high in absolute terms, and that the analysis is not adjusted for multiple testing. We do not draw strong conclusions from it, but it warrants being mentioned.

Mobile money account ownership tracks the adoption pattern. In the pooled sample, neither treatment significantly increases account holding. The split, however, isolates the response. The incentive generates a 17.5 pp increase in account holding among at-or-below-secondary respondents ($p = 0.010$), a group with substantially lower baseline mobile money penetration. This indicates that the incentive operated on two margins simultaneously: switching payment mode among existing account holders, and prompting new account creation among less digitally connected households.

Assets. The asset split also shows clear dose dependence, but in the opposite direction from education. The incentive effect on adoption is larger for below-median-asset households at the first dose (13.6 pp, $p = 0.003$ vs. 10.5 pp, $p = 0.028$), at the second dose (20.2 pp, $p < 0.001$ vs. 11.7 pp, $p = 0.058$), and converges further in the post-treatment cycle (23.4 pp vs. 20.0 pp, both $p < 0.001$). The information treatment shows the same pattern: significant for below-median-asset respondents at every cycle, but not for the above-median group.

As with education, the incentive generates a significant increase in mobile money account ownership among below-median-asset respondents (17.2 pp, $p = 0.019$), with no comparable effect among the above-median-asset group. Together, the education and asset results suggest that the financial incentive overcame infrastructural rather than purely behavioral barriers among less digitally connected households.

Age. The age split reveals differential adoption *timing* rather than differential responsiveness, and it is most clearly visible across the three doses. At the first dose, below-median-age respondents are more responsive: the incentive increases adoption by 15.5 pp ($p < 0.001$) among younger respondents but has no significant effect among older ones (7.1 pp, $p = 0.18$). At the second dose, the pattern reverses sharply: older respondents now show the larger effect (25.2 pp, $p < 0.001$ vs. 10.6 pp, $p = 0.042$ for younger), and the same reversal appears in the information coefficient

(14.4 pp, $p = 0.002$ vs. 4.2 pp, $p = 0.42$). By February the two groups have converged (incentive effects of 21.7 pp and 22.6 pp, both $p < 0.001$).

Stated plans corroborate this timing interpretation. The information treatment raises plans by 30.8 pp among younger respondents ($p < 0.01$) but only 9.7 pp among older ones ($p > 0.10$). Younger respondents' intentions respond easily to information even when their first-dose behavior already reflected it, while older respondents' plans and behavior move together only after the second dose. The most natural interpretation is that younger, more digitally familiar respondents adopted on the first offer while older respondents required reinforcement from a second offer before acting; both groups ultimately formed digital payment habits, but through different adoption trajectories.

Summary. Five patterns emerge from the heterogeneity analysis. First, the incentive generates significant adoption effects across nearly every subgroup at both doses, while the information treatment is overall less effective. Second, incentive effects on adoption are typically *larger* in the post-treatment cycle than during either treatment cycle, consistent with delayed adoption and habit formation. Third, the financial incentive prompted new mobile money account creation among less digitally connected households (lower education, fewer assets), suggesting that it overcame infrastructural as well as behavioral barriers. Fourth, the gap between stated plans and realized behavior is wider for women than for men, pointing to remaining practical constraints. Fifth, the age dimension shows differential timing rather than differential responsiveness, with younger respondents adopting at the first dose and older respondents only at the second dose, before both groups converge in the post-treatment cycle.

4.7 Mechanisms

The incentive generates two empirical patterns: a treatment effect that is roughly twice as large as the information effect during the offer window, and a treatment effect that grows rather than decays in the post-incentive billing cycle. We argue that both follow naturally from a simple fixed-cost-of-adoption model: digital payment requires a one-time investment (learning the interface, registering an account, completing a first transaction) after which the per-transaction cost is lower than in-person payment; the incentive lowers the effective price of crossing that fixed cost; and once crossed, households stay on the digital channel because doing so is now the cheaper option.

Three pieces of evidence support this interpretation against the alternatives. First, although digital payment is sticky once adopted, this stickiness is not what differentiates the treatment arms in February. Table 4.4 decomposes February adoption into persistence among January adopters and new adoption among January non-adopters. Persistence rates are high in all three arms (60.0% in control, 70.6% in information, and 71.1% in incentive), confirming that the technology itself is sticky for households who try it once. Persistence rates are nearly identical between the two treatment arms (70.6% vs. 71.1%), so much of the February gap between information and incentive must come from differential conversion of January non-adopters rather than from differential stickiness once adopted.

The incentive arm converted January non-adopters into February adopters at 22.0%, compared with 15.6% in information and 7.5% in control. This is the channel through which the February treatment effect grows. We interpret this pattern as an *intention-action gap*: the incentive raised the rate at which households were willing to pay the upfront fixed cost, but a non-trivial share did not act on that intention until after the offer had been withdrawn. Habit formation explains why first-cycle adopters in all three arms continue paying digitally, but it does not differentiate the treatment arms; the incentive's larger February effect comes from the fact that it produced more first-cycle adopters *and* continued to convert non-adopters even after the financial offer ended.

Second, we observe *active learning across channels*: households consolidate onto the lowest-cost option once they have tried the technology. After the first incentive offer, new adopters split roughly evenly between telebirr and the CBE smartphone app. By the second billing cycle, telebirr had become the dominant channel (13 pp treatment effect, $p < 0.001$) while the CBE app effect became insignificant (Figure 4.4). Telebirr is accessible via a USSD code on any mobile phone, while the CBE app requires a smartphone and bank-app registration. The migration toward telebirr across cycles is consistent with households learning that a lower-cost digital channel exists and switching to it once they have paid the fixed cost of trying digital payment at all. The first incentive financed experimentation; the second consolidated households on the channel with the lowest ongoing transaction cost.

Third, survey evidence on self-reported barriers (Figure 4.9) confirms that the treatments operated on different components of the bundled fixed cost. The estimates reveal an interesting divergence: the information treatment successfully targets the primary cognitive constraint, generating a significant reduction in respondents citing a lack of knowledge (“do not know”) as a barrier. The incentive treatment, however, does not reduce this knowledge gap; in fact, it significantly increases the likelihood of respondents citing the system as “too complicated” and having “network issues.” This suggests that inducing actual trial exposes users to real system frictions that information alone glosses over. Instead, the incentive's primary attitudinal impact is a sharp reduction in the psychological attachment to physical receipts. The data indicates that while information alleviates self-reported lack of know-how, the financial incentive is required to break the behavioral habit of needing physical proof of payment, pushing users across the threshold even as it forces them to confront the platform's complexities.

A direct alternative is that the incentive worked because it provided a salient, time-limited reason to transact, with the dollar amount itself doing little work. We cannot fully separate this from the fixed-cost story with the present design, but two observations are difficult to reconcile with a pure salience account. The incentive reduced in-person bank-branch payments meaningfully (CBE branch -9.4 pp, Awash Teller -9.4 pp in the second cycle), suggesting that adopters substituted away from in-person rather than treating digital payment as a novelty conducted alongside their existing routine. And the post-incentive adoption growth, with no remaining offer or framing in place, indicates that what changed during the treatment was something durable about households' assessment of digital payment such that they had now done it and learned that the per-transaction cost going forward is lower than continuing in-person.

What the data establish is that a temporary, modest financial incentive moved households across a fixed adoption threshold, that crossing the threshold was sufficient to lock in continued digital use

without further intervention, and that the consolidation onto the lowest-cost digital channel reflects ongoing optimization rather than a one-time response to the incentive itself. The implication for policy design is that the returns to interventions that successfully induce a first digital transaction are larger than treatment-period adoption alone would suggest.

4.8 Expert Predictions

We also collected data from 44 experts through the Social Science Prediction Platform (SSPP). Each forecaster was given the experimental design, the sample-selection criteria, the treatment dosages, and was asked to predict treatment effects for both arms on the primary digital-adoption outcome across all three billing cycles, as well as on a small set of secondary outcomes. Forecasts were elicited as point estimates, and we report below the mean across forecasters together with the gap between the forecast and the realized average treatment effect (ATE). This exercise serves two purposes: it provides a benchmark against which to gauge whether our findings are surprising relative to expert priors, and it documents which features of the response were anticipated and which were not. Figure 4.10 plots, for each outcome and arm, the gap between the mean forecast and the realized ATE together with a 95% confidence interval on that difference.⁷

For each outcome and treatment arm, we compare the mean forecast $\bar{\beta}$ to the realized ATE $\hat{\beta}$ using a two-sample Wald test, $z = (\bar{\beta} - \hat{\beta}) / \sqrt{\widehat{SE}(\bar{\beta})^2 + \widehat{SE}(\hat{\beta})^2}$, where $\widehat{SE}(\bar{\beta})$ is obtained from a 1,000-replication nonparametric bootstrap over forecasters and $\widehat{SE}(\hat{\beta})$ is the heteroskedasticity-robust standard error from the ITT regression in equation (4.1). Independence of the two samples justifies adding the variances, and we report two-sided p -values from the standard normal distribution.

For digital adoption, forecasters were broadly accurate about the direction and rough magnitude of the information treatment but systematically misjudged the trajectory of the incentive arm. The information treatment's realized December effect (6.8 pp) is close to the forecast (9.4 pp, $p = 0.407$), as are the January (realized 9.3 pp vs. forecast 11.0 pp, $p = 0.635$) and February (realized 13.3 pp vs. forecast 10.9 pp, $p = 0.575$) comparisons. None of these differences is statistically distinguishable from zero, indicating that experts anticipated the information arm's trajectory reasonably well across all three billing cycles.

The incentive arm, by contrast, exhibits a pattern of mispredicted timing. In December, forecasters expected an effect of 22.0 pp, nearly double the realized 11.6 pp (difference 10.3 pp, $p = 0.009$): expert priors seem to have implicitly assumed that a 50% incentive would generate large immediate take-up, whereas the actual response was more gradual. By January, the forecast and realized effects had partially converged (23.4 vs. 16.7 pp, $p = 0.133$), and by February the pattern reversed entirely: forecasters expected the incentive effect to decay once the incentive was withdrawn (predicted 17.6 pp), yet the realized effect grew to 22.5 pp (difference -5.0 pp, $p = 0.306$). The expert panel thus encoded a standard price-incentive mental model in which incentives act fast and fade fast, and the data depart from that model in both directions: slower

⁷Table C.4 in the appendix shows the exact estimates as well.

onset and durable persistence. This is consistent with the intention-action gap and active-learning mechanisms discussed in Section 4.7, neither of which appears to have featured prominently in forecasters' priors.

The secondary outcomes reveal additional dimensions on which expert priors diverged from realized effects. Forecasters were close to the ATE on stated plans to adopt e-payment under the information treatment (forecast 18.3 pp, realized 19.3 pp, $p = 0.856$) but underestimated the incentive arm's effect on plans (forecast 22.1 pp, realized 31.4 pp, $p = 0.084$), a pattern broadly consistent with the realized adoption dynamics: the incentive was a more powerful intention-shifter than experts anticipated. Forecasts on mobile-money account ownership were directionally correct but somewhat optimistic, particularly for the information arm: experts expected a 7.3 pp increase, but the realized effect was essentially zero (1.0 pp, $p = 0.251$), with the incentive arm's realized effect (7.4 pp) close to but below the mean forecast (10.9 pp, $p = 0.524$). Finally, forecasters substantially overestimated time savings: the predicted reductions of roughly 11 and 10 minutes for the information and incentive arms exceeded the realized effects of 1.3 ($p = 0.002$) and 5.0 minutes ($p = 0.069$), respectively.⁸

Two takeaways emerge from the forecasting exercise. First, experts correctly anticipated the direction and rough magnitude of the information treatment's effect on adoption but systematically mistimed the incentive arm's trajectory, expecting too much in the first month of treatment and too little after the incentive was withdrawn. Second, the gap between forecast and realized effects is largest precisely on the dimensions where the mechanism we propose in Section 4.7 (intention-action gap reinforced by active learning across channels) departs from a standard price-incentive narrative, suggesting that the persistence finding represents a genuine update on prior beliefs among the experts in our sample, rather than a confirmation of widely held expectations.

4.9 Cost Effectiveness

We conduct a back-of-envelope cost-effectiveness analysis to compare the two interventions on a common metric: cost per additional household induced to adopt digital payment. We decompose costs into delivery costs and, for the incentive arm, the direct transfer value. Both arms required an enumerator physical visit in the first billing cycle (December); they differ in how the second-cycle reminder was delivered and whether a monetary transfer was offered.

The information treatment was delivered by trained enumerators who visited households in person, provided step-by-step instructions, and left printed materials. Enumerator compensation was \$18.54 per day, with an output of three household visits, implying a delivery cost of \$6.18 per household per visit. The intervention was re-delivered in person in the second billing cycle (January), so the total delivery cost is \$12.36 per household. There was no monetary transfer; therefore, the \$6.18 figure represents the total cost per visit.

The incentive arm also received a physical enumerator visit in the first cycle (December), at the same per-visit cost of \$6.18. The second-cycle reminder, however, was delivered by phone rather

⁸We elicited forecasts for treatment effects on trust in e-payment but exclude this outcome from the comparison as several forecasters interpreted the slider as a posterior level rather than a change.

than in person: enumerators were paid \$18.54 per day and contacted 25 households, implying an enumerator time cost of \$0.74 per household. Combined with an average call cost of \$0.03 per household, we get a total second-round delivery cost of \$0.77 and a total delivery cost of \$6.95 per household.⁹ Figure 4.11 plots the distribution of the incentive cost among households that paid digitally.

We compute cost per additional adopter in two ways. The first divides per-household cost by the January ITT adoption effect (after both intervention offers), capturing the cost of inducing adoption during the treatment window. The second uses the February ITT effect (the first post-treatment billing cycle). Because digital payment is nearly an absorbing state in the AAWSA customer universe (96.6% of digital payers in March 2025 remained digital ten months later), the February estimate is arguably the more policy-relevant figure.

The results, reported in Table 4.5, reveal a clear asymmetry. The information arm costs \$12.36 per household and induced adoption among 9.3% of targeted households during the treatment period, implying a cost of \$133.1 per induced adopter. The incentive arm costs \$7.57 per household and induced adoption among 16.7%, implying a cost of \$45.3 per adopter, less than half the cost of the information arm. In the post-treatment period, cost per durable adopter falls to \$93.1 (information) and \$33.6 (incentive). The incentive is therefore both more effective in absolute terms and substantially more cost-effective per household converted. These estimates are conservative: they ignore downstream savings to AAWSA from reduced manual reconciliation and branch congestion, as well as the time and transport-cost savings that accrue to households upon switching (the incentive arm reports spending approximately five fewer minutes paying their bill at endline).

4.10 Conclusion

This paper provides experimental evidence on what moves persistent in-person payers onto digital channels for a recurring government service. Among almost 600 households in Addis Ababa that had paid their water bills exclusively at bank branches for at least six consecutive months, a 50% incentive on electronic payment offered for two consecutive billing cycles increased digital adoption by 11.6 pp after the first offer and 16.7 pp after the second, nearly tripling the control-group rate. In the first post-treatment billing cycle the effect grew to 22.5 pp ($p < 0.01$). An information-only treatment also delivered twice increased adoption by 6.8, 9.3, and 13.3 pp across the same cycles, but consistently produced about half the incentive's effect. Neither intervention reduced bill payment compliance. The treatments shifted how households paid, not whether they paid.

We draw three policy implications. First, the incentive effect was driven largely by uptake of *telebirr*, a USSD-based mobile money platform accessible on basic handsets. The binding constraint for most non-adopters was not access or capability but the upfront cost of changing an established routine, which a temporary financial incentive was sufficient to overcome. Second,

⁹Since the transfer was conditional on actual digital payment, 136 of the 198 incentive-arm households never redeemed the offer and received no transfer; among the 62 households that redeemed the offer in at least one cycle, the average cumulative transfer was approximately \$1.97.

administrative data from AAWSA's full customer base show that digital payment is nearly an absorbing state: 96.6% of customers who paid digitally in March 2025 continued to do so ten months later. If this persistence extends to experimentally induced adopters beyond the one-month post-treatment we observe, then even a one-time temporary incentive generates lasting reductions in transaction costs for both households and the utility. Third, the incentive delivered larger and more durable adoption gains at substantially lower per-household cost: roughly \$33.6 per durable adopter versus \$93.1 for information, before accounting for downstream administrative savings to the utility from reduced in-person reconciliation.

We turn to discuss the study's limitations. The sample is drawn from a single large city and targets a specific population of persistent non-adopters, limiting external validity to other settings and populations. The relatively small sample also leaves the study underpowered for some secondary outcomes and subgroup analyses. Furthermore, we observe one post-intervention billing cycle showing that treatment effects persisted and grew after incentive withdrawal; whether this persistence extends over longer horizons remains an open question that we plan to address by tracking these households over additional billing cycles. Finally, the results do not speak to the non-paying segment of AAWSA's customer base because our sampling frame excluded them by construction.

Overall, digital infrastructure for citizen-state transactions is now widely available, but its operational benefits are realized only when citizens use it. For habitual in-person users, a temporary, modest incentive on the digital channel appears sufficient to move adoption permanently, at a per-adopter cost low enough that even rough estimates of the administrative savings to the utility likely exceed it. The returns to interventions that successfully induce a first digital transaction may therefore be substantially larger than a static analysis of the incentive period alone would suggest. More broadly, the results suggest that the bundled fixed cost of switching to a digital channel for a recurring transaction has both cognitive and financial components, and that a temporary financial incentive does work that information alone cannot, particularly when the post-incentive response is allowed to play out beyond the offer window.

4.11 Figures and Tables

Table (4.1) Transition of Payment Modes

March Payment Mode	January Payment Mode	
	In-Person	Digital
In-Person (March)	6204 (31.4%)	13580 (68.6%)
Digital (March)	7638 (3.4%)	218735 (96.6%)

Notes: This table reports payment mode transitions between March 2025 and January 2026 among residential customers with a cleanly classified channel in both months. Percentages represent row shares. $N = 246,157$.

Table (4.2) Primary Effects (Admin Data)

Outcome	Control mean	Information (T1)	Incentive (T2)	T1=T2 p-value	N
<i>Panel A: Digital payment adoption</i>					
Paid digitally (Dec)	0.068	0.068** (0.030)	0.116*** (0.033)	0.189	590
Paid digitally (Jan)	0.089	0.093*** (0.034)	0.167*** (0.038)	0.075	590
Paid digitally (Feb)	0.119	0.133*** (0.040)	0.225*** (0.044)	0.056	541
<i>Panel B: Bill payment compliance</i>					
Paid bill (Dec)	0.947	0.017 (0.021)	0.007 (0.022)	0.607	590
Paid bill (Jan)	0.953	-0.003 (0.022)	-0.023 (0.024)	0.389	590
Paid bill (Feb)	0.926	0.005 (0.020)	-0.029 (0.024)	0.114	590
<i>Panel C: Consumption and billing</i>					
Water consumption (Dec)	9.078	0.533 (0.751)	-0.021 (0.696)	0.451	552
Water consumption (Jan)	9.394	0.223 (0.772)	0.775 (0.789)	0.490	557
Bill amount (Dec, USD)	1.932	0.315 (0.338)	0.196 (0.333)	0.742	564
Bill amount (Jan, USD)	2.314	0.037 (0.346)	0.382 (0.365)	0.363	557

Notes: OLS estimates of intent-to-treat effects. Column (1) reports the control group mean. Columns (2) and (3) report coefficients on the Information and Incentive treatment indicators with robust standard errors in parentheses. All regressions include randomization strata fixed effects. Column (4) reports the p -value from a test of equality of the two treatment effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table (4.3) Secondary Effects (Survey Data)

Outcome	Control mean	Information (T1)	Incentive (T2)	T1= T2 p-value	N
Plans to e-pay next month	0.360	0.193*** (0.052)	0.314*** (0.050)	0.017	540
Willing to e-pay if made easy	0.805	-0.032 (0.045)	0.014 (0.043)	0.295	486
Trust in e-payment (5=high)	3.760	-0.185 (0.142)	0.165 (0.136)	0.011	540
E-payment saves time (5=high)	4.371	-0.003 (0.094)	0.206** (0.084)	0.013	540
Confident with mobile (5=high)	3.577	0.131 (0.136)	0.128 (0.133)	0.985	540
Has mobile money account	0.537	0.010 (0.053)	0.074 (0.053)	0.218	540
Transport cost (USD)	0.024	0.009 (0.009)	-0.007 (0.007)	0.070	490
Time to pay in person (min)	27.030	-1.335 (2.428)	-4.968** (2.336)	0.061	490
Heard about e-pay training	0.046	0.012 (0.023)	-0.003 (0.023)	0.527	540
Heard about e-pay incentive	0.029	-0.007 (0.016)	-0.018 (0.016)	0.407	540

Notes: OLS estimates of intent-to-treat effects. Column (1) reports the control group mean. Columns (2) and (3) report coefficients on the Information and Incentive treatment indicators with robust standard errors in parentheses. All regressions include randomization strata fixed effects. Column (4) reports the p -value from a test of equality of the two treatment effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table (4.4) Payment Transitions from January to February

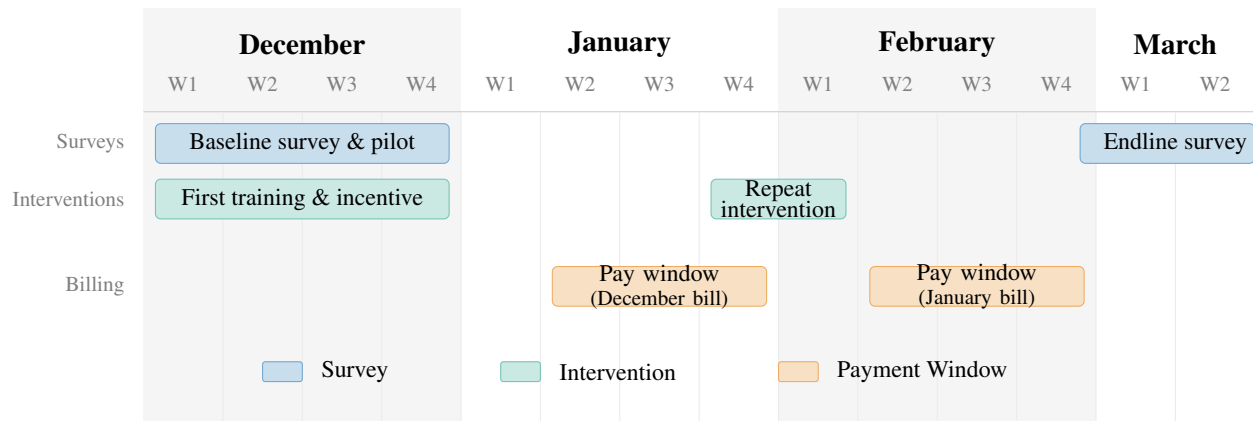
January Payment Mode	In-Person (Feb)		Digital (Feb)	
	N	%	N	%
<i>Panel A: Control arm</i>				
In-person (Jan)	149	92.5%	12	7.5%
Digital (Jan)	6	40.0%	9	60.0%
<i>Panel B: Information arm</i>				
In-person (Jan)	130	84.4%	24	15.6%
Digital (Jan)	10	29.4%	24	70.6%
<i>Panel C: Incentive arm</i>				
In-person (Jan)	103	78.0%	29	22.0%
Digital (Jan)	13	28.9%	32	71.1%

Notes: Row shares are computed among households with non-missing February payment mode.

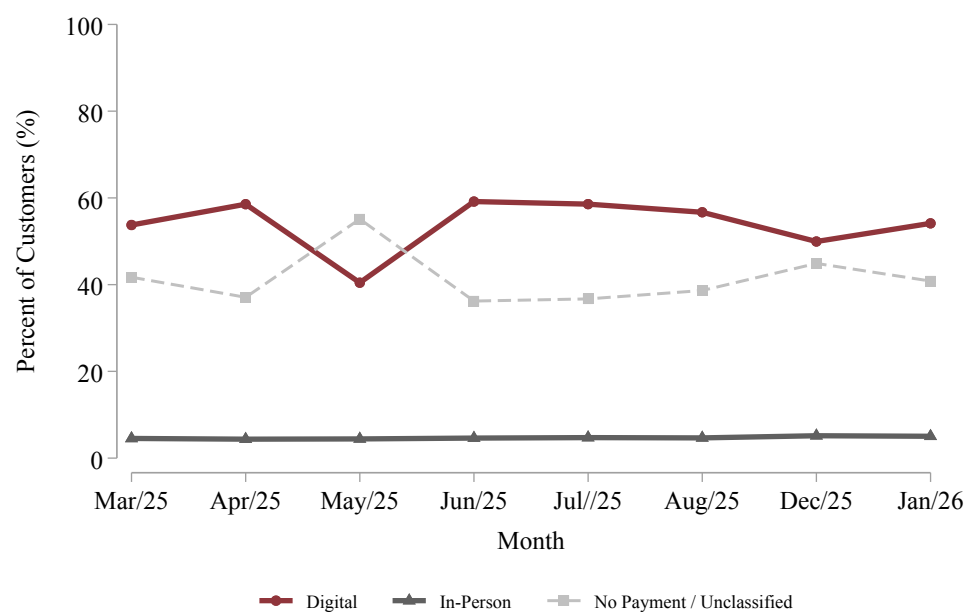
Table (4.5) Cost Effectiveness Estimates

	Information	Incentive
<i>Panel A: Costs per household targeted</i>		
Round 1 in-person visit	\$6.18	\$6.18
Round 2 in-person visit	\$6.18	—
Round 2 phone (enumerator time)	—	\$0.74
Round 2 call cost	—	\$0.03
Direct transfer (conditional)	—	\$1.97
Total cost per HH	\$12.36	\$7.57
<i>Panel B: Treatment-period adopters (January ITT)</i>		
Adoption effect (pp)	9.3	16.7
Cost per induced adopter	\$133.1	\$45.3
<i>Panel C: Post-treatment durable adopters (February ITT)</i>		
Adoption effect (pp)	13.3	22.5
Cost per durable adopter	\$93.1	\$33.6

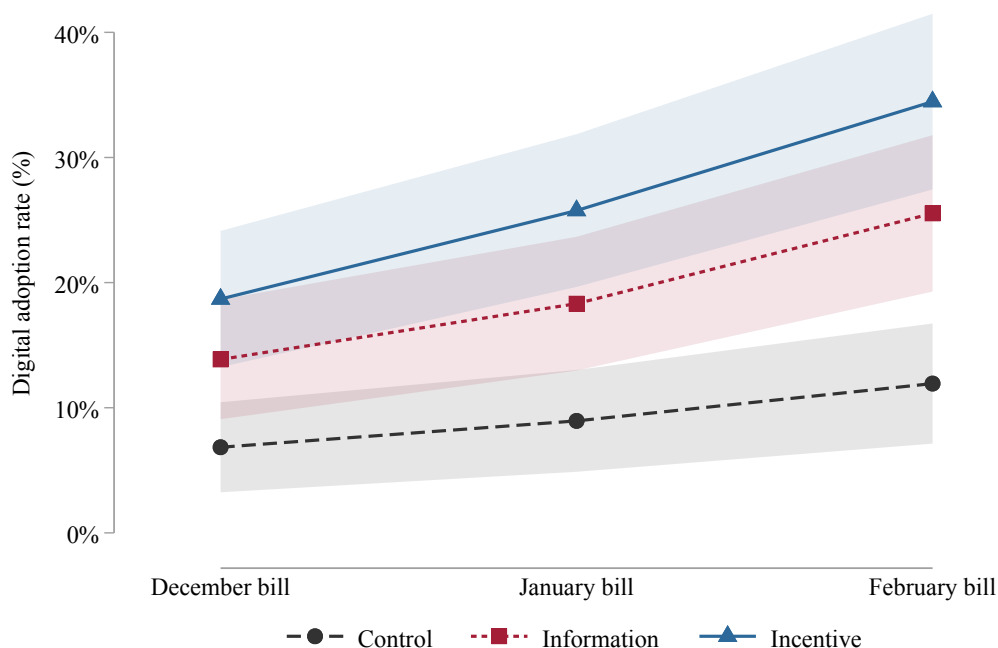
Notes: Enumerator wage: \$18.54/day throughout. In-person visits: 3 households/day (\$6.18/visit); both arms received a Round 1 visit. The information arm received a second in-person visit in Round 2. The incentive arm's Round 2 reminder was by phone: enumerator time cost \$18.54/25 calls = \$0.74 plus \$0.03 call cost, totalling \$0.77 in Round 2 delivery and \$6.95 overall. The direct transfer equals 50% of the December bill (mean \$2.37) and 50% of the January bill (mean \$2.73), winsorized at the 99th percentile and converted at 151 ETB/USD, conditional on digital payment in each cycle; 136 of 198 incentive-arm households never redeemed and received no transfer. Cost per adopter divides total per-household cost by the ITT adoption effect in percentage points. Panel B uses the January ITT; Panel C uses the February ITT, the first post-treatment cycle.

Figure (4.1) Experimental Timeline

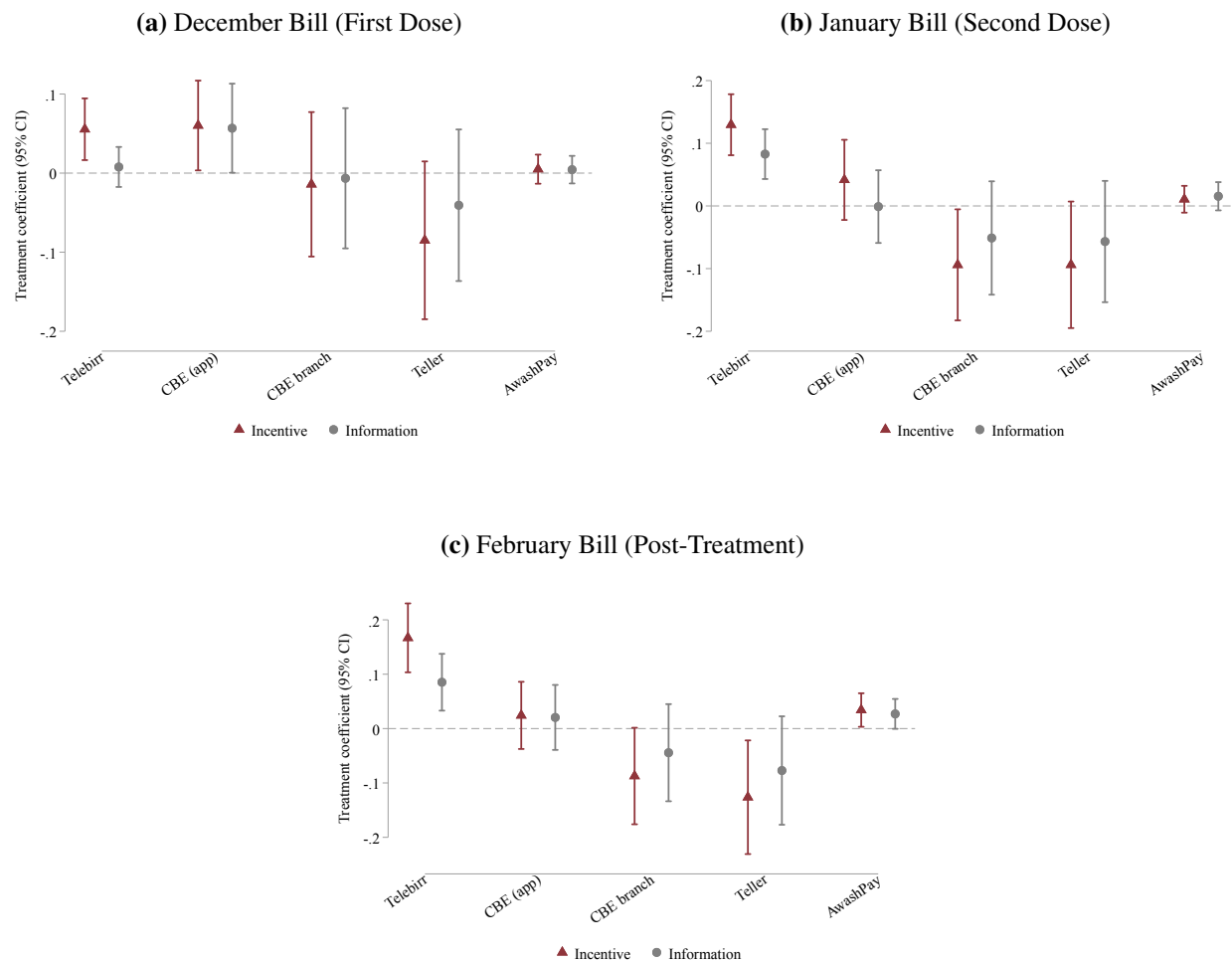
Notes: This figure shows the study timeline.

Figure (4.2) Payment Trends

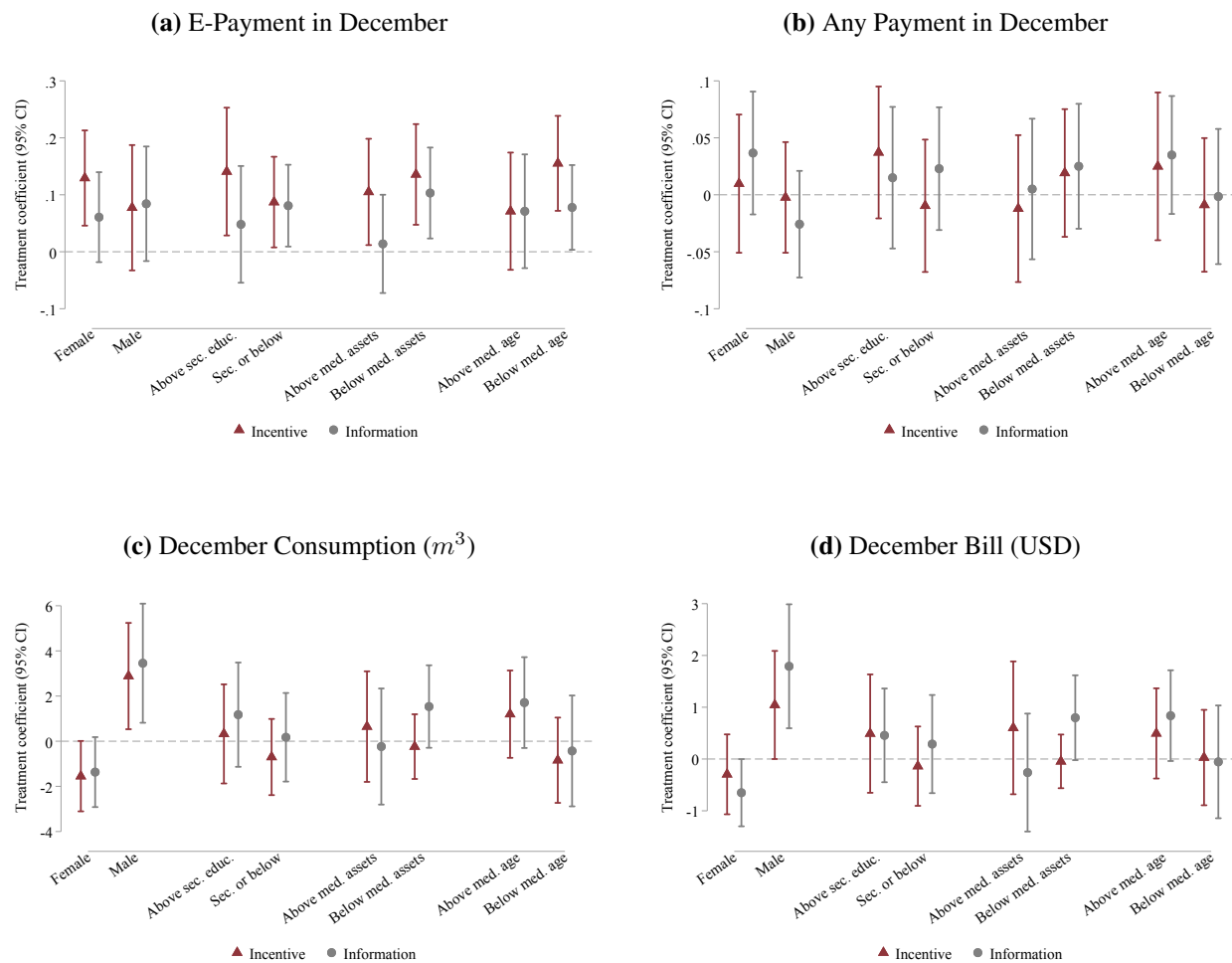
Notes: This figure decomposes the payment behavior of the full residential customer universe (approximately 553,000 accounts) across available administrative data months. The solid line plots the share paying via a digital channel (CBE, CBE Birr, CBE Internet, or Awash Bank). The dark line near the bottom plots the share paying in person at a bank front desk (Teller at Awash Bank or CBE BR at CBE). The dashed line captures the residual: customers who did not pay or whose channel field is missing.

Figure (4.3) Digital Adoption Rates by Arm

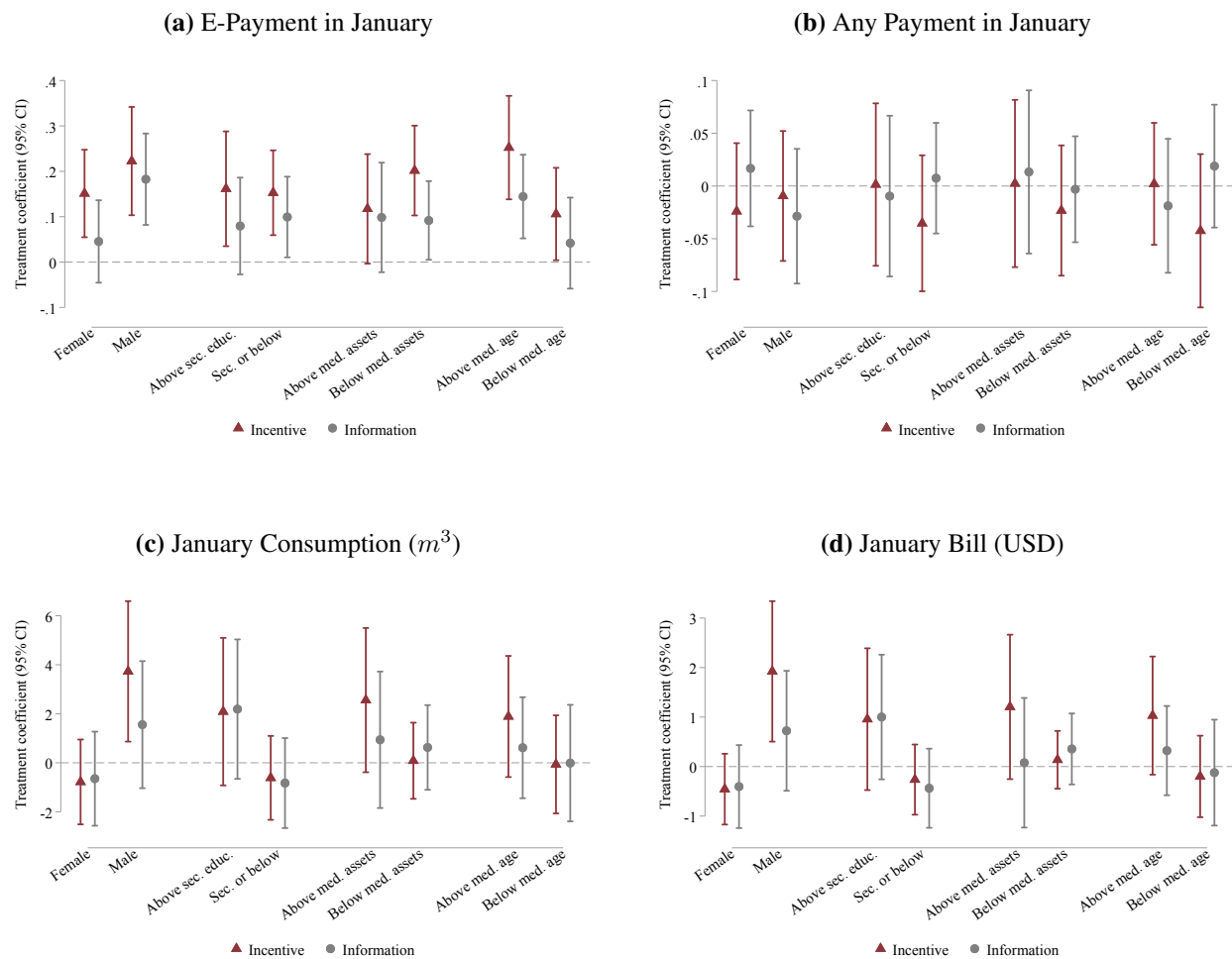
Notes: This figure plots the share of households paying their water bill digitally, by treatment arm and billing cycle. Shaded bands represent 95% confidence intervals from the raw data. Note that the actual payment date for each bill is the subsequent month, so the payment for February was made in March, roughly one month after the second-time interventions took place.

Figure (4.4) Treatment Effect by Payment Channel

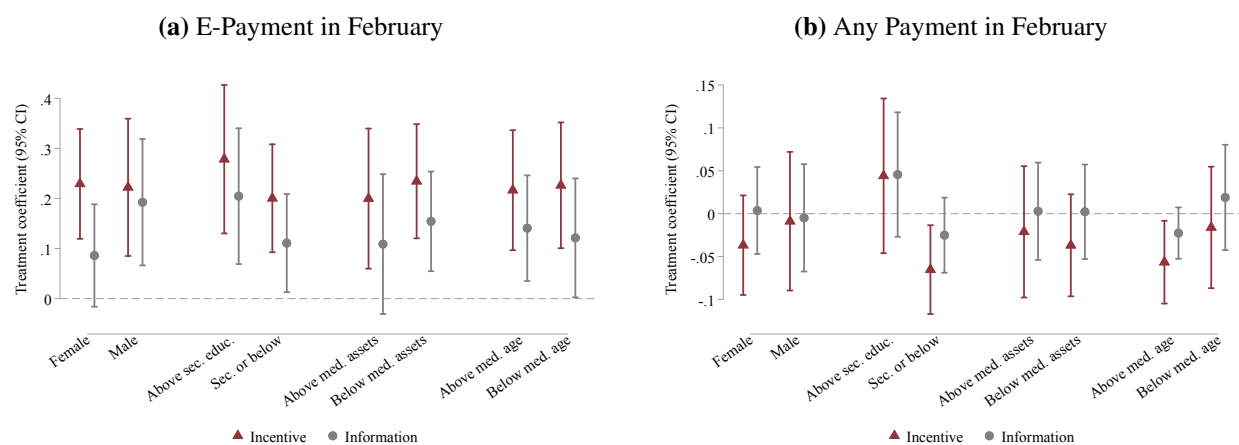
Notes: Each panel plots intent-to-treat coefficients from separate OLS regressions of each channel indicator on the two treatment dummies, with strata fixed effects and robust standard errors. telebirr, CBE (app), and AwashPay are digital channels; CBE branch (in-person payment at a CBE front desk) and Teller (in-person payment at an Awash Bank front desk) are in-person channels. Error bars show 95% confidence intervals. Panel (a) captures the effect of the first incentive offer alone (December bill, paid in January); Panel (b) captures the cumulative effect after the incentive was re-offered the following month (January bill, paid in February); Panel (c) captures the effect after both incentives have ended (February bill, paid in March).

Figure (4.5) Heterogeneous Effects on Primary Outcomes (First Dose)

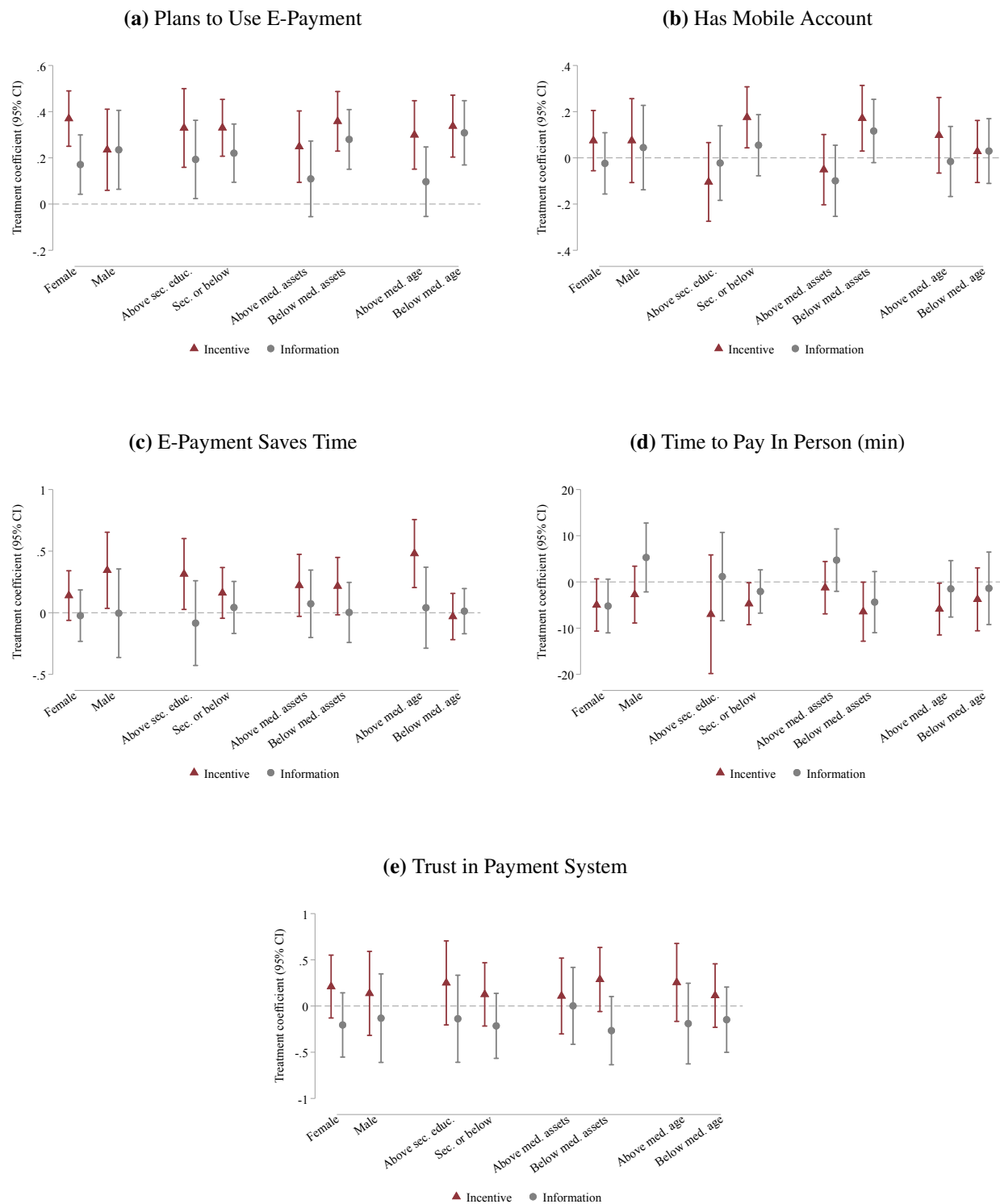
Notes: Each panel plots the incentive (maroon triangles) and information (gray circles) treatment coefficients with 95% confidence intervals, estimated separately by subgroup. December outcomes capture the effect of the first incentive offer only, before the January reminder. All regressions include strata fixed effects and use robust standard errors.

Figure (4.6) Heterogeneous Effects on Primary Outcomes (Second Dose)

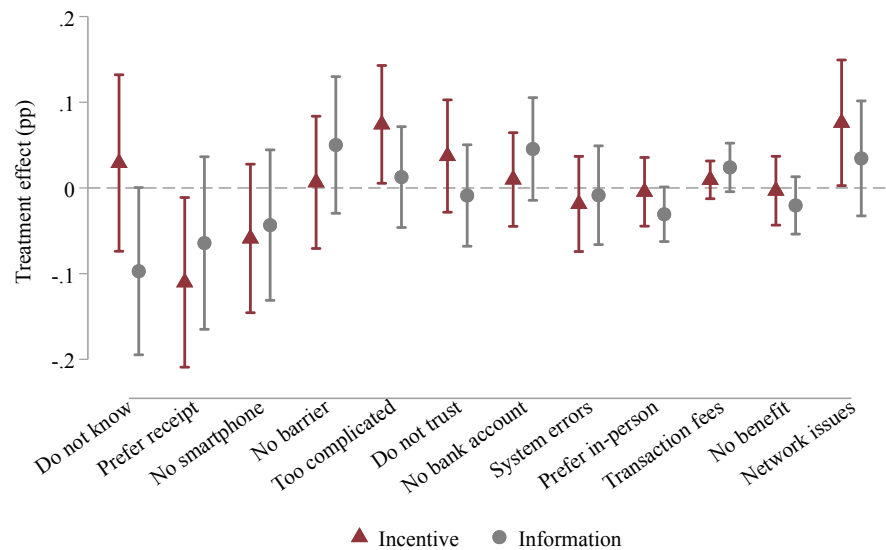
Notes: Each panel plots the incentive (maroon triangles) and information (gray circles) treatment coefficients with 95% confidence intervals, estimated separately by subgroup. January outcomes capture the cumulative effect after two incentive offers (December and January reminder). All regressions include strata fixed effects and use robust standard errors.

Figure (4.7) Heterogeneous Effects on Primary Outcomes (Post-Treatment)

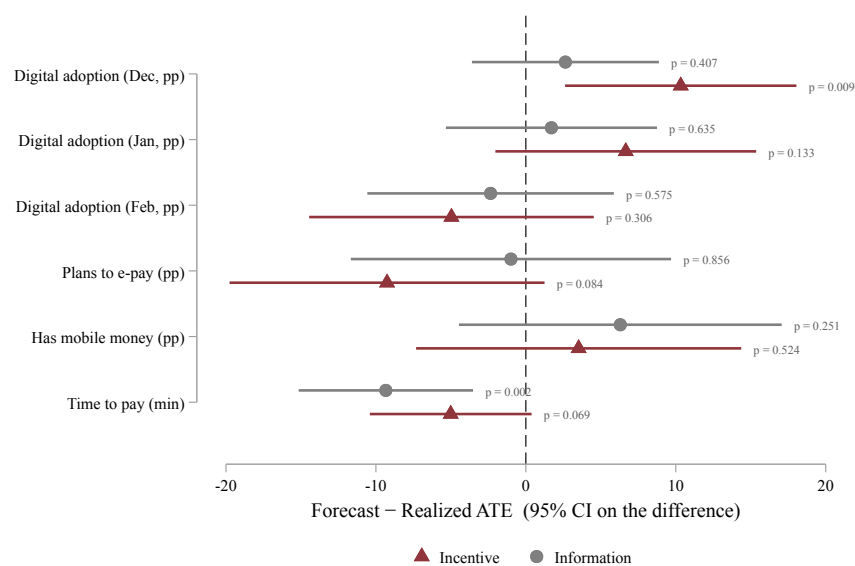
Notes: Each panel plots the incentive (maroon triangles) and information (gray circles) treatment coefficients with 95% confidence intervals, estimated separately by subgroup. February outcomes capture the first billing cycle after the incentive period ended (February bill, paid in March). Consumption and bill amount panels are omitted as February administrative data do not include these variables. All regressions include strata fixed effects and use robust standard errors.

Figure (4.8) Heterogeneous Effects on Secondary Outcomes

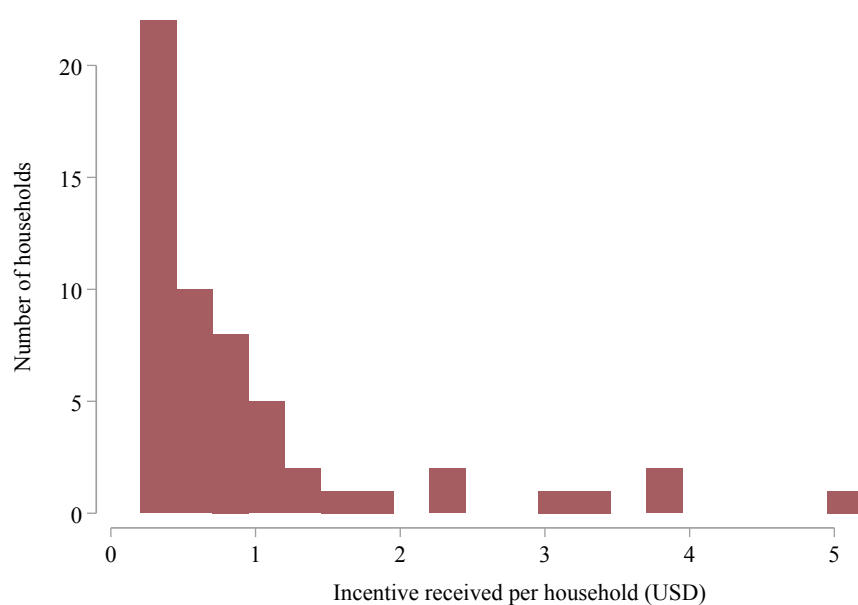
Notes: Each panel plots the incentive (maroon triangles) and information (gray circles) treatment coefficients with 95% confidence intervals, estimated separately by subgroup. All regressions include strata fixed effects and use robust standard errors.

Figure (4.9) Treatment Effects on Barriers to Digital Payment

Notes: This figure plots the intent-to-treat (ITT) estimates of the information (gray circles) and incentive (maroon triangles) treatments on the likelihood of citing specific barriers to digital payment at endline. Coefficients are estimated using ordinary least squares (OLS) regressions with strata fixed effects and robust standard errors. Error bars represent 95% confidence intervals.

Figure (4.10) Forecasts vs. Realized Treatment Effects

Notes: This figure shows differences between average forecasts and realized ATE for the primary outcome and three secondary outcomes. Horizontal lines are 95% confidence intervals.

Figure (4.11) Distribution of Incentive Transfers Among Incentive Arm Redeemers

Notes: Digital payers only ($N = 62$; 136 in-person payers excluded). There are 6 high-cost outliers above the 90th percentile that are not shown. Transfer equals 50% of the water bill in each billing cycle where the household paid digitally.

Bibliography

- Abajian, A., C. Cole, B. K. Jack, K. C. Meng, and M. Visser (2025). Dodging day zero: Drought, adaptation, and inequality in Cape Town. *Journal of the European Economic Association* 0(0), 1–34.
- Abramitzky, R. and L. Boustan (2024). Immigrants and their children assimilate into us society and the us economy, both in the past and today. *PNAS Nexus*.
- Aggarwal, S., V. Brailovskaya, and J. Robinson (2020). Mobile money and the path to financial inclusion: Evidence from a field experiment in Malawi. *Working Paper*.
- Agness, D. (2023). Housing and human capital: Condominiums in Ethiopia. *3ie Registry*.
- Aker, J. C., R. Boumnijel, A. McClelland, and N. Tierney (2016). Payment mechanisms and anti-poverty programs: Evidence from a mobile money cash transfer experiment in niger. *Economic Development and Cultural Change* 65(1), 1–37.
- Al-Rkebat, R. A. (2026). Strategic resilience: A comprehensive analysis of Jordan’s water security, digital transformation, and infrastructure modernization. *Journal of Water Resource and Protection* 18(1), 70–83.
- Alesina, A., A. Miano, and S. Stantcheva (2023). Immigration and redistribution. *The Review of Economic Studies* 90(1), 1–39.
- Allcott, H., B. B. Lockwood, and D. Taubinsky (2019). Regressive sin taxes, with an application to the optimal soda tax. *The Quarterly Journal of Economics* 134(3), 1557–1626.
- Alrababa’h, A., A. Dillon, S. Williamson, J. Hainmueller, D. Hangartner, and J. Weinstein (2021). Attitudes toward migrants in a highly impacted economy: Evidence from the Syrian refugee crisis in Jordan. *Comparative Political Studies* 54(1), 33–76.
- Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and early training projects. *Journal of the American Statistical Association* 103(484), 1481–1495.
- Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 91(434), 444–455.

- Ash, N. and C. Huang (2018). Using the compact model to support host states and refugee self-reliance. *World Refugee Council Research Paper Series*.
- Auriol, E. and M. Warlters (2012). The marginal cost of public funds and tax reform in Africa. *Journal of Development Economics* 97(1), 58–72.
- Aygün, A. H., M. G. Kirdar, M. Koyuncu, and Q. Stoeffler (2024). Keeping refugee children in school and out of work: Evidence from the world’s largest humanitarian cash transfer program. *Journal of Development Economics* 168, 103266.
- Bachas, P., P. Gertler, S. Higgins, and E. Seira (2021). How debit cards enable the poor to save more. *Journal of Finance* 76(4), 1913–1957.
- Bahar, D., A. M. Ibáñez, and S. V. Rozo (2021). Give me your tired and your poor: Impact of a large-scale amnesty program for undocumented refugees. *Journal of Development Economics* 151, 102652.
- Baird, S., J. De Hoop, and B. Özler (2013). Income shocks and adolescent mental health. *Journal of Human Resources* 48(2), 370–403.
- Baird, S., C. McIntosh, and B. Özler (2013). Cash or condition? evidence from a cash transfer experiment. *American Economic Journal: Applied Economics* 5(2), 1–26.
- Baisa, B., L. W. Davis, S. W. Salant, and W. Wilcox (2010). The welfare costs of unreliable water service. *Journal of Development Economics* 92(1), 1–12.
- Balboni, C., J. Boehm, and M. Waseem (2026). Firm adaptation in production networks: Evidence from extreme weather events in Pakistan. *American Economic Review*. Forthcoming.
- Banerjee, A., E. Duflo, C. Imbert, S. Mathew, and R. Pande (2020). Can e-governance reduce capture of public programs? experimental evidence from India’s employment guarantee. *NBER Working Paper* (22762).
- Barahona, N., F. A. Gallego, and J.-P. Montero (2020). Vintage-specific driving restrictions. *The Review of Economic Studies* 87(4), 1646–1682.
- Barnwal, P. (2024). Curbing leakage in public programs: Evidence from India’s direct benefit transfer policy. *American Economic Review* 114(12), 3812–3846.
- Barreca, A., K. Clay, O. Deschênes, M. Greenstone, and J. S. Shapiro (2016). Adapting to climate change: The remarkable decline in the US temperature-mortality relationship over the twentieth century. *Journal of Political Economy* 124(1), 105–159.
- Barron, K., H. Harmgart, S. Huck, S. O. Schneider, and M. Sutter (2023). Discrimination, narratives, and family history: An experiment with Jordanian host and Syrian refugee children. *The Review of Economics and Statistics* 105(4), 1008–1016.

- Barzel, Y. (1974, April). A theory of rationing by waiting. *The Journal of Law & Economics* 17(1), 73–95.
- Baseler, T., T. Ginn, R. Hakiza, H. Ogude-Chambert, and O. Woldemikael (2023). Can redistribution change policy views?: Aid and attitudes toward refugees in Uganda. *Center for Global Development, Working Paper 645*.
- Baumol, W. J. (1977). On the proper cost tests for natural monopoly in a multiproduct industry. *The American Economic Review* 67(5), 809–822.
- Beltramo, T., F. Nimoh, and O. Matthew (2024). Financial security, climate shocks and social cohesion. *Center for Economic Policy Research, Working Paper*.
- Bivins, A. W., T. Sumner, E. Kumpel, G. Howard, O. Cumming, I. Ross, K. Nelson, and J. Brown (2017). Estimating infection risks and the global burden of diarrheal disease attributable to intermittent water supply using QMRA. *Environmental Science & Technology* 51(13), 7542–7551.
- Borenstein, S. (2012). The redistributive impact of nonlinear electricity pricing. *American Economic Journal: Economic Policy* 4(3), 56–90.
- Burgess, R., M. Greenstone, N. Ryan, and A. Sudarshan (2020). The consequences of treating electricity as a right. *Journal of Economic Perspectives* 34(1), 145–69.
- Burke, M., S. M. Hsiang, and E. Miguel (2015). Global non-linear effect of temperature on economic production. *Nature* 527(7577), 235–239.
- Calonico, S., M. D. Cattaneo, and M. H. Farrell (2020). Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs. *The Econometrics Journal* 23(2), 192–210.
- Calonico, S., M. D. Cattaneo, and R. Titiunik (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica* 82(6), 2295–2326.
- Carleton, T., A. Jina, M. Delgado, M. Greenstone, T. Houser, S. Hsiang, A. Hultgren, R. E. Kopp, K. E. McCusker, I. Nath, J. Rising, A. Rode, H. K. Seo, A. Viaene, J. Yuan, and A. T. Zhang (2022). Valuing the global mortality consequences of climate change accounting for adaptation costs and benefits. *Quarterly Journal of Economics* 137(4), 2037–2105.
- Cattaneo, M. D., M. Jansson, and X. Ma (2020). Simple local polynomial density estimators. *Journal of the American Statistical Association* 115(531), 1449–1455.
- Cellhay, P. A., P. J. Gertler, P. Giovagnoli, and C. Vermeersch (2019). Long-run effects of temporary incentives on medical care productivity. *American Economic Journal: Applied Economics* 11(3), 92–127.

- Chetty, R., N. Hendren, and L. F. Katz (2016, April). The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *American Economic Review* 106(4), 855–902.
- Christodoulou, S. and A. Agathokleous (2012). A study on the effects of intermittent water supply on the vulnerability of urban water distribution networks. *Water Science and Technology: Water Supply* 12(4), 523–530.
- Cole, S., T. Sampson, and B. Zia (2011). Prices or knowledge? what drives demand for financial services in emerging markets? *Journal of Finance* 66(6), 1933–1967.
- Crowne, D. P. and D. Marlowe (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology* 24(4), 349–354.
- Davis, L. W. (2008). The effect of driving restrictions on air quality in Mexico City. *Journal of Political Economy* 116(1), 38–81.
- Davis, L. W. and L. Kilian (2011). The allocative cost of price ceilings in the U.S. residential market for natural gas. *Journal of Political Economy* 119(2), 212–241.
- Deacon, R. T. and J. Sonstelie (1985). Rationing by waiting and the value of time: Results from a natural experiment. *Journal of Political Economy* 93(4), 627–647.
- DellaVigna, S., N. Otis, and E. Vivalt (2020). Forecasting the results of experiments: Piloting an elicitation strategy. Working paper, National Bureau of Economic Research.
- DellaVigna, S. and D. Pope (2018). What motivates effort? Evidence and expert forecasts. *The Review of Economic Studies* 85(2), 1029–1069.
- Demirgüç-Kunt, A., L. Klapper, D. Singer, and S. Ansar (2022). The Global Findex Database 2021: Financial inclusion, digital payments, and resilience in the age of COVID-19. *World Bank Policy Research Working Paper*.
- Deschênes, O., M. Greenstone, and J. S. Shapiro (2017). Defensive investments and the demand for air quality: Evidence from the nox budget program. *American Economic Review* 107(10), 2958–2989.
- Devoto, F., E. Duflo, P. Dupas, W. Parienté, and V. Pons (2012). Happiness on tap: Piped water adoption in urban Morocco. *American Economic Journal: Economic Policy* 4(4), 68–99.
- Diamond, P. A. (1975). A many-person Ramsey tax rule. *Journal of Public Economics* 4(4), 335–342.
- Diamond, R., T. McQuade, and F. Qian (2019). The effects of rent control expansion on tenants, landlords, and inequality: Evidence from San Francisco. *American Economic Review* 109(9), 3365–94.

- Dieter, C. A., M. A. Maupin, R. R. Caldwell, M. A. Harris, T. I. Ivahnenko, J. K. Lovelace, N. L. Barber, and K. S. Linsey (2018). Estimated use of water in the united states in 2015. USGS Circular 1441, U.S. Geological Survey, Reston, Virginia. Supersedes USGS Open-File Report 2017-1131.
- Donna, J. D. and J.-A. Espín-Sánchez (2023). The illiquidity of water markets. *Review of Economic Studies*. Accepted.
- Dupas, P. (2014). Short-run subsidies and long-run adoption of new health products: Evidence from a field experiment. *Econometrica* 82(1), 197–228.
- Dustmann, C., K. Vasiljeva, and A. Damm (2019). Refugee migration and electoral outcomes. *The Review of Economic Studies* 86(5), 2035–2091.
- Egger, D., J. Haushofer, E. Miguel, P. Niehaus, and M. Walker (2022). General equilibrium effects of cash transfers: experimental evidence from Kenya. *Econometrica* 90(6), 2603–2643.
- Ericson, K. M. M. (2014). Consumer inertia and firm pricing in the Medicare Part D prescription drug insurance exchange. *American Economic Journal: Economic Policy* 6(1), 38–64.
- Feldstein, M. S. (1972). Distributional equity and the optimal structure of public prices. *The American Economic Review* 62(1), 32–36.
- Ferraro, P. J. and M. K. Price (2013). Using nonpecuniary strategies to influence behavior: Evidence from a large-scale field experiment. *Review of Economics and Statistics* 95(1), 64–73.
- Frech, H. E. and W. C. Lee (1987). The welfare cost of rationing-by-queuing across markets: Theory and estimates from the u.s. gasoline crises. *The Quarterly Journal of Economics* 102(1), 97–108.
- Galiani, S., P. Gertler, and E. Schargrodsky (2005). Water for life: The impact of the privatization of water services on child mortality. *Journal of Political Economy* 113(1), 83–120.
- Ghosn, F., A. Braithwaite, and T. S. Chu (2019). Violence, displacement, contact, and attitudes toward hosting refugees. *Journal of Peace Research* 56(1), 118–133.
- Glaeser, E. L. and E. F. P. Luttmer (2003). The misallocation of housing under rent control. *The American Economic Review* 93(4), 1027–1046.
- Guiteras, R., J. Levinsohn, and A. M. Mobarak (2015). Encouraging sanitation investment in the developing world: A cluster-randomized trial. *Science* 348(6237), 903–906.
- Guo, R., J. Yi, J. Zhang, and N. Zhang (2025). Rationed fertility: Treatment effect heterogeneity in the child quantity-quality trade-off. *Journal of Political Economy* 133(10), 3349–3386.

- Hagen-Zanker, J., M. Ulrichs, and R. Holmes (2018). What are the effects of cash transfers for refugees in the context of protracted displacement? Findings from Jordan. *International Social Security Review* 71(2), 57–77.
- Hagerty, N. (2023). What holds back water markets? transaction costs and the gains from trade. Working paper.
- Handel, B. R. (2013). Adverse selection and inertia in health insurance markets: When nudging hurts. *American Economic Review* 103(7), 2643–2682.
- Haushofer, J., J. Reisinger, and J. Shapiro (2015). Your gain is my pain: Negative psychological externalities of cash transfers. Working Paper.
- Haushofer, J. and J. Shapiro (2016). The short-term impact of unconditional cash transfers to the poor: Experimental evidence from Kenya. *The Quarterly Journal of Economics* 131(4), 1973–2042.
- Haushofer, J. and J. Shapiro (2018). The long-term impact of unconditional cash transfers: experimental evidence from Kenya. *Busara Center for Behavioral Economics Working Paper, Nairobi, Kenya*.
- Hidrobo, M., J. Hoddinott, A. Peterman, A. Margolies, and V. Moreira (2014). Cash, food, or vouchers? Evidence from a randomized experiment in northern Ecuador. *Journal of Development Economics* 107, 144–156.
- Hsiao, A. (2024). Sea level rise and urban adaptation in Jakarta. Working paper, Stanford University.
- Ibáñez, A., A. Moya, M. A. Ortega, S. Roza, and M. J. Urbina Florez (2024). Life out of the shadows: The impacts of regularization program on the lives of forced migrants. *Journal of the European Economic Association*.
- Ito, K. (2014). Do consumers respond to marginal or average price? Evidence from nonlinear electricity pricing. *The American Economic Review* 104(2), 537–563.
- Jacobsen, M. R., C. R. Knittel, J. M. Sallee, and A. A. van Benthem (2020). The use of regression statistics to analyze imperfect pricing policies. *Journal of Political Economy* 128(5), 1826–1876.
- Jeuland, M., M. Pucilowski, D. Hudner, A. Smith, T. Thompson, C. Seybolt, and A. Wyatt (2020). Impact evaluation of the MCA Jordan compact: Endline report. Technical report, Social Impact, Washington, DC. Prepared for the Millennium Challenge Corporation.
- Klassert, C., J. Yoon, K. Sigel, B. Klauer, S. Talozzi, T. Lachaut, P. Selby, S. Knox, N. Avisse, A. Tilmant, J. J. Harou, D. Mustafa, B. B. Josué Medellín-Azuara, H. Zhang, E. Gawel, and S. M. Gorelick (2023). Unexpected growth of an illegal water market. *Nature Sustainability* 6, 1406–1417.

- Kling, J. R., J. B. Liebman, and L. F. Katz (2007). Experimental analysis of neighborhood effects. *Econometrica* 75(1), 83–119.
- Kremer, M., J. Leino, E. Miguel, and A. P. Zwane (2011). Spring cleaning: Rural water impacts, valuation, and property rights institutions. *Quarterly Journal of Economics* 126(1), 145–205.
- Kumar, T. (2021). Home-price subsidies increase local-level political participation in urban India. *Journal of Politics*. Forthcoming.
- Kumpel, E., C. Woelfle-Erskine, I. Ray, and K. L. Nelson (2017). Measuring household consumption and waste in unmetered, intermittent piped water systems. *Water Resources Research* 53(4), 3220–3235.
- Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies* 76(3), 1071–1102.
- Lee, J. N., J. Morduch, S. Ravindran, and A. S. Shonchoy (2022). Narrowing the gender gap in mobile banking. *Journal of Economic Behavior & Organization* 193, 276–293.
- Lehmann, C. and D. Masterson (2014). Emergency Economies: the Impact of Cash Assistance in Lebanon. Technical report, International Rescue Committee.
- Lehmann, M. C. and D. T. Masterson (2020). Does aid reduce anti-refugee violence? Evidence from Syrian refugees in Lebanon. *American Political Science Review* 114(4), 1335–1342.
- Levinson, A. and E. Silva (2022, May). The electric gini: Income redistribution through energy prices. *American Economic Journal: Economic Policy* 14(2), 341–65.
- Lindsay, C. M. and B. Feigenbaum (1984, June). Rationing by waiting lists. *The American Economic Review* 74(3), 404–417.
- Ludwig, J., G. J. Duncan, L. A. Gennetian, L. F. Katz, R. C. Kessler, J. R. Kling, and L. Sanbonmatsu (2013, May). Long-term neighborhood effects on low-income families: Evidence from moving to opportunity. *American Economic Review* 103(3), 226–31.
- Madajewicz, M., A. Pfaff, A. van Geen, J. Graziano, I. Hussein, H. Momotaj, R. Sylvi, and H. Ahsan (2007). Can information alone change behavior? response to arsenic contamination of groundwater in Bangladesh. *Journal of Development Economics* 84(2), 731–754.
- McRae, S. (2015). Infrastructure quality and the subsidy trap. *American Economic Review* 105(1), 35–66.
- Ministry of Water and Irrigation (2023). National water strategy 2023–2040. Technical report, Ministry of Water and Irrigation, Hashemite Kingdom of Jordan, Amman, Jordan.

- Ministry of Water and Irrigation and USAID (2025). IT roadmap for the water sector in Jordan (2026–2030). Strategic planning document, Hashemite Kingdom of Jordan, Amman, Jordan. Outlines CAPEX requirements for digital transformation and smart metering.
- Miyahuna (2023). Annual report 2022. Technical report, Jordan Water Company - Miyahuna, Amman, Jordan.
- Moussa, W., N. Salti, A. Irani, R. Al Mokdad, Z. Jamaluddine, J. Chaaban, and H. Ghattas (2022). The impact of cash transfers on Syrian refugee children in Lebanon. *World Development* 150, 105711.
- Muralidharan, K., P. Niehaus, and S. Sukhtankar (2016). Building state capacity: Evidence from biometric smartcards in India. *American Economic Review* 106(10), 2895–2929.
- Neary, J. P. and K. W. S. Roberts (1980). The theory of household behavior under rationing. *European Economic Review* 13(1), 25–42.
- NRC (2015). In search of a home: Access to adequate housing in Jordan. Technical report, Norwegian Refugee Council.
- Okunogbe, O. and V. Pouliquen (2022). Technology, taxation, and corruption: Evidence from the introduction of electronic tax filing. *American Economic Journal: Economic Policy* 14(1), 341–372.
- Özler, B., Ç. Çelik, S. Cunningham, P. F. Cuevas, and L. Parisotto (2021). Children on the move: Progressive redistribution of humanitarian cash transfers among refugees. *Journal of Development Economics* 153, 102733.
- Park, S.-H. and H. Y. Yu (2021). How useful is the center for epidemiologic studies depression scale in screening for depression in adults? an updated systematic review and meta-analysis. *Psychiatry Research* 302, 114037.
- Quattrochi, J., G. Bisimwa, P. Van Der Windt, and M. Voors (2022). Cash-like vouchers improve psychological well-being of vulnerable and displaced persons fleeing armed conflict. *PNAS Nexus* 1(3), pgac101.
- Rafey, W. (2023). Droughts, deluges, and (river) diversions: Valuing market-based water reallocation. *American Economic Review* 113(2), 430–71.
- Ramsey, F. P. (1927). A contribution to the theory of taxation. *The Economic Journal* 37(145), 47–61.
- Reiss, P. C. and M. W. White (2005). Household electricity demand, revisited. *The Review of Economic Studies* 72(3), 853–883.

- Riley, E., A. S. Shonchoy, and R. D. Osei (2025). Incentives and endorsement for technology adoption: Evidence from mobile banking in Ghana. *Journal of Development Economics* 176, 103511.
- Royer, H., M. Stehr, and J. Sydnor (2015). Incentives, commitments, and habit formation in exercise: Evidence from a field experiment with workers at a fortune-500 company. *American Economic Review* 105(4), 1341–1367.
- Rozo, S. V., A. Quintana, and M. J. Urbina (2023). The electoral consequences of easing the integration of forced migrants. Policy research working paper, World Bank.
- Rozo, S. V. and M. Sviatschi (2021). Is a refugee crisis a housing crisis? only if housing supply is unresponsive. *Journal of Development Economics* 148, 102563.
- Ryan, N. and A. Sudarshan (2022). Rationing the commons. *Journal of Political Economy* 130(1), 210–257.
- Saez, E. (2002). Optimal income transfer programs: Intensive versus extensive labor supply responses. *The Quarterly Journal of Economics* 117(3), 1039–1073.
- Saez, E. and S. Stantcheva (2016). Generalized social marginal welfare weights for optimal tax theory. *The American Economic Review* 106(1), 24–45.
- Sallee, J. M. (2022). Pigou creates losers: On the implausibility of achieving pareto improvements from efficiency-enhancing policies. Energy Institute Working Paper 302, Energy Institute at Haas, University of California Berkeley.
- Schaner, S. (2018). The persistent power of behavioral change: Long-run impacts of temporary savings subsidies for the poor. *American Economic Journal: Applied Economics* 10(3), 67–100.
- Szabó, A. and G. Ujhelyi (2015). Reducing non-payment for public utility services: Experimental evidence from South Africa. *Journal of Development Economics* 117, 20–31.
- Szabó, A. (2015). The value of free water: Analyzing South Africa’s free basic water policy. *Econometrica* 83(5), 1913–1961.
- Tauchmann, H. (2014). Lee (2009) treatment-effect bounds for nonrandom sample selection. *The Stata Journal* 14(4), 884–894.
- Taylor, D. D. J., A. H. Slocum, and A. J. Whittle (2018). Analytical scaling relations to evaluate leakage and intrusion in intermittent water supply systems. *PLOS ONE* 13(5), e0196887.
- Timmins, C. (2002). Measuring the dynamic efficiency costs of regulators’ preferences: Municipal water utilities in the arid west. *Econometrica* 70(2), 603–629.
- Tobin, J. (1952). A survey of the theory of rationing. *Econometrica* 20(4), 521–553.

- UNHCR (2023a). Jordan. Technical report, United Nations High Commissioner for Refugees.
- UNHCR (2023b). Syria. Technical report, United Nations High Commissioner for Refugees.
- UNHCR (2023c). Unhcr global trends. Technical report, United Nations High Commissioner for Refugees.
- UNHCR (2025). Refugee statistics. Accessed April 11, 2026.
- United Nations (2025). World urbanization prospects 2025: Summary of results. Technical Report UN DESA/POP/2025/TR/NO. 12, United Nations, Department of Economic and Social Affairs, Population Division, New York.
- USAID Water Governance Activity (2022, May). National non-revenue water strategy for Jordan's water sector (2022–2040). Technical report, Ministry of Water and Irrigation, Hashemite Kingdom of Jordan. Prepared by Chemonics International Inc. under the USAID Water Governance Activity (WGA).
- Utilities Performance Monitoring Unit (2023). Jordan water utilities monitoring report 2022. Annual monitoring report, Ministry of Water and Irrigation, Hashemite Kingdom of Jordan, Amman, Jordan.
- Weitzman, M. L. (1991, June). Price distortion and shortage deformation, or what happened to the soap? *The American Economic Review* 81(3), 401–414.
- WHO and UNICEF (2025). Progress on household drinking-water, sanitation and hygiene 2000–2024: Special focus on inequalities. Technical report, World Health Organization and UNICEF, Geneva.
- Woerner, W., B. Fleitlich-Bilyk, R. Martinussen, J. Fletcher, G. Cucchiaro, P. Dalgarrondo, M. Lui, and R. Tannock (2004). The strengths and difficulties questionnaire overseas: evaluations and applications of the sdq beyond Europe. *European Child & Adolescent Psychiatry* 13, ii47–ii54.
- World Health Organization (2023). Drinking-water. Fact sheet. Accessed April 11, 2026.
- Zhou, Y.-Y. (2019). How refugee resentment shapes national identity and citizen participation in Africa. In *APSA 2018 Annual Meeting Paper*.

Appendix A

Appendix to Rationing by Intermittency

A.1 Solution Algorithm

Algorithm 1: Numerical Computation of Counterfactual Equilibria

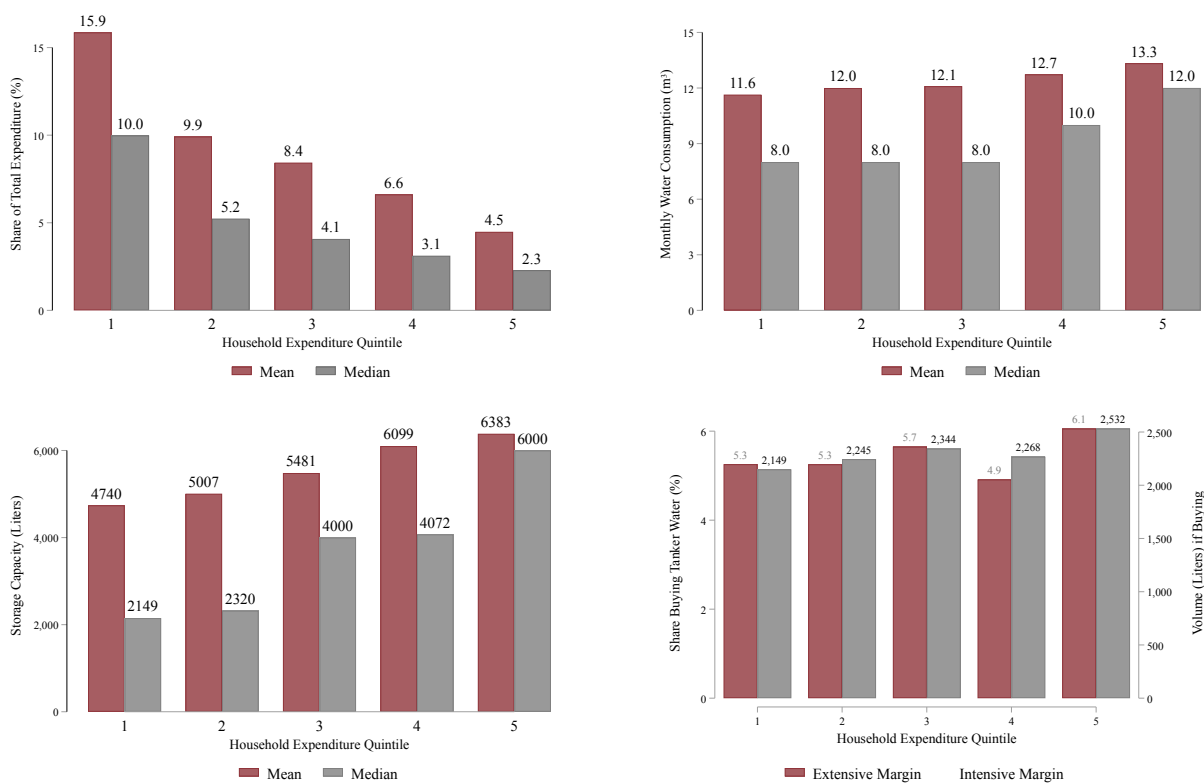
```

1
2 EvaluatePolicy( $p, n, \bar{q}, \delta$ ) Solve  $K_i^*$  from FOC:  $\mathbb{E}[\mu_{K,i}] = g(K_i^*) \quad \forall i$ ;
3 for  $m = 1, \dots, M$  do
4    $q_i^* \leftarrow \hat{\mu}_i + \omega_i^{(m)}$ ;
5   Classify  $\mathcal{R}_1/\mathcal{R}_2/\mathcal{R}_3 \rightarrow$  compute  $q_i^{(m)}, \text{GU}_i^{(m)}$ ;
6  $V_i \leftarrow \mathbb{E}_M[\text{GU}_i] - p \mathbb{E}_M[q_i] - \frac{g}{2}(K_i^*)^2$ ;
7  $\mathcal{W} \leftarrow \sum_i g_i V_i + (p - \Lambda) \sum_i \mathbb{E}_M[q_i] - F \cdot N$ ;
8 _____;
9 /* Search over instruments, enforcing  $\iota(n) Q \leq S$  */;
10 Third-Best ( $p = \bar{p}, \bar{q} = \infty$ );
11 for  $n \in \{\Delta n, 2\Delta n, \dots, T/\delta\}$  do
12    $\mathcal{W} \leftarrow \text{EvaluatePolicy}(\bar{p}, n, \infty, \delta) \rightarrow$  keep best feasible  $n$ 
13 First-Best ( $\bar{q} = \infty$ );
14 for  $n \in \{\Delta n, 2\Delta n, \dots, T/\delta\}$  do
15    $p^* \leftarrow \arg \max_p \mathcal{W}(p, n) \rightarrow$  if  $\iota Q > S$ :  $p^* \leftarrow$  root of  $\iota Q(p) = S$ ;
16    $\mathcal{W} \leftarrow \text{EvaluatePolicy}(p^*, n, \infty, \delta) \rightarrow$  keep best feasible  $(p, n)$ 
17 Second-Best ( $p = \bar{p}$ );
18 for  $n \in \{\Delta n, 2\Delta n, \dots, T/\delta\}$  do
19   EvaluatePolicy( $\bar{p}, n, \infty, \delta$ )  $\rightarrow$  if  $\iota Q > S$ :  $\bar{q}^* \leftarrow$  root of  $\iota Q(\bar{q}) = S$ ;
20    $\mathcal{W} \leftarrow \text{EvaluatePolicy}(\bar{p}, n, \bar{q}^*, \delta) \rightarrow$  keep best feasible  $(n, \bar{q})$ 

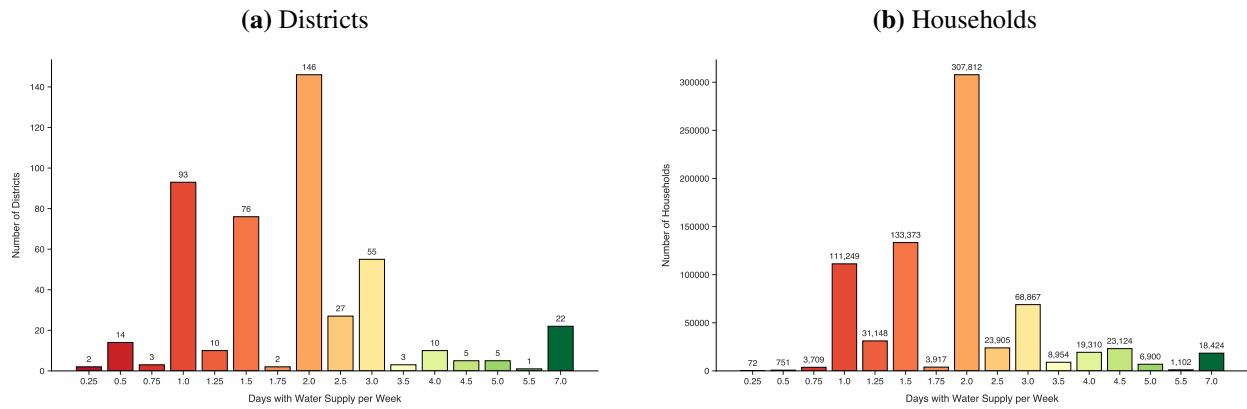
```

A.2 Additional Figures and Tables

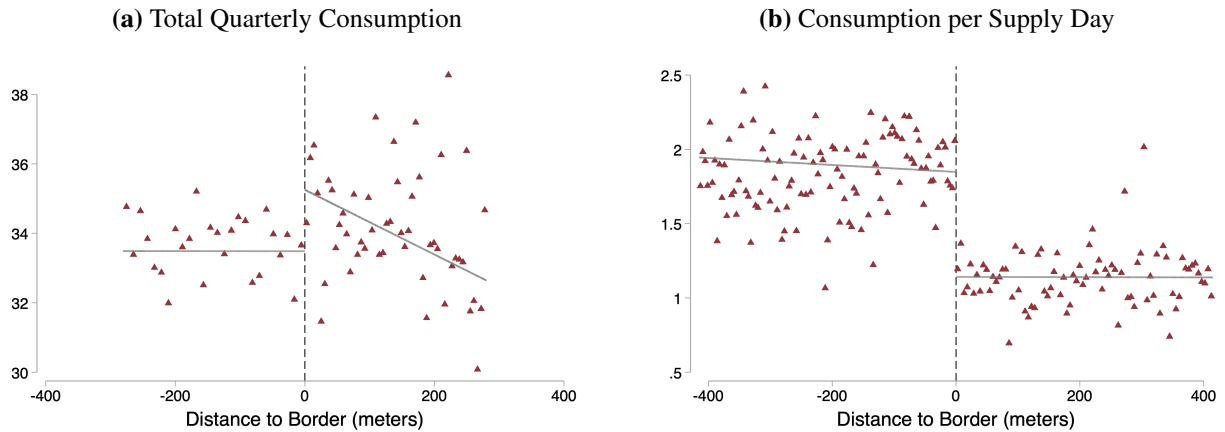
Figure (A.1) Patterns by Expenditure Quintile



Notes: Households are sorted into quintiles based on total monthly expenditure. Top-left: water budget share by expenditure quintile (mean and median). Top-right: piped water consumption levels. Bottom-left: mean storage capacity. Bottom-right: extensive margin (share purchasing, left axis) and intensive margin (volume conditional on purchasing, right axis) of private tanker water. The mechanical relationship in the top-left panel (water share has expenditure in the denominator) should be interpreted with care.

Figure (A.2) Distribution of Water Supply Days

Notes: Water is supplied intermittently through designated “water zones” established by Miyahuna, Amman’s utility. Allocation across zones reflects topography, customer density, the presence of critical infrastructure (e.g., hospitals), and historical precedent. Panel (a) shows the distribution of water zones by supply period. Panel (b) shows the distribution of geocoded households mapped to each zone. The household and water zone counts are closely aligned within each supply-period bin.

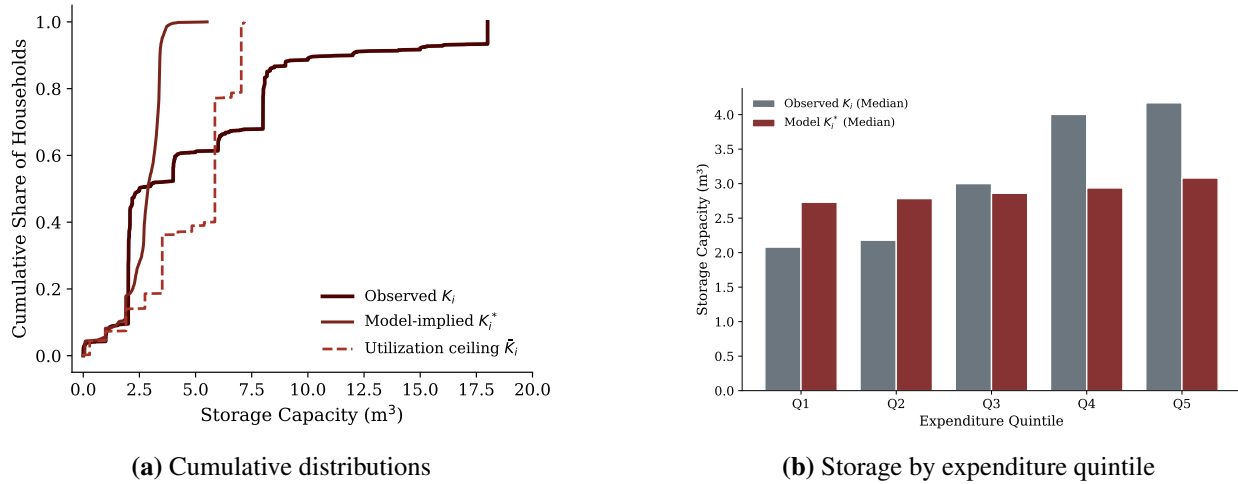
Figure (A.3) Change in Water Consumption at the Supply Boundary

Notes: Binned averages of water consumption within the CER-optimal bandwidth, with local linear fits on each side of the cutoff. Panel (a) plots total quarterly consumption (m^3 , winsorized at the 99th percentile). Panel (b) plots the network withdrawal rate, defined as weekly consumption divided by supply days per week, i.e., $(\text{Quarterly Consumption}/13)/n_{d(i)}$, winsorized at the 99th percentile. Households on the high-supply side consume modestly more water overall but their consumption per supply day is lower, indicating they distribute consumption across the expanded supply window. This pattern is consistent with storage constraints binding (Case 2) or with unconstrained households (Case 1) spreading a fixed demand over more days.

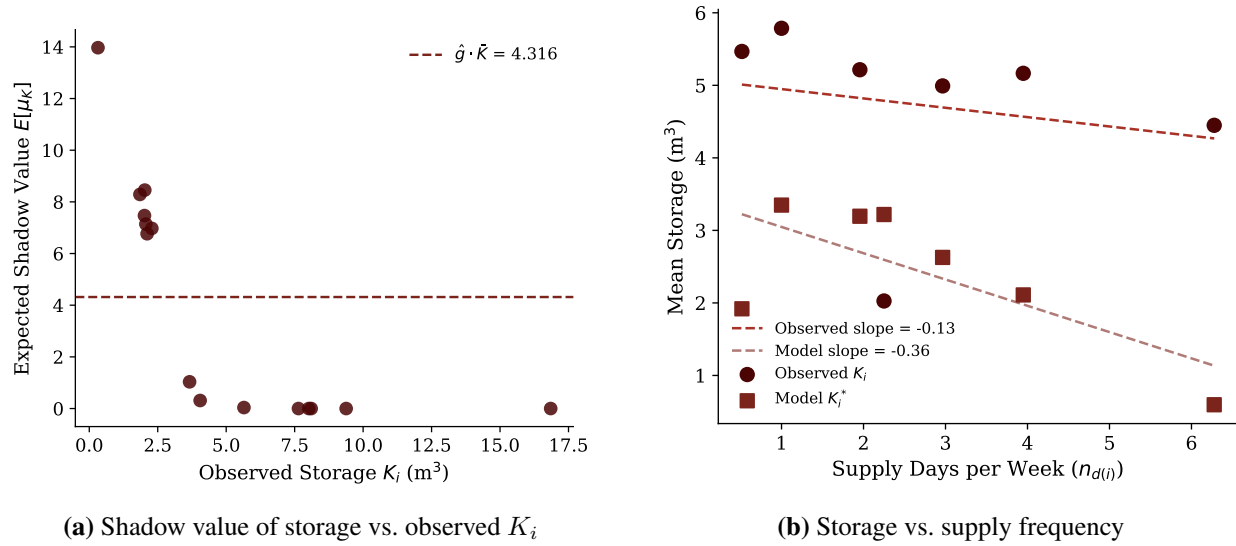
Table (A.1) Econometric Regimes

Regime	Condition	Observed q_{it}	Information about q_{it}^*
$\mathcal{R}_1$: Unconstrained	$q_{it} < \min\{\bar{q}_i^K, \phi n_{d(i)}\}$	$q_{it} = q_{it}^*$	Fully revealed
$\mathcal{R}_2$: Storage-constrained	$q_{it} \geq \bar{q}_i^K$ and $q_{it} < \phi n_{d(i)}$	$q_{it} = q_{it}^* \tau_i + K_i$	Revealed via transformation
$\mathcal{R}_3$: Pipe-constrained	$q_{it} \geq \phi n_{d(i)}$	$q_{it} = \phi n_{d(i)}$	Right-censored

Notes: $\bar{q}_i^K \equiv K_i/(1 - \tau_i)$ is the smoothing threshold. $\phi n_{d(i)}$ is the pipe ceiling. $\tau_i = n_{d(i)}/T$. Regime $\mathcal{R}_3$ combines Cases 3 and 4 from Table A.8.

Figure (A.4) Observed and Model-Optimal Storage Capacity

Notes: Panel (a): CDFs of observed storage K_i , model-implied optimal K_i^* , and the utilization ceiling $\bar{K}_i$. Most households hold substantially more storage than the weekly-smoothing optimum, consistent with storage capacity being determined primarily by housing characteristics (Table A.3) rather than by the demand-side fundamentals the model captures. Panel (b): median observed and model-optimal storage by expenditure quintile. The gap between observed and optimal widens with income, reflecting the positive correlation between income and housing size documented in Table A.3. Because our welfare baseline uses the model-implied K_i^* , which understates actual storage costs, the estimated welfare gains from reform are conservative.

Figure (A.5) Storage, Shadow Value, and Supply Response

Notes: Panel (a): expected shadow value $\mathbb{E}[\mu_K]$ declines steeply in observed storage; households with $K_i > 5 \text{ m}^3$ face near-zero marginal returns. Panel (b): observed and model-predicted mean storage against supply days per week; the slopes are closely matched, supporting the comparative statics (Proposition 2).

Table (A.2) Data Sources and Identification

	Billing Universe	Household Survey
Observations	~910,000 HH	~3,500 HH
Time periods	20 quarters	4 survey waves
Metered consumption q	✓	✓
Bill / average price p	✓	—
Storage capacity K_i	—	✓
Hours of supply	—	✓
Income and demographics	—	✓
Identifies	b	$\alpha_0, \beta, \sigma_\varepsilon$

Notes: The billing universe provides tariff variation to identify the demand slope b via IV. The household survey provides storage capacity and demographics for the structural MLE. The billing-matched panel ($\sim 4,500 \text{ HH} \times 8$ quarters) identifies the transitory component σ_ε .

Table (A.3) Determinants of Observed Storage Capacity

	(1) Housing	(2) Model	(3) Combined
Expenditure ($\times 100$ JD)	0.129*** (0.025)		0.054 (0.118)
Amman (=1)	-0.019 (0.198)		-0.056 (0.202)
Supply days/week ($n_{d(i)}$)		-0.225*** (0.065)	-0.225*** (0.065)
Flow rate (ϕ_i)		0.220 (0.220)	0.616 (0.887)
Latent demand ($\hat{\mu}_i$)		1.144*** (0.262)	0.592 (1.246)
Weekly consumption (q_i^w)		0.272*** (0.055)	0.293*** (0.072)
Wave FE	Yes	Yes	Yes
N	2325	2325	2325
R^2	0.011	0.031	0.032

Notes: Dependent variable: observed storage capacity K_i (m³). Col. (1): housing and demographic characteristics with wave FE. Col. (2): structural model variables with wave FE. Col. (3): combined. Expenditure divided by 100 for readability. Standard errors in parentheses. ***, **, *: 1%, 5%, 10%.

Table (A.4) Observed vs. Model-Implied Storage Capacity

	Mean	SD	P25	Median	P75	N
Observed K_i	5.24	4.71	2.00	2.50	8.00	2,325
Model K_i^*	2.70	0.87	2.37	2.88	3.35	2,325
Utilization ceiling $\bar{K}_i$	4.87	1.98	3.52	5.86	5.86	2,325
Excess capacity ($K_i - \bar{K}_i$) ⁺	2.15	3.79	0.00	0.00	2.32	2,325
Corr(K_i^* , K_i)	0.040					

Notes: K_i is the observed household storage capacity from the survey. K_i^* is the model-implied optimal capacity from the first-stage FOC. $\bar{K}_i \equiv \phi n_{d(i)}(1 - \tau_i)$ is the maximum capacity that can be filled during supply periods, the utilization ceiling from Proposition 5. Excess capacity is $\max(K_i - \bar{K}_i, 0)$.

Table (A.5) Oversized Storage by Income Quintile

	Q1	Q2	Q3	Q4	Q5	All
Mean K_i (m ³)	4.6	4.8	5.2	6.0	6.1	5.2
Mean $\bar{K}_i$ (m ³)	4.9	4.8	5.0	4.8	4.9	4.9
% oversized ($K_i > \bar{K}_i$)	42.0%	44.7%	46.3%	51.6%	55.3%	47.1%
Mean excess (m ³)	1.7	1.9	2.0	2.8	2.8	2.2

Notes: Proposition 5 states that no household at an interior storage optimum would hold capacity exceeding $\bar{K}_i \equiv \phi n_{d(i)}(1 - \tau_i)$, since additional capacity can never be filled during the supply window. “Oversized” is the fraction violating this condition. Income quintiles are based on total household expenditure.

Table (A.6) Common Tank Sizes

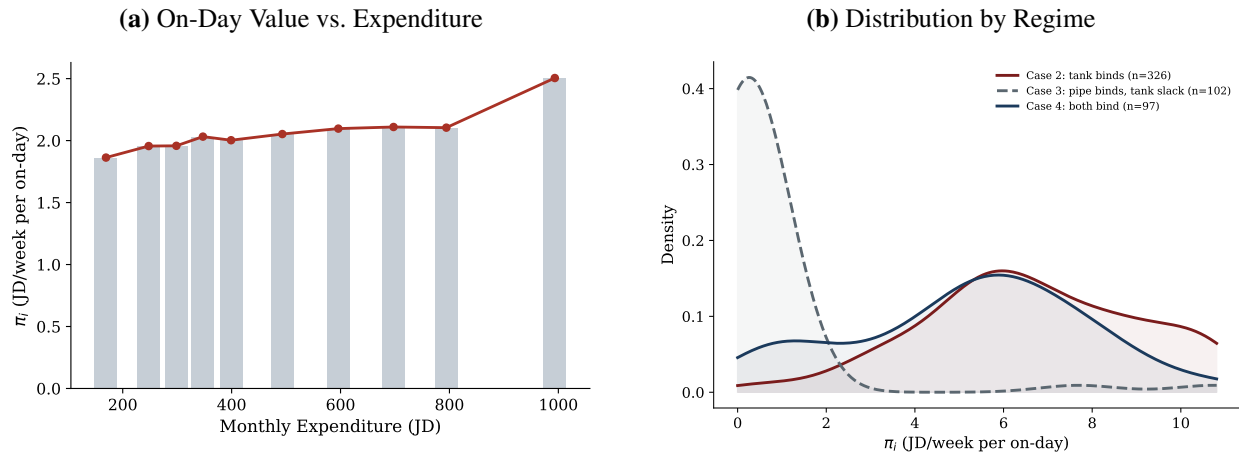
Tank Size (m ³)	Households	Share (%)
2.0	650	28.0
2.1	170	7.3
4.0	146	6.3
8.0	300	12.9
18.0	156	6.7
Top 5 sizes	1,422	61

Notes: Tank sizes are rounded to the nearest 100 liters. The concentration at a small number of discrete sizes is consistent with standardized tank markets where households select from a fixed menu of available capacities rather than optimizing on a continuous margin.

Table (A.7) Observed vs. Predicted Moments

	Data			Model			Corr
	Mean	Median	SD	Mean	Median	SD	
Consumption q_i (m ³ /wk)	2.60	1.85	1.84	3.04	3.08	0.85	0.160
Storage K_i (m ³)	5.24	2.50	4.71	2.70	2.88	0.87	0.040
Withdrawal rate q_i/n (m ³ /d)	1.96	1.11	2.65	1.96	1.56	1.13	0.543

Notes: Model predictions use each household’s observed supply schedule $n_{d(i)}$, storage capacity K_i , and MLE-estimated latent demand $\hat{\mu}_i$. Corr is the cross-sectional Pearson correlation between observed and predicted values.

Figure (A.6) Private Marginal Value of an On-Day

Notes: Panel (a): mean π_i (private value of an additional supply day) within vigintiles of monthly expenditure. The weakly positive gradient explains why $\bar{g}_\pi \approx 1$. Panel (b): kernel densities of π_i for tank-constrained ($\mathcal{R}_2$) and doubly-constrained (Case 4) households. The dashed line marks the per-household social marginal cost of a unit of water. Unconstrained households ($\pi_i = 0$) are omitted.

A.3 Omitted Proofs

Household Problem Solution

Optimal Smoothing. The first-order condition for an interior solution requires equating the marginal utility of consumption across states:

$$u'_i(q_i^{\text{on}}) = u'_i(q_i^{\text{off}}) = p \quad (\text{A.1})$$

The household can achieve perfect smoothing only if the required off-period consumption volume does not exceed storage capacity:

$$(T - n)q_i^{\text{off}} \leq K_i \quad (\text{A.2})$$

Defining $\tau \equiv n/T$ as the share of on-days, the smoothing threshold is:¹

$$\bar{q}_i^K \equiv \frac{K_i}{1 - \tau} \quad (\text{A.3})$$

This is the maximum total consumption that the tank can support with perfect temporal smoothing. When desired consumption $q_i^* > \bar{q}_i^K$, the tank constraint binds, forcing the household to reduce off-day consumption to $q_i^{\text{off}} = K_i/(T - n)$, thereby incurring a utility penalty from volatility.

Definition 10 (Unconstrained Demand). For household i with preference realization ω_i , define the unconstrained total demand q_i^* as the solution to $u'_i(q_i^*/T) = p$, i.e.,

$$q_i^* \equiv T \cdot (u'_i)^{-1}(p) \quad (\text{A.4})$$

where $(u'_i)^{-1}$ denotes the inverse of the marginal utility function. Under strict concavity, q_i^* is unique and decreasing in p .

Solution: Four Cases. The interaction of the pipe constraint ($q_i \leq \phi n$) and the smoothing constraint ($q_i \leq \bar{q}_i^K$) generates four regimes. Table A.8 summarizes the results.

Case 1: Unconstrained. Neither pipe nor tank binds.

Condition: Aggregate consumption $q_i^* \leq \phi n$ and the aggregate stored amount $q_i^*(1 - \tau) \leq K_i$.

$$q_i = q_i^* \quad (\text{A.5})$$

Consumption is perfectly smoothed at the unconstrained optimum, with total off-period consumption volume being equal to $q_i^*(1 - \tau)$ and that of on-period being $q_i^* \cdot \tau$.

Case 2: Storage-constrained only. Tank binds but pipe does not.

Condition: $q_i^*(1 - \tau) > K_i$ and the pipe constraint is slack.

¹Substituting the perfect smoothing condition $q_i^{\text{off}} = q_i/T$: $(T - n)\frac{q_i}{T} \leq K_i \implies (1 - \frac{n}{T})q_i \leq K_i$. This simplifies to: $(1 - \tau)q_i \leq K_i \implies q_i \leq \frac{K_i}{1 - \tau}$

The household sets off-period consumption to the tank limit K_i and optimizes on-days consumption. The on-days daily rate is $(q_i - K_i)/n$, and the off-days rate is $K_i/(T - n)$.

$$q_i = q_i^* \tau + K_i \quad (\text{A.6})$$

Total consumption is the sum of on-days optimal volume ($q_i^* \tau$) and the full tank (K_i). Note that $q_i < q_i^*$: the household does not exhaust all pipe water. The marginal value of on-days consumption equals the price p , so further on-days consumption is not worthwhile, yet the tank is full. The shadow price of tank capacity is proportional to the gap in marginal utility of consumption rates across sub-periods.

Case 3: Pipe-constrained only. Pipe binds but tank does not.

Condition: $q_i^* > \phi n$ and $\phi n(1 - \tau) \leq K_i$.

The household exhausts all pipe water and smooths it perfectly:

$$q_i = \phi n \quad (\text{A.7})$$

In this case, off-period consumption is $\phi n(1 - \tau)$. The consumption rate is $\phi \cdot \tau$ in both sub-periods. The shadow price of pipe capacity is given by the difference between optimal unconstrained consumption and the binding pipe consumption.

Case 4: Both bind. Pipe and tank are simultaneously binding.

Condition: $q_i^* > \phi n$ and $\phi n(1 - \tau) > K_i$.

The household consumes all pipe water but cannot smooth it:

$$q_i = \phi n \quad (\text{A.8})$$

On-days rate: $(\phi n - K_i)/n$ and off-days rate: $K_i/(T - n)$. Consumption is unsmooth even though all pipe water is used. Both shadow prices are positive.

Table (A.8) Household Classification

Case	Conditions	q_i	Smooth?	Multiplier
1. Unconstrained	$q_i^* \leq \phi n, q_i^*(1 - \tau) \leq K_i$	q_i^*	Yes	0
2. Storage	$q_i^* \tau + K_i \leq \phi n, q_i^*(1 - \tau) > K_i$	$q_i^* \tau + K_i$	No	$\mu_K > 0$
3. Pipe	$q_i^* > \phi n, \phi n(1 - \tau) \leq K_i$	ϕn	Yes	$\mu_\phi > 0$
4. Both bind	$q_i^* > \phi n, \phi n(1 - \tau) > K_i$	ϕn	No	$\mu_K, \mu_\phi > 0$

Taken together, Case 4 splits the welfare cost of intermittency into two channels - a quantity short-fall ($\mu_\phi > 0$) and a smoothing loss ($\mu_K > 0$) - each of which is individually smaller than the corresponding single-constraint shadow price in Cases 2 and 3. The *total* welfare cost, however, is the largest of the three regimes, as the household faces both constraints simultaneously.

Table (A.9) Shadow Prices by Regime

	q_i^{on}	q_i^{off}	μ_ϕ	μ_K
Case 1 (none)	$\frac{q_i^*}{T}$	$\frac{q_i^*}{T}$	0	0
Case 2 (tank only)	$\frac{q_i^*}{T}$	$\frac{K_i}{T-n}$	0	$u'_i\left(\frac{K_i}{T-n}\right) - p$
Case 3 (pipe only)	$\phi\tau$	$\phi\tau$	$u'_i(\phi\tau) - p$	0
Case 4 (both)	$\frac{\phi n - K_i}{n}$	$\frac{K_i}{T-n}$	$u'_i\left(\frac{\phi n - K_i}{n}\right) - p$	$u'_i\left(\frac{K_i}{T-n}\right) - u'_i\left(\frac{\phi n - K_i}{n}\right)$

Table (A.10) Private Marginal Value of an Additional Day by Regime

Case	Binding	L_i	$\mu_{\phi,i}\phi$	π_i
1. Unconstrained	Neither	0	0	0
2. Storage-constrained	Tank only	> 0	0	L_i
3. Pipe-constrained	Pipe only	0	> 0	$\mu_{\phi,i}^{(3)}\phi$
4. Both bind	Pipe + Tank	> 0	> 0	$L_i + \mu_{\phi,i}^{(4)}\phi$

Optimal Storage. The interior first-order condition equates the expected shadow value to the marginal cost of tank capacity ($\frac{\partial V_i}{\partial K_i} = 0$):

$$\mathbb{E}[\partial W_i / \partial K_i] = \Gamma'_i(K_i^*) \quad (\text{A.9})$$

Note that the partial derivative of the indirect utility (inside the expectation) is simply the shadow price $\mu_K(\omega)$.

The two components have distinct economic content and activate in different regimes: In Case 3, rates are equalized at $\phi\tau$ despite the pipe binding (the tank is large enough to absorb $\phi n(1 - \tau)$), so $L_i = 0$: the household suffers from insufficient total volume but not from temporal distortion. The entire value of an additional on-day comes through the quantity-expansion channel.

Propositions

I will assume in this proof that the instantaneous utility u_i is quadratic $u_i(q) = (a_i + \omega_i)q - \frac{b}{2}q^2$ where q is the daily consumption rate.

Proof of Proposition 1. Consider two on-days t, t' with $t < t'$ and suppose $q_{i,t} > q_{i,t'}$. Since on-days are contiguous, transferring $\varepsilon > 0$ of consumption from day t to day t' is feasible: reducing q_t by ε weakly increases $K_{i,t+1}$, and since all intermediate days are also on-days (by contiguity), the additional stored water propagates forward to day t' , where $q_{t'}$ can increase by ε . By strict concavity of u_i , this transfer strictly increases U_i , contradicting optimality.

For off-days, the argument is identical: off-days $n + 1, \dots, T$ are also contiguous by Assumption 2. Transferring consumption from a high- q off-day to a low- q off-day is feasible because the intervening storage levels adjust accordingly, and strict concavity ensures a welfare improvement. $\square$

Proof of Proposition 2. Higher n affects the return to storage through two channels. First, the smoothing threshold $\bar{q}_i^K = K_i T / (T - n)$ rises for any given K_i , reducing the probability that the tank binds (the critical shock $\bar{\omega}_i$ increases). Second, the insurance threshold $K_i(n) = q_i^*(T - n) / T$ falls, meaning less storage is needed for full smoothing. Both effects reduce the expected marginal value of storage. By the implicit function theorem applied to (A.9), $dK_i^* / dn < 0$.

At the extensive margin, when n rises enough that $K_i(n)$ falls below K_i^* , the household jumps discretely from Case 2 to Case 1, and optimal storage may drop to zero. $\square$

Proof of Proposition 3. Higher $\bar{p}$ reduces unconstrained demand q_i^* , lowering the insurance threshold $K_i = q_i^*(T - n) / T$ and reducing the probability that the tank binds. Both the probability-weighted shadow value and the shadow price in each binding state fall. By the implicit function theorem, $dK_i^* / dp < 0$. $\square$

Proof of Corollary 1. At $n = T$, there are no off-days ($T - n = 0$), so the smoothing threshold $\bar{q}_i^K = K_i T / (T - n) \rightarrow \infty$ for any $K_i > 0$. The tank never binds regardless of demand. Since $\Gamma'_i(K) > 0$ for all $K > 0$, the household sets $K_i^* = 0$. $\square$

Proposition 5 (Non-Existence of Case 3 at Interior Optimum). *Suppose storage costs satisfy Assumption 1. Then at the optimal storage level K_i^* , no household operates in Case 3 (pipe-constrained, tank slack) with positive probability: $\mathbb{P}(i \in \Omega_\phi \mid K_i = K_i^*) = 0$.*

Proof of Proposition 5. Case 3 requires the tank to be large enough to smooth the full pipe throughput: $K_i \geq \phi n (1 - \tau)$. I show that no household chooses such a large tank at an interior optimum.

At any $K \geq \phi n (1 - \tau)$, the tank is slack for every realization of ω . To see this, note that consumption is bounded above by the pipe ceiling, $q_i \leq \phi n$, so the off-period storage demand satisfies $(T - n) q_i^{\text{off}} \leq \phi n (1 - \tau) \leq K$. Hence the tank constraint does not bind in any state, and the shadow price is zero: $\mu_K(\omega; K) = 0$ for all ω . Taking expectations, $\mathbb{E}_\omega[\mu_K(\omega; K)] = 0$ on the entire region $K \geq \phi n (1 - \tau)$.

The storage FOC at an interior optimum requires $\mathbb{E}_\omega[\mu_K(\omega; K^*)] = \Gamma'(K^*)$. Because $\Gamma'(K) > 0$ for all $K > 0$ (Assumption 1) while $\mathbb{E}[\mu_K] = 0$ throughout $[\phi n (1 - \tau), \infty)$, the FOC cannot hold anywhere in this region. Hence $K_i^* < \phi n (1 - \tau)$. By the Case 3 definition, this implies $\mathbb{P}(i \in \Omega_\phi \mid K_i = K_i^*) = 0$.

Intuitively, a tank large enough to smooth the full pipe throughput would never bind, so its marginal insurance value is zero. Since the marginal cost of capacity is strictly positive, the household always stops investing before reaching this threshold. $\square$

Derivation of the Demand Functions

I will assume in this proof that the instantaneous utility u_i is quadratic $u_i(q) = (a_i + \omega_i)q - \frac{b}{2}q^2$ where q is the daily consumption rate.

By Proposition 1, within each sub-period the household equalizes daily consumption rates. The problem therefore reduces to choosing total consumption q_i and total stored volume $s_i \equiv (T - n) q_i^{\text{off}}$, where the on-period volume is $q_i - s_i$ and the daily rates are $q_i^{\text{on}} = (q_i - s_i)/n$ and $q_i^{\text{off}} = s_i/(T - n)$.

Objective in volume notation. Substituting the rate definitions into the separable objective $n u_i(q_i^{\text{on}}) + (T - n) u_i(q_i^{\text{off}})$ with $u_i(r) = (a_i + \omega_i)r - \frac{b}{2}r^2$, the quadratic terms become:

$$n \frac{b}{2} \left(\frac{q_i - s_i}{n} \right)^2 + (T - n) \frac{b}{2} \left(\frac{s_i}{T - n} \right)^2 = \frac{b}{2} \left[\frac{(q_i - s_i)^2}{n} + \frac{s_i^2}{T - n} \right]$$

and the linear terms sum to $(a_i + \omega_i) q_i$. Subtracting the expenditure $p q_i$ and defining $\theta_i \equiv a_i + \omega_i - p$, the household solves:

$$\max_{q_i, s_i} \theta_i q_i - \frac{b}{2} \left[\frac{(q_i - s_i)^2}{n} + \frac{s_i^2}{T - n} \right] \quad (\text{A.10})$$

subject to:

$$q_i \leq \phi n \quad (\mu_\phi \geq 0) \quad (\text{A.11})$$

$$s_i \leq K_i \quad (\mu_K \geq 0) \quad (\text{A.12})$$

$$0 \leq s_i \leq q_i \quad (\text{A.13})$$

Lagrangian.

$$\mathcal{L} = \theta_i q_i - \frac{b}{2} \left[\frac{(q_i - s_i)^2}{n} + \frac{s_i^2}{T - n} \right] + \mu_\phi (\phi n - q_i) + \mu_K (K_i - s_i) \quad (\text{A.14})$$

First-Order Conditions.

$$\frac{\partial \mathcal{L}}{\partial q_i} = \theta_i - b \frac{q_i - s_i}{n} - \mu_\phi = 0 \quad (\text{A.15})$$

$$\frac{\partial \mathcal{L}}{\partial s_i} = b \frac{q_i - s_i}{n} - b \frac{s_i}{T - n} - \mu_K = 0 \quad (\text{A.16})$$

In rate notation:

$$\theta_i - b q_i^{\text{on}} = \mu_\phi \quad (\text{A.17})$$

$$b(q_i^{\text{on}} - q_i^{\text{off}}) = \mu_K \quad (\text{A.18})$$

Equation (A.17): the gap between the marginal utility of on-day consumption and the price equals μ_ϕ . Equation (A.18): the gap in marginal utilities across sub-periods equals μ_K .

Complementary Slackness.

$$\mu_\phi \geq 0, \quad \mu_\phi(\phi n - q_i) = 0 \quad (\text{A.19})$$

$$\mu_K \geq 0, \quad \mu_K(K_i - s_i) = 0 \quad (\text{A.20})$$

The four cases correspond to $(\mu_\phi, \mu_K) \in \{0, >0\} \times \{0, >0\}$.

Case 1: Unconstrained ($\mu_\phi = 0, \mu_K = 0$). *Step 1 (Storage)*. Setting $\mu_K = 0$ in (A.18): $q_i^{\text{on}} = q_i^{\text{off}}$, i.e., $(q_i - s_i)/n = s_i/(T - n)$. Cross-multiplying: $(q_i - s_i)(T - n) = s_i n$, hence $q_i(T - n) = s_i T$, giving:

$$s_i = q_i(1 - \tau) \quad (\text{A.21})$$

Step 2 (Consumption). Setting $\mu_\phi = 0$ in (A.15) and substituting $q_i - s_i = q_i \tau$:

$$\theta_i = b \frac{q_i \tau}{n} = \frac{b q_i}{T} \implies \boxed{q_i = \frac{T \theta_i}{b} \equiv q_i^*} \quad (\text{A.22})$$

Step 3 (Verification). Interior when $q_i^* \leq \phi n$ and $q_i^*(1 - \tau) \leq K_i$. Both rates equal q_i^*/T : consumption is perfectly smoothed.

Case 2: Storage-constrained ($\mu_\phi = 0, \mu_K > 0$). The tank binds: $s_i = K_i$.

Step 1 (Consumption). With $\mu_\phi = 0$ in (A.15):

$$\theta_i = b \frac{q_i - K_i}{n} \implies q_i - K_i = \frac{n \theta_i}{b} = q_i^* \tau \implies \boxed{q_i = q_i^* \tau + K_i} \quad (\text{A.23})$$

On-day rate: $q_i^{\text{on}} = q_i^*/T$ (unconstrained). Off-day rate: $q_i^{\text{off}} = K_i/(T - n)$ (constrained).

Step 2 (Shadow price). From (A.18):

$$\mu_K = b \left[\frac{q_i^*}{T} - \frac{K_i}{T - n} \right] = \frac{b}{T} (q_i^* - \bar{q}_i^K) > 0 \quad (\text{A.24})$$

where $\bar{q}_i^K \equiv K_i/(1 - \tau)$ is the smoothing threshold. Positive when $q_i^*(1 - \tau) > K_i$.

Step 3 (Verification). Pipe slack requires $q_i^* \tau + K_i \leq \phi n$.

Case 3: Pipe-constrained ($\mu_\phi > 0, \mu_K = 0$). The pipe binds: $q_i = \phi n$.

Step 1 (Storage). $\mu_K = 0$ gives $q_i^{\text{on}} = q_i^{\text{off}}$. Substituting $q_i = \phi n$: $(\phi n - s_i)/n = s_i/(T - n)$, hence $s_i = \phi n(1 - \tau)$.

Step 2 (Rates). Both rates equal $\phi\tau$. Total consumption:

$$\boxed{q_i = \phi n} \quad (\text{A.25})$$

Step 3 (Shadow price). From (A.17):

$$\mu_\phi = \theta_i - b\phi\tau > 0 \quad (\text{A.26})$$

Positive when $q_i^* > \phi n$, since $\theta_i = b q_i^*/T > b\phi n/T = b\phi\tau$.

Step 4 (Verification). Tank slack requires $\phi n(1 - \tau) \leq K_i$.

Case 4: Both bind ($\mu_\phi > 0, \mu_K > 0$). $q_i = \phi n$ and $s_i = K_i$.

Step 1 (Rates). $q_i^{\text{on}} = (\phi n - K_i)/n$ and $q_i^{\text{off}} = K_i/(T - n)$. Since $\phi n(1 - \tau) > K_i$: $q_i^{\text{on}} > q_i^{\text{off}}$.

$$\boxed{q_i = \phi n} \quad (\text{A.27})$$

Step 2 (Shadow prices).

$$\mu_\phi = \theta_i - b \frac{\phi n - K_i}{n} > 0 \quad (\text{A.28})$$

$$\mu_K = b \left[\frac{\phi n - K_i}{n} - \frac{K_i}{T - n} \right] > 0 \quad (\text{A.29})$$

Step 3 (Positivity of μ_ϕ). The on-day rate $(\phi n - K_i)/n < \phi\tau < q_i^*/T$, so $\theta_i - b(\phi n - K_i)/n > \theta_i - b q_i^*/T = 0$.

Summary. The four cases are mutually exclusive and exhaustive, determined by whether latent demand q_i^* exceeds the pipe ceiling ϕn and whether it exceeds the smoothing threshold $\bar{q}_i^K \equiv K_i/(1 - \tau)$.

Table (A.11) Household Classification

Case	Condition	q_i	Smoothed?	Shadow Price
1. Unconstrained	$q_i^* \leq \phi n, q_i^*(1 - \tau) \leq K_i$	q_i^*	Yes	$\mu_\phi = \mu_K = 0$
2. Storage-constrained	$q_i^*(1 - \tau) > K_i, q_i^*\tau + K_i \leq \phi n$	$q_i^*\tau + K_i$	No	$\mu_K > 0$
3. Pipe-constrained	$q_i^* > \phi n, \phi n(1 - \tau) \leq K_i$	ϕn	Yes	$\mu_\phi > 0$
4. Both bind	$q_i^* > \phi n, \phi n(1 - \tau) > K_i$	ϕn	No	$\mu_\phi, \mu_K > 0$

Derivation of Marginal Value of an On-Day

Proof. I will assume in this proof that the instantaneous utility u_i is quadratic $u_i(q) = (a_i + \omega_i)q - \frac{b}{2}q^2$ where q is the daily consumption rate.

For a given realization ω_i , the household's second-stage indirect utility is $W_i(K_i, \omega_i, n) \equiv \max_{q_i, s_i} \mathcal{L}_i$, where $\mathcal{L}_i$ is (A.14) and n enters explicitly through the sub-period lengths and the pipe constraint ϕn . By the envelope theorem, $\partial W_i / \partial n = \partial \mathcal{L}_i / \partial n|_{q_i^*, s_i^*}$.

The Lagrangian, written in rate form for clarity:

$$\mathcal{L}_i = n u_i(q_i^{\text{on}}) + (T - n) u_i(q_i^{\text{off}}) - p q_i + \mu_\phi(\phi n - q_i) + \mu_K(K_i - s_i) \quad (\text{A.30})$$

Differentiating w.r.t. n (holding q_i and s_i fixed, so $q_i^{\text{on}} = (q_i - s_i)/n$ and $q_i^{\text{off}} = s_i/(T - n)$ both depend on n):

$$\frac{\partial \mathcal{L}_i}{\partial n} = u_i(q_i^{\text{on}}) + n u_i'(q_i^{\text{on}}) \cdot \frac{\partial q_i^{\text{on}}}{\partial n} - u_i(q_i^{\text{off}}) + (T - n) u_i'(q_i^{\text{off}}) \cdot \frac{\partial q_i^{\text{off}}}{\partial n} + \mu_\phi \phi \quad (\text{A.31})$$

The rate derivatives are:

$$\frac{\partial q_i^{\text{on}}}{\partial n} = -\frac{q_i - s_i}{n^2} = -\frac{q_i^{\text{on}}}{n}, \quad \frac{\partial q_i^{\text{off}}}{\partial n} = \frac{s_i}{(T - n)^2} = \frac{q_i^{\text{off}}}{T - n} \quad (\text{A.32})$$

Substituting:

$$\frac{\partial W_i}{\partial n} = u_i(q_i^{\text{on}}) - u_i'(q_i^{\text{on}}) q_i^{\text{on}} - \left[u_i(q_i^{\text{off}}) - u_i'(q_i^{\text{off}}) q_i^{\text{off}} \right] + \mu_\phi \phi \quad (\text{A.33})$$

For the quadratic utility $u(q) = (a_i + \omega_i)q - \frac{b}{2}q^2$:

$$u(q) - u'(q)q = (a_i + \omega_i)q - \frac{b}{2}q^2 - [(a_i + \omega_i) - bq]q = \frac{b}{2}q^2 \quad (\text{A.34})$$

Substituting into (A.33), the marginal value of an on-day *conditional on a single realization* ω_i is:

$$\frac{\partial W_i}{\partial n} = \underbrace{\frac{b}{2}[(q_i^{\text{on}})^2 - (q_i^{\text{off}})^2]}_{L_i(\omega) \text{ (smoothing dividend)}} + \underbrace{\mu_\phi(\omega) \phi}_{\text{quantity expansion}} \quad (\text{A.35})$$

Since the annualized storage cost $\Gamma_i(K_i)$ does not depend on n directly, the ex ante value function $V_i(K_i) = \mathbb{E}_\omega[W_i(K_i, \omega, n)] - \Gamma_i(K_i)$ satisfies:

$$\pi_i \equiv \frac{dV_i}{dn} \Big|_{K_i} = \mathbb{E}_\omega \left[\frac{\partial W_i}{\partial n} \right] = \mathbb{E}_\omega [L_i(\omega) + \mu_{\phi,i}(\omega) \phi] \quad (\text{A.36})$$

which is equation (2.8) in the main text. The expectation integrates over the regime-specific expressions: $L_i(\omega) = 0$ and $\mu_{\phi,i}(\omega) = 0$ in Case 1; $L_i > 0$ and $\mu_{\phi,i} = 0$ in Case 2; $L_i = 0$ and $\mu_{\phi,i} > 0$ in Case 3; and both positive in Case 4. $\square$

Price and Duration Responses by Case

I record the derivatives $\partial q_i/\partial p$ and $dq_i/dn|_{K_i}$ that enter the planner's FOCs.

Price responses. Using $q_i^* = T(a_i + \omega_i - p)/b$:

Case 1 ($q_i = q_i^*$): $\partial q_i/\partial p = \partial q_i^*/\partial p = -T/b$.

Case 2 ($q_i = q_i^*\tau + K_i$): $\partial q_i/\partial p = \tau \partial q_i^*/\partial p = -\tau T/b = -n/b$.

Cases 3 & 4 ($q_i = \phi n$): $\partial q_i/\partial p = 0$.

Duration responses (holding K_i fixed). Case 1 ($q_i = q_i^*$, independent of n): $dq_i/dn = 0$.

Case 2 ($q_i = q_i^*\tau + K_i = q_i^*n/T + K_i$): $dq_i/dn = q_i^*/T > 0$.

Cases 3 & 4 ($q_i = \phi n$): $dq_i/dn = \phi > 0$.

$$\left. \frac{dq_i}{dn} \right|_{K_i} = \begin{cases} 0 & i \in \Omega_1 \\ q_i^*/T & i \in \Omega_K \\ \phi & i \in \Omega_\phi \cup \Omega_B \end{cases} \quad (\text{A.37})$$

We then connect this to the RDD as follows:

	$\left. \frac{\partial q_i}{\partial n} \right _{K_i}$	$\frac{\partial q_i}{\partial K_i}$	$\frac{dK_i^*}{dn}$	$\frac{dq_i}{dn}$
Case 1 (Unconstrained)	0	0	< 0	0
Case 2 (Tank binds)	q_i^*/T	1	< 0	$q_i^*/T + dK_i^*/dn$
Cases 3 & 4 (Pipe binds)	ϕ	0	< 0	ϕ

The key observation is that the indirect channel operates *only* through Case 2 households. In Case 2, $q_i = q_i^*\tau + K_i$, so an additional unit of storage capacity translates one-for-one into additional consumption ($\partial q_i/\partial K_i = 1$). Since $dK_i^*/dn < 0$ (households with more frequent supply shed storage), this indirect effect is negative: storage adaptation *partially offsets* the direct consumption gain from additional supply days. In Cases 3 and 4, consumption is pinned at ϕn regardless of K_i , so storage adaptation has no effect on observed consumption. In Case 1, neither channel is active.

Proof of Theorem 1: First-Best Allocation

Proof. The planner maximizes the weighted welfare function:

$$\mathcal{W} = \sum_{i=1}^N g_i V_i + (p - \Lambda) Q - F \quad (\text{A.38})$$

where $V_i = \mathbb{E}[\text{GU}_i - p q_i] - \Gamma_i(K_i^*)$, $\Lambda = (c + \lambda) \iota(n)$, and $Q = \sum_i \mathbb{E}[q_i]$.

Part (i): FOC with respect to p .

Differentiating with respect to p :

$$\frac{\partial \mathcal{W}}{\partial p} = \sum_i g_i \frac{\partial V_i}{\partial p} + Q + (p - \Lambda) \frac{\partial Q}{\partial p} = 0 \quad (\text{A.39})$$

Roy's identity gives $\partial V_i / \partial p = -\mathbb{E}[q_i]$, so:

$$\sum_i g_i \frac{\partial V_i}{\partial p} = - \sum_i g_i \mathbb{E}[q_i] = -\bar{g}_q Q \quad (\text{A.40})$$

where $\bar{g}_q \equiv \sum_i g_i \mathbb{E}[q_i] / Q$ is the consumption-weighted welfare weight.

Substituting into (A.39):

$$-\bar{g}_q Q + Q + (p - \Lambda) \frac{\partial Q}{\partial p} = 0 \quad (\text{A.41})$$

$$(1 - \bar{g}_q) Q + (p - \Lambda) \frac{\partial Q}{\partial p} = 0 \quad (\text{A.42})$$

Rearranging (A.42):

$$(p - \Lambda) \frac{\partial Q}{\partial p} = (\bar{g}_q - 1) Q \quad (\text{A.43})$$

Substituting the elasticity $\varepsilon_{Q,p} \equiv (p/Q)(\partial Q / \partial p)$, i.e., $\partial Q / \partial p = \varepsilon_{Q,p} Q / p$:

$$(p - \Lambda) \frac{\varepsilon_{Q,p}}{p} = \bar{g}_q - 1 \quad (\text{A.44})$$

$$p \varepsilon_{Q,p} - \Lambda \varepsilon_{Q,p} = (\bar{g}_q - 1) p \quad (\text{A.45})$$

$$p [\varepsilon_{Q,p} - (\bar{g}_q - 1)] = \Lambda \varepsilon_{Q,p} \quad (\text{A.46})$$

$$p = \frac{\Lambda \varepsilon_{Q,p}}{\varepsilon_{Q,p} - (\bar{g}_q - 1)} \quad (\text{A.47})$$

Since $\varepsilon_{Q,p} < 0$, write $\varepsilon_{Q,p} = -|\varepsilon_{Q,p}|$:

$$p^* = \frac{\Lambda^* |\varepsilon_{Q,p}|}{|\varepsilon_{Q,p}| + \bar{g}_q - 1} = \frac{\Lambda^*}{1 + \frac{\bar{g}_q - 1}{|\varepsilon_{Q,p}|}} \quad (\text{A.48})$$

Part (ii): FOC with respect to n .

Taking the total derivative of SWF with respect to n :

$$\frac{d\mathcal{W}}{dn} = \underbrace{\sum_i g_i \frac{dV_i}{dn}}_{(A)} + \underbrace{(p - \Lambda) \frac{dQ}{dn} \Big|_{K_i}}_{(B)} - \underbrace{(c + \lambda) \iota'(n) Q}_{(C)} = 0 \quad (\text{A.49})$$

Term (A): Weighted household envelope. By the household envelope theorem ($\partial V_i / \partial K_i \Big|_{K_i^*} = 0$ at the private optimum):

$$\frac{dV_i}{dn} = \frac{\partial V_i}{\partial n} \Big|_{K_i} + \underbrace{\frac{\partial V_i}{\partial K_i} \Big|_{K_i^*}}_{=0} \cdot \frac{dK_i^*}{dn} = \pi_i \quad (\text{A.50})$$

Hence:

$$(A) = \sum_i g_i \pi_i \quad (\text{A.51})$$

Terms (B) and (C): Unchanged. (B) = $(p - \Lambda) \sum_i \mathbb{E}[dq_i/dn|_{K_i}]$ and (C) = $(c + \lambda) \iota'(n) Q$.

Applying the optimal price. At $p^* = \Lambda^* / (1 + (\bar{g}_q - 1)/|\varepsilon_{Q,p}|)$, the term $(p^* - \Lambda^*)$ is generally nonzero. The full FOC is:

$$\boxed{\sum_{i=1}^N g_i \pi_i(n^*) + (p^* - \Lambda^*) \sum_{i=1}^N \mathbb{E} \left[\frac{dq_i}{dn} \right] = (c + \lambda) \iota'(n^*) Q} \quad (\text{A.52})$$

Note: When $\bar{g}_q = 1$, $p^* = \Lambda^*$ and $(p^* - \Lambda^*) = 0$, recovering the unweighted result. When $\bar{g}_q \neq 1$, the fiscal wedge term $(p^* - \Lambda^*)$ is nonzero and contributes to the supply duration FOC even in the First-Best. $\square$

Proof of Theorem 2 (Second-Best Allocation)

Proof. The planner maximizes $\mathcal{W}$ over $(\bar{q}, n)$ at fixed price $\bar{p}$, using:

$$\mathcal{W} = \sum_i g_i V_i + (\bar{p} - \Lambda) Q - F \quad (\text{A.53})$$

where $\Delta_p \equiv \Lambda - \bar{p} > 0$.

Part (i): Optimal Quota.

Differentiating with respect to $\bar{q}$, only quota-constrained households respond. For $i \in \Omega_{\bar{q}}$, consumption $q_i = \bar{q}$ (smoothed at $\bar{q}/T$), so $\partial V_i / \partial \bar{q} = \mathbb{E}[u'_i(\bar{q}/T)] - \bar{p}$ and $\partial \mathbb{E}[q_i] / \partial \bar{q} = 1$.

From the full welfare function:

$$\frac{\partial \mathcal{W}}{\partial \bar{q}} = \sum_{i \in \Omega_{\bar{q}}} \left[g_i \mathbb{E}[u'_i(\bar{q}/T)] + (1 - g_i) \bar{p} - \Lambda \right] = 0 \quad (\text{A.54})$$

Rewriting by collecting g_i terms:

$$\begin{aligned} &= \sum_{i \in \Omega_{\bar{q}}} g_i \left[\mathbb{E}[u'_i(\bar{q}/T)] - \bar{p} \right] + \sum_{i \in \Omega_{\bar{q}}} (\bar{p} - \Lambda) = 0 \\ &= \sum_{i \in \Omega_{\bar{q}}} g_i \left[\mathbb{E}[u'_i(\bar{q}/T)] - \bar{p} \right] - \Delta_p |\Omega_{\bar{q}}| = 0 \end{aligned} \quad (\text{A.55})$$

Solving:

$$\boxed{\frac{1}{|\Omega_{\bar{q}}|} \sum_{i \in \Omega_{\bar{q}}} g_i \mathbb{E}[u'_i(\bar{q}^*/T) - \bar{p}] = \Delta_p} \quad (\text{A.56})$$

Utilitarian case ($g_i = 1$ for all i):

$$\frac{1}{|\Omega_{\bar{q}}|} \sum_{i \in \Omega_{\bar{q}}} \mathbb{E}[u'_i(\bar{q}^*/T)] = \Lambda^{SB} \quad (\text{A.57})$$

Part (ii): Optimal Supply Duration.

The total derivative of $\mathcal{W}$ with respect to n :

$$\frac{d\mathcal{W}}{dn} = \underbrace{\frac{\partial \mathcal{W}}{\partial n} \Big|_{K_i}}_{\text{Direct}} + \underbrace{\sum_{i=1}^N \frac{\partial \mathcal{W}}{\partial K_i} \cdot \frac{dK_i^*}{dn}}_{\text{Indirect}} = 0 \quad (\text{A.58})$$

Direct effect.

$$\left. \frac{\partial \mathcal{W}}{\partial n} \right|_{K_i} = \underbrace{\sum_i g_i \pi_i}_{(a)} - \underbrace{\Delta_p \sum_i \mathbb{E} \left[\left. \frac{dq_i}{dn} \right|_{K_i} \right]}_{(b)} \underbrace{-(c + \lambda) \iota'(n) Q}_{(c)} \quad (\text{A.59})$$

Indirect effect (adaptation correction). The social marginal value of storage at the household's private optimum:

$$\begin{aligned} \left. \frac{\partial \mathcal{W}}{\partial K_i} \right|_{K_i^*} &= g_i \underbrace{\left. \frac{\partial V_i}{\partial K_i} \right|_{K_i^*}}_{= 0 \text{ (private FOC)}} + (\bar{p} - \Lambda) \mathbb{E} \left[\left. \frac{\partial q_i}{\partial K_i} \right] \right. \\ &= -\Delta_p \cdot \mathbb{P}(i \in \Omega_K) \end{aligned} \quad (\text{A.60})$$

Summing:

$$\mathcal{S}(n) \equiv -\Delta_p \sum_{i=1}^N \mathbb{P}(i \in \Omega_K) \cdot \frac{dK_i^*}{dn} > 0 \quad (\text{A.61})$$

Combining.

$$\boxed{\underbrace{\sum_{i=1}^N g_i \pi_i(n^{SB}) + \mathcal{S}(n^{SB})}_{\text{Effective Marginal Benefit}} = \underbrace{(c + \lambda) \iota'(n^{SB}) Q}_{\text{Leakage Cost}} + \underbrace{\Delta_p \sum_{i=1}^N \mathbb{E} \left[\left. \frac{dq_i}{dn} \right|_{K_i} \right]}_{\text{Fiscal Wedge}}} \quad (\text{A.62})$$

□

Proof of Proposition 4 (Third-Best Allocation)

Proof. The Third-Best is the Second-Best with $\bar{q} \rightarrow \infty$. The FOC structure is identical; only the constrained sets expand.

FOC. With no quota, high-demand households that would have been capped at $\bar{q}^*$ now belong to the physically constrained sets. The weighted supply duration FOC is:

$$\sum_{i=1}^N g_i \pi_i(n^{TB}) + \mathcal{S}(n^{TB}) = (c + \lambda) \iota'(n^{TB}) Q + \Delta_p \sum_{i=1}^N \mathbb{E} \left[\frac{dq_i}{dn} \Big|_{K_i} \right] \quad (\text{A.63})$$

Fiscal wedge comparison. Since $\Omega_K^{SB} \subseteq \Omega_K^{TB}$ and $\Omega_\phi^{SB} \cup \Omega_B^{SB} \subseteq \Omega_\phi^{TB} \cup \Omega_B^{TB}$, the fiscal wedge is strictly larger in the Third-Best:

$$\text{FD}^{TB} - \text{FD}^{SB} = \Delta_p \sum_{i=1}^N \left(\mathbb{E} \left[\frac{q_i^*}{T} \cdot \mathbb{1}\{i \in \Omega_{\bar{q}} \cap \Omega_K^{TB}\} \right] + \phi \cdot \mathbb{P}(i \in \Omega_{\bar{q}} \cap (\Omega_\phi^{TB} \cup \Omega_B^{TB})) \right) > 0 \quad (\text{A.64})$$

Duration ordering. The effective marginal benefit $\text{MB}(n) = \sum_i g_i \pi_i(n) + \mathcal{S}(n)$ is decreasing in n . At any common n , $\text{MC}^{TB}(n) > \text{MC}^{SB}(n) > \text{MC}^{FB}(n)$, implying $n^* > n^{SB} > n^{TB}$. $\square$

Proof of Theorem 3 (Welfare Change)

Proof. Recall the planner's objective function from Equation 2.12:

$$\mathcal{W}(\Theta) = \sum_{i=1}^N g_i V_i + (p - \Lambda)Q - F \quad (\text{A.65})$$

where $Q = \sum_{i=1}^N \mathbb{E}[q_i]$ is aggregate expected consumption and $V_i = \mathbb{E}[\text{GU}_i - p q_i] - \Gamma_i(K_i^*)$ is consumer surplus net of storage costs.

Substituting V_i into the objective and grouping terms into four distinct economic components—weighted consumer surplus ($\text{CS}_i \equiv \text{GU}_i - p q_i$), government fiscal balance, storage costs, and fixed infrastructure costs:

$$\mathcal{W}(\Theta) = \underbrace{\sum_{i=1}^N g_i \mathbb{E}[\text{CS}_i]}_{\text{Weighted CS}} + \underbrace{(pQ - \Lambda Q)}_{\text{Fiscal Balance}} - \underbrace{\sum_{i=1}^N g_i \Gamma_i(K_i^*)}_{\text{Weighted Storage Costs}} - \underbrace{F}_{\text{Fixed Costs}} \quad (\text{A.66})$$

For any counterfactual regime $r \in \{TB, SB, *\}$ with policy $\Theta^r = (p^r, \bar{q}^r, n^r)$, the aggregate welfare change $\Delta \mathcal{W}^{SQ \rightarrow r} = \mathcal{W}^r - \mathcal{W}^{SQ}$ is the term-by-term difference:

$$\begin{aligned} \Delta \mathcal{W}^{SQ \rightarrow r} = & \underbrace{\sum_{i=1}^N g_i \mathbb{E}[\Delta \text{CS}_i^r]}_{\text{(I) Change in Weighted CS}} + \underbrace{\left[(p^r Q^r - \bar{p} Q^{SQ}) - (\Lambda^r Q^r - \Lambda^{SQ} Q^{SQ}) \right]}_{\text{(II) Fiscal Change}} \\ & - \underbrace{\sum_{i=1}^N g_i [\Gamma_i(K_i^r) - \Gamma_i(K_i^{SQ})]}_{\text{(III) Weighted Storage Cost Change}} - \underbrace{(F^r - F^{SQ})}_{\text{(IV) Infrastructure Cost Change}} \end{aligned} \quad (\text{A.67})$$

where $\Delta \text{CS}_i^r \equiv (\text{GU}_i^r - p^r q_i^r) - (\text{GU}_i^{SQ} - \bar{p} q_i^{SQ})$. This gives (2.21).

The remainder of the proof evaluates term (I) for each regime by decomposing the change in consumer surplus into its underlying price, quantity, and duration effects. Terms (II)–(IV) evaluate directly from the equilibrium objects reported in Table 2.5.

Term (I) for the Third-Best ($p^{TB} = \bar{p}, \bar{q}^{TB} = \infty$). The only instrument that changes is supply duration n . By the envelope theorem (Definition 3), the marginal change in consumer surplus with respect to n is π_i , so for small $\Delta n \equiv n^{TB} - n^{SQ}$:

$$\Delta \text{CS}_i^{TB} \approx \pi_i(n^{SQ}) \cdot (n^{TB} - n^{SQ}) \quad (\text{A.68})$$

Because $\pi_i \geq 0$ and $n^{TB} \geq n^{SQ}$ (from Proposition 4), this is a weak consumer-surplus gain. Empirically, $n^{TB} = n^{SQ}$ so this effect is zero.

Term (I) for the Second-Best ($p^{SB} = \bar{p}$, $\bar{q}^{SB}$ binds for high-demand households). I partition the population into quota-constrained ($\Omega_{\bar{q}}$) and unconstrained ($\notin \Omega_{\bar{q}}$) households.

Quota-constrained households. The forced reduction in consumption $\Delta q_i \equiv q_i^{SQ} - \bar{q}^{SB} > 0$ generates a private utility loss. I calculate it using a second-order Taylor expansion around q_i^{SQ} ; because the utility is strictly quadratic, the expansion is exact:

$$GU_i(q_i^{SQ}) - GU_i(\bar{q}^{SB}) = \bar{p} \Delta q_i + \frac{b}{2T} (\Delta q_i)^2 \quad (\text{A.69})$$

using $GU'_i(q_i^{SQ}) = \bar{p}$ (from the household's FOC under the subsidized price) and $GU''_i(q) = -b/T$ (from the quadratic utility). The corresponding private consumer surplus change is:

$$\Delta CS_i^{SB} = -\frac{b}{2T} (\Delta q_i)^2 \quad (\text{A.70})$$

since the Harberger triangle is the only welfare effect once the transfer $\bar{p} \Delta q_i$ is netted against the change in expenditure $\bar{p} q_i^{SB} - \bar{p} q_i^{SQ} = -\bar{p} \Delta q_i$.

Unconstrained households. For $i \notin \Omega_{\bar{q}}$, the welfare change depends only on the duration change:

$$\Delta CS_i^{SB} \approx \pi_i(n^{SQ}) \cdot (n^{SB} - n^{SQ}) \quad (\text{A.71})$$

Term (I) for the First-Best ($p^* = \Lambda^*$, $\bar{q}^* = \infty$). The transition involves simultaneous changes in price and supply duration. By Roy's identity, $\partial CS_i / \partial p = -q_i$; by the envelope theorem, $\partial CS_i / \partial n = \pi_i$. A second-order Taylor expansion in p (exact for linear demand) combined with a first-order expansion in n yields:

$$\mathbb{E}[\Delta CS_i^*] \approx \underbrace{-\mathbb{E}[q_i^{SQ}](p^* - \bar{p}) + \frac{1}{2} \mathbb{E}\left[-\frac{\partial q_i}{\partial p}\right] (p^* - \bar{p})^2}_{\text{price effect (transfer + Harberger triangle)}} + \underbrace{\mathbb{E}[\pi_i(n^{SQ})] \cdot (n^* - n^{SQ})}_{\text{duration effect}} \quad (\text{A.72})$$

Because $p^* \gg \bar{p}$, the price effect is large and negative; weighted by g_i , it imposes the heavy consumer-surplus penalty that the fiscal gain in term (II) must overcome.

Signs of the Remaining Terms. Term (II) flips sign across regimes:

- *Third-Best/Second-Best* ($p^r = \bar{p}$): term (II) reduces to $\bar{p} \Delta Q - (\Lambda^r Q^r - \Lambda^{SQ} Q^{SQ})$, which is typically positive because the quota or supply-duration change reduces subsidized consumption and thereby reduces the fiscal gap.
- *First-Best* ($p^r = p^* \gg \bar{p}$): term (II) is strictly positive and large, reflecting the revenue windfall from market-clearing pricing.

Term (III) is typically positive (i.e., a welfare gain) because $K_i^r < K_i^{SQ}$ in all three reforms: the quota (Second-Best) and the extended duration (all reforms) both induce storage disinvestment

(Proposition 2), eliminating deadweight capital. Term (IV) varies: zero for the Third-Best and the First-Best on degraded pipes, positive for the Second-Best (smart metering cost) and the First-Best with rehabilitation (amortized capital).

Combining, the net welfare change $\Delta \mathcal{W}^{SQ \rightarrow r}$ is positive if and only if the fiscal gain in term (II) plus the storage-cost saving in term (III) outweigh the consumer-surplus loss in term (I) and the infrastructure cost in term (IV). $\square$

Derivation of the Scarcity Rent

Although the main model is static, the scarcity parameter ν is derived from the underlying dynamic optimization problem of a social planner managing an exhaustible resource (the aquifer). Consider a planner maximizing the discounted stream of social welfare over infinite time $t \in \{0, \dots, \infty\}$. Let S_t denote the remaining stock of the resource at the start of period t , and q_t denote extraction. The transition equation is $S_{t+1} = S_t - q_t$.

The planner's value function satisfies the Bellman equation:

$$V(S_t) = \max_{q_t} \{u(q_t) - cq_t + \beta V(S_{t+1})\} \quad (\text{A.73})$$

subject to $S_{t+1} = S_t - q_t$.

Let ν_t be the multiplier associated with the resource constraint. This ν_t represents the shadow value of an additional unit of stock (marginal user cost). The FOC is:

$$u'(q_t) - c - \beta V'(S_{t+1}) = 0 \quad (\text{A.74})$$

By the Envelope Theorem, the marginal value of the stock is the scarcity rent in the next period: $V'(S_{t+1}) = \nu_{t+1}$.² Thus:

$$u'(q_t) = c + \beta \nu_{t+1} \quad (\text{A.75})$$

In the current period equilibrium, define the scarcity rent $\nu \equiv \beta \nu_{t+1}$. This yields the pricing rule:

$$p_t = c + \nu \quad (\text{A.76})$$

This derivation shows that the efficient price in a static snapshot must include ν to internalize the intertemporal opportunity cost of depletion. The static model treats this ν as an exogenous parameter inherited from the dynamic path.

²Formally, let the Lagrangian for the optimization problem at $t + 1$ be

$$\mathcal{L}_{t+1} = u(q_{t+1}) - cq_{t+1} + \beta V(S_{t+2}) + \nu_{t+1}(S_{t+1} - q_{t+1} - S_{t+2})$$

The Envelope Theorem states that the total derivative of the value function with respect to the state variable S_{t+1} is equal to the partial derivative of the Lagrangian with respect to that variable, evaluated at the optimum. Since S_{t+1} appears only in the constraint term, we have $V'(S_{t+1}) = \frac{\partial \mathcal{L}_{t+1}}{\partial S_{t+1}} = \nu_{t+1}$. Intuitively, the marginal welfare gain from relaxing the stock constraint by one unit is exactly equal to the shadow price of that constraint.

A.4 More Details on the RDD

Data Construction

The analysis dataset is constructed by merging the billing universe (Section 4.4) with spatial distance calculations. For each geocoded household, I compute the shortest perpendicular distance to the nearest boundary separating zones with *different* supply intensities, excluding boundary segments between zones that share the same schedule. Each household is classified as being on the “high” side (more supply days) or “low” side (fewer supply days) of its nearest boundary. The signed running variable is:

$$d_i = \begin{cases} +\text{distance}_i & \text{if household } i \text{ is on the high-supply side} \\ -\text{distance}_i & \text{if household } i \text{ is on the low-supply side} \end{cases} \quad (\text{A.77})$$

I restrict the sample to households within 3,000 meters of a qualifying boundary, as more distant households are uninformative for the local comparison and distort bandwidth selection. To account for the fact that treatment intensity varies across boundaries, I define the supply-day gap at each border as $\Delta S_b = |S_{\text{high}} - S_{\text{low}}|$ and construct four gap bins: $(0, 1]$, $(1, 2]$, $(2, 3]$, and $3+$ days.

Weekly consumption is constructed as quarterly consumption divided by 13. The network withdrawal rate is defined as weekly consumption divided by the household’s supply days per week:

$$\text{Withdrawal Rate}_i = \frac{Q_i/13}{n_{d(i)}} \quad (\text{A.78})$$

where $n_{d(i)}$ is the number of supply days per week in household i ’s zone. Both the outcome (quarterly consumption, winsorized at the 99th percentile) and the withdrawal rate (winsorized at the 99th percentile) are trimmed to limit the influence of outliers on bandwidth selection.

Estimation Details

All RDD specifications follow Calonico et al. (2020) with a local linear polynomial ($p = 1$), triangular kernel weights, and coverage error rate–optimal bandwidth selection. Standard errors are clustered at the water zone level (approximately 400 clusters), which is the level at which supply schedules are assigned.

Every specification includes border-pair fixed effects and year-quarter fixed effects as covariates. Border-pair fixed effects ensure that the treatment effect is identified from within-boundary comparisons of households on opposite sides of the same border, absorbing any level differences across borders (e.g., altitude, neighborhood wealth). Year-quarter fixed effects absorb aggregate time trends in consumption.

The pooled specification estimates a single treatment effect across all qualifying boundaries. I also estimate separate RDD regressions within each supply-day gap bin to examine whether the consumption response scales with treatment intensity. The reported estimates are bias-corrected coefficients $\hat{\tau}_{\text{bc}}$ with robust confidence intervals.

Density Test for Manipulation

I implement the density test of Cattaneo et al. (2020) to detect manipulation of the running variable, which does not require prebinning the data. If households sort across boundaries to obtain more favorable supply schedules, the density of the running variable should exhibit a discontinuity at the cutoff.

The test evaluates whether the density $f(x)$ of the running variable X_i is continuous at the cutoff $c = 0$. The null hypothesis is:

$$H_0 : f_R(c) = f_L(c) \quad (\text{A.79})$$

where $f_R(c) = \lim_{x \rightarrow c^+} f(x)$ and $f_L(c) = \lim_{x \rightarrow c^-} f(x)$.

The density estimators are obtained as the slope coefficients from local polynomial regressions of the empirical distribution function on either side of the cutoff. Specifically, the coefficient vector $\hat{\beta}_{p,L}(h) = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p)'$ solves:

$$\hat{\beta}_{p,L}(h) = \arg \min_{\beta} \sum_{i=1}^n \mathbb{1}(X_i < c) \left\{ \hat{F}(X_i) - \sum_{j=0}^p \beta_j (X_i - c)^j \right\}^2 K_h(X_i - c) \quad (\text{A.80})$$

and similarly for $\hat{\beta}_{p,R}(h)$ using observations with $X_i \geq c$. The density estimators are $\hat{f}_{p,L}(h) = \hat{\beta}_1$ and $\hat{f}_{p,R}(h) = \hat{\beta}_1$ from the left and right regressions respectively, where $\hat{F}(X_i)$ denotes the leave-one-out empirical CDF, $K_h(u) = K(u/h)/h$ with $K(\cdot)$ a kernel function, and h is a positive bandwidth.

The manipulation test statistic is:

$$T_p(h) = \frac{\hat{f}_{p,R}(h) - \hat{f}_{p,L}(h)}{\hat{V}_p(h)} \stackrel{a}{\sim} N(0, 1) \quad (\text{A.81})$$

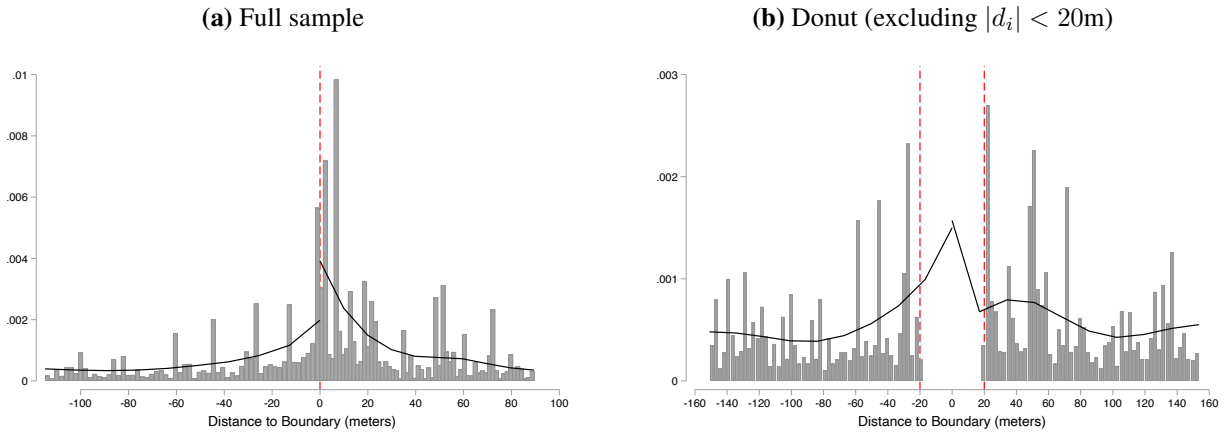
where $\hat{V}_p^2(h) = \hat{V}\{\hat{f}_{p,R}(h)\} + \hat{V}\{\hat{f}_{p,L}(h)\}$ pools the variance estimates from both sides.

Implementation details: I use local quadratic polynomials ($p = 2$) with robust bias correction following Calonico et al. (2014), which employs a higher-order polynomial ($q = p + 1 = 3$) to estimate and correct for the leading-order bias. The procedure selects MSE-optimal bandwidths separately for each side of the cutoff. Standard errors are calculated using the jackknife procedure.

Results. Figure A.7 displays the density test results. The test statistic is $T = 116.19$, strongly rejecting the null of no discontinuity. Inspection of the density plot reveals that the excess mass is concentrated within approximately 5 meters of the boundary on the high-supply side. Two explanations are consistent with this pattern. First, households may sort across boundaries to obtain more favorable supply schedules, which would violate the continuity assumption. Second, at very fine spatial scales, geocoding measurement error may systematically misassign households near the boundary, creating spurious bunching that does not reflect actual sorting.

To address this concern, I implement a donut specification that excludes all households within 20 meters of the boundary. As shown in the right panel of Figure A.7, the density imbalance is substantially reduced under the donut, consistent with either removing the marginal sorters or correcting for measurement error at fine spatial scales. All main RDD results are robust to this donut restriction.

Figure (A.7) Manipulation Density Test



Notes: Density estimates of the running variable (signed distance to the nearest supply-schedule boundary) using the Cattaneo et al. (2020) procedure with local quadratic polynomials ($p = 2$) and robust bias correction ($q = 3$). Panel (a): full sample. The test statistic is $T = 116.19$, rejecting continuity. The excess mass is concentrated within 5 meters of the cutoff. Panel (b): donut specification excluding households within 20 meters of the boundary. The discontinuity is substantially reduced.

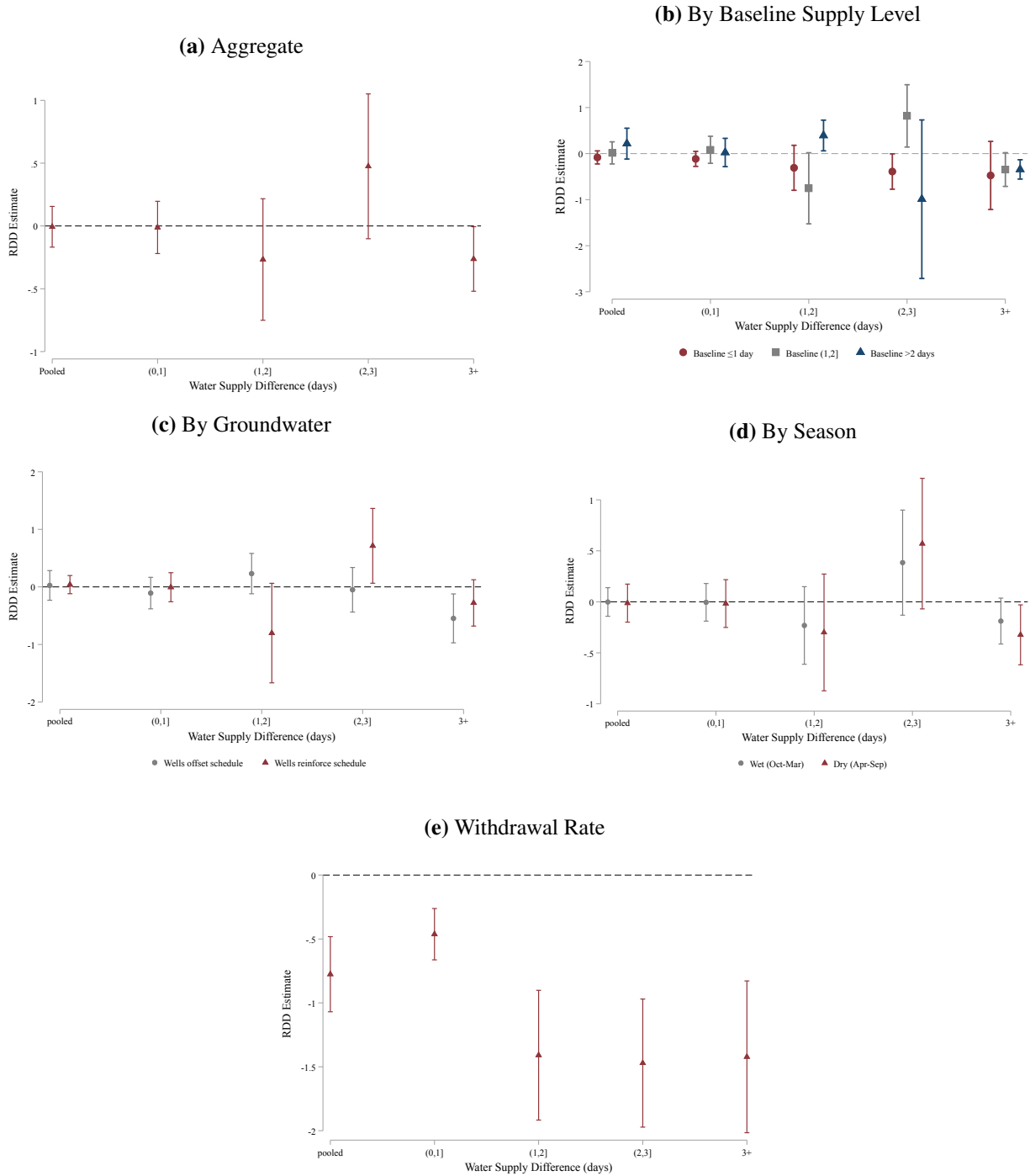
Network Withdrawal Rate

The withdrawal rate (weekly consumption divided by supply days) provides a diagnostic for the binding constraint regime. The structural model (Section 2.5) predicts distinct responses depending on which constraint is active:

Regime	Withdrawal Rate q_i/n	Predicted RDD on Rate
Case 1 (unconstrained)	q_i^*/n	Negative (denominator rises)
Case 2 (tank binds)	$(q_i^*\tau + K_i)/n$	Negative (tank is fixed)
Cases 3–4 (pipe binds)	ϕ (constant)	Zero (rate is mechanical)

A negative RDD coefficient on the withdrawal rate is consistent with Cases 1 and 2 dominating the boundary population: households on the high-supply side spread their consumption (which is

fixed or only modestly higher) over more days, reducing the rate. A zero coefficient would indicate that pipe-constrained households dominate. The results (Figure A.8, panel e) show weakly negative estimates, supporting the interpretation that most boundary households are not pipe-constrained, consistent with the large unconstrained share (77.4%) identified in the structural estimation.

Figure (A.8) Impact of Supply Intensity on Consumption (Donut)

Notes: Each point is a separate RDD estimate for water consumption per week with 95% CIs clustered at the water zone level. Observations within 20 meters from the boundary are excluded. Panels (a)–(c) group borders by the *gap* in supply days between adjacent zones. Panel (d) instead stratifies by the baseline (low-side) supply level, estimating gap-specific effects separately within each baseline bin. Panel (e) shows impacts on withdrawal rate, defined as consumption per supply day. “Dry”: April–September; “Wet”: October–March. “Reinforce”: the high-supply side also has higher local well production; “Offset”: the low-supply side does.

A.5 More Details on the Price Instrument

I use administrative records from Jordan’s primary water utility covering the universe of households over 20 quarters from 2020 Q1 to 2024 Q4. The data include quarterly water consumption (in cubic meters) and total bills (in Jordanian dinars). Average price is calculated as total bill divided by total consumption.

Sample Construction

The analysis sample is constructed through the following steps.

1. Start with all household-quarter observations (approximately $910,000 \text{ households} \times 20 \text{ quarters}$).
2. Winsorize consumption and bills at the 95th percentile to limit the influence of outliers.
3. Drop observations with missing consumption or bills.
4. Generate lagged variables: consumption twelve months prior ($Q_{i,t-4}$), and average price twelve months prior ($P_{i,t-4}$). This step loses the first 2–4 quarters per household.
5. Construct simulated instruments (described below), which requires both current-period and $t - 4$ tariff schedule estimates, losing an additional set of early quarters for which no $t - 4$ comparison period exists.
6. Create year-quarter $\times$ consumption-decile fixed effects by assigning each household to one of ten within-quarter quantile bins based on its lagged consumption $Q_{i,t-4}$.

The final estimation sample represents approximately 46% of the initial sample because the instrument requires observing each household at three time points: the current period (t), twelve months ago ($t - 4$), and twelve months ago ($t - 4$).

Instrument Construction

The instrument isolates exogenous variation in prices driven by policy changes in the tariff schedule between period $t - 4$ and period t , evaluated at a consumption level that is predetermined relative to the current demand shock. The construction proceeds in three steps.

Step 1: Estimate the price–quantity mapping within each quarter. For each quarter t in the sample, I estimate the relationship between average price and consumption using a cubic polynomial:

$$P_{it} = \gamma_0^{(t)} + \gamma_1^{(t)} Q_{it} + \gamma_2^{(t)} Q_{it}^2 + \gamma_3^{(t)} Q_{it}^3 + u_{it} \quad (\text{A.82})$$

where P_{it} is average price and Q_{it} is quarterly consumption. This regression is estimated separately within each quarter, so the coefficients $\gamma^{(t)}$ capture the quarter-specific tariff schedule mapping

consumption levels to average prices. I require at least 20 observations per quarter for estimation; quarters with fewer observations are dropped.

Step 2: Construct simulated prices at a predetermined consumption level. Using the estimated coefficients from (A.82), I predict each household's average price under two tariff schedules: the current schedule and the schedule from twelve months earlier, both evaluated at the same predetermined consumption level $Q_{i,t-4}$:

$$\widehat{AP}_t(Q_{i,t-4}) = \hat{\gamma}_0^{(t)} + \hat{\gamma}_1^{(t)}Q_{i,t-4} + \hat{\gamma}_2^{(t)}Q_{i,t-4}^2 + \hat{\gamma}_3^{(t)}Q_{i,t-4}^3 \quad (\text{A.83})$$

$$\widehat{AP}_{t-4}(Q_{i,t-4}) = \hat{\gamma}_0^{(t-4)} + \hat{\gamma}_1^{(t-4)}Q_{i,t-4} + \hat{\gamma}_2^{(t-4)}Q_{i,t-4}^2 + \hat{\gamma}_3^{(t-4)}Q_{i,t-4}^3 \quad (\text{A.84})$$

The key feature is that both predictions are evaluated at the *same* consumption level $Q_{i,t-4}$, which is predetermined relative to the current demand shock ε_{it} . The only source of variation between $\widehat{AP}_t$ and $\widehat{AP}_{t-4}$ is therefore the change in the tariff schedule coefficients across the two periods.

Step 3: Form the instrument. Following Ito (2014), I construct the simulated price instrument

$$Z_{it}^p = \widehat{AP}_t(Q_{i,t-4}) - \widehat{AP}_{t-4}(Q_{i,t-4}) \quad (\text{A.85})$$

where $\widehat{AP}_t(Q)$ is the predicted average price from a cubic polynomial of average price on quantity estimated within each quarter t , evaluated at household i 's consumption twelve months prior ($Q_{i,t-4}$). The instrument isolates variation in prices driven by policy changes in the tariff schedule (between $t-4$ and t), evaluated at a consumption level that is predetermined relative to the current demand shock ε_{it} . The twelve-month lag balances relevance against the exclusion restriction: it is recent enough to predict current tariff placement but sufficiently distant to break correlation with transitory demand shocks.

Estimation

All specifications are estimated in long differences to eliminate household fixed effects. The endogenous variable is the four-quarter difference in average price, $\Delta P_{it} \equiv P_{it} - P_{i,t-4}$, and the dependent variable is the corresponding difference in weekly consumption, $\Delta Q_{it} \equiv Q_{it} - Q_{i,t-4}$. The estimating equation is:

$$\Delta Q_{it} = \gamma \Delta P_{it} + \alpha_i + \lambda_{t,d} + \varepsilon_{it} \quad (\text{A.86})$$

where α_i are household fixed effects and $\lambda_{t,d}$ are year-quarter $\times$ consumption-decile fixed effects. The consumption deciles are constructed within each quarter based on lagged consumption $Q_{i,t-4}$, so they absorb any level differences across consumption groups within each period. Standard errors are clustered at the household level throughout.

I instrument ΔP_{it} with Z_{it}^p from equation (2.33). The first-stage regression is:

$$\Delta P_{it} = \pi_0 + \pi_1 Z_{it}^p + \alpha_i + \lambda_{t,d} + v_{it} \quad (\text{A.87})$$

The coefficient $\hat{\gamma}$ from the IV regression is the slope of demand in units of m^3/week per JD/m^3 . At sample means, the implied elasticity is $\hat{\gamma} \cdot \bar{p}/\bar{q} = -0.035$. Because the structural model implies linear demand (Assumption 4) with $\partial q/\partial p = -T/b$, the slope parameter b is recovered directly:

$$\hat{b} = -\frac{T}{\hat{\gamma}}.$$

This mapping identifies b in quantity units without reference to sample means.

Results

Table (A.12) Instrumental Variables Estimates of the Demand Slope

	(1) First Stage ΔP	(2) Reduced Form ΔQ	(3) IV ΔQ
Simulated $\Delta \hat{p}$ (JD/m^3)	0.9736*** (0.0160)	-0.1238*** (0.0083)	
Δp (JD/m^3)			-0.1272*** (0.0085)
Observations	8,642,055	8,642,055	8,642,055
R-squared	0.333	0.496	0.020

Notes: Each column reports IV estimates of the price elasticity of demand using predetermined consumption at a different lag to construct the simulated price instrument. Column (2) is the baseline ($t - 4$) specification. The endogenous regressor is the four-quarter change in average price. All specifications include household fixed effects and year-quarter $\times$ consumption-decile fixed effects. Standard errors in parentheses, clustered at the household level. *, **, ***: significance at 10%, 5%, 1%.

Table (A.13) Robustness to Alternative Lag Lengths for the Instrument

	(1) $t - 2$ (6 mo.)	(2) $t - 4$ (12 mo.)	(3) $t - 6$ (18 mo.)
Δp (JD/m ³)	-0.1582*** (0.0194)	-0.1272*** (0.0085)	-0.2173*** (0.0290)
Observations	8,418,465	8,642,055	7,293,316
First-stage F	600.83	3696.54	334.44
Implied $b = -T/\hat{\gamma}$	44.26	55.04	32.21

Notes: Each column reports IV estimates of the price elasticity of demand using predetermined consumption at a different lag to construct the simulated price instrument. Column (2) is the baseline ($t - 4$) specification. The endogenous regressor is the four-quarter log change in average price. All specifications include household fixed effects and year-quarter $\times$ consumption-decile fixed effects. Standard errors in parentheses, clustered at the household level. *, **, ***: significance at 10%, 5%, 1%.

Appendix B

Appendix to Housing Subsidies for Forced Migrants

B.1 Estimates Correcting for Non-Random Attrition

Table (B.1) Pooled Treatment Effects on Primary Outcomes (with Lee Bounds)

Outcome	TOT (se)	FDR q-value	Lower Bound	Upper Bound	Control (sd)	N	Rounds
Panel A: Primary Outcomes							
Overall Housing Quality (Z-Score)	0.31 (0.20)	[0.19]	0.21	0.71***	-0.01 (1.00)	4,218	1, 2, 3
Total Monthly Housing Expenditures (USD PPP)	-84.71*** (33.10)	[0.07]	-171.45***	-38.40	228.80 (208.80)	1,421	2
Food Consumption (Log USD PPP)	-0.02 (0.07)	[0.63]	-0.28***	0.36***	4.41 (0.72)	2,753	2, 3
Log Total Consumption (Log USD PPP)	-0.03 (0.08)	[0.63]	-0.26***	0.27***	8.73 (0.55)	1,422	2
CESD Score (Higher: Less Depression)	-0.10 (0.13)	[0.49]	-0.53***	0.45***	0.01 (1.00)	4,207	1, 2, 3
SDQ Score (Higher: Better Child Wellbeing)	-0.34** (0.15)	[0.07]	-0.55***	-0.04	-0.01 (1.00)	1,782	2, 3
Panel B: Other Outcomes							
At least 30 JD (95 USD PPP) in savings (=1)	0.08** (0.04)	[0.04]	-0.15***	0.11***	0.10 (0.30)	2,770	2, 3
Respondent Hunger Last Week (=1)	0.19*** (0.05)	[0.00]	0.02	0.33***	0.38 (0.49)	4,168	1, 2, 3
Adult Hunger Last Week (=1)	0.12*** (0.05)	[0.04]	-0.04	0.24***	0.36 (0.48)	4,099	1, 2, 3
Child Hunger Last Week (=1)	0.16*** (0.05)	[0.01]	-0.03	0.25***	0.25 (0.43)	3,741	1, 2, 3

Notes: This table reports 5 primary outcomes plus per capita food consumption (not pre-specified), pooled across as many rounds as were collected for each outcome. The table shows that treatment had no impact on housing quality on average over all 3 rounds, reduced housing expenditures at endline, had no impact on food consumption over endline and followup, had no impact on depression over all 3 rounds, and reduced child socio-emotional wellbeing over rounds 2 and 3. Overall Housing Quality is defined as a normalized housing quality index that includes indicators for quality floors, roofs, and walls, indicators for access to grid electricity and piped water, and the number of people per room. Total monthly housing expenditures is only reported for endline in the pooled table due to variations in measurement of housing expenditures across rounds and due to when treatment ended. Housing expenditure was measured including assistance payments at midline, and excluding assistance payments at endline since most recipients had stopped receiving payments but were still subject to rent freezes at part of the program. Results on housing expenditures in each round are reported in the appendix. Total consumption, including food and non-food expenses, was only measured at endline due to the length of the module. Reported in this table is food consumption, which was measured at endline and followup. Total food consumption is the log of the sum of food consumed in last seven days (Cereals and cereal products, Live animals, meat, and other parts of slaughtered land animals; Fish and other seafood; Milk, other dairy products, and eggs; Oils and fats; Fruits and nuts; Vegetables, tubers, pulses; Sugar and desserts; Ready-made food and other food products (baby food, spices)), plus home-produced foods produced in the last 7 days, plus assistance received in the last 30 days divided by 4.33. CES-D was measured in all rounds with exactly the same questionnaire. It includes 10 questions on depressive symptoms. Individuals who skipped more than 2 questions were marked as missing. The continuous scores were standardized with respect to the control group. The Child Strengths and Difficulties questionnaire (SDQ) was administered regarding a randomly selected child aged 3-8 at endline. At follow-up, the SDQ was administered regarding the same child. If not surveyed previously or if that child is no longer in the household, the survey was administered with respect to a randomly selected child age 4-9. The continuous scores were randomized with respect to the control group. At least 30 JD in savings (=1) is an indicator equal to one if the respondent answered yes to "Do you currently have at least 30 JDs in personal savings you can draw from in an emergency? Whether or not it is in a bank?". Respondent, adult, and child hunger last week are indicators equal to 1 if the respondent, other adults in the household or children in the household (respectively) went to bed hungry at least once in the last 7 days, zero otherwise. The lower and upper bounds refer to the ones proposed in Lee (2009) to correct for attrition. In round 3, there are only 70 treated households that answered "yes" to having at least 30 JD, but 84 were needed for the Lee bounds. To overcome this issue, all 70 who answered by "yes" are replaced by "no" in round 3 when finding the lower bound. Round 2 does not suffer from this problem, which is why we still have some variation to run the regression.

Table (B.2) Significant Impacts at Midline (with Lee Bounds)

Outcome	TOT	(se)	FDR q-values	Lower Bound	Upper Bound	Control mean	(sd)	N
Panel A: Housing Improvements								
Quality Roof (=1)	0.24***	(0.09)	[0.08]	0.23**	0.33***	0.71	(0.45)	1,610
Clean water (=1)	0.21***	(0.06)	[0.01]	0.07	0.25***	0.17	(0.38)	1,610
Occupied Rooms	0.51***	(0.17)	[0.04]	0.14	0.79***	2.85	(1.05)	1,610
Panel B: Food Aid and Insecurity								
Food consumption (aid) USD PPP	-53.05**	(22.40)	[0.02]	-72.09***	-46.97**	218.51	(201.68)	1,569
Food (aid) USD PPP (<i>Per capita</i>)	-9.73***	(3.44)	[0.01]	-13.03***	-8.73***	35.88	(29.19)	1,569
Respondent hunger last week (=1)	0.23***	(0.07)	[0.01]	0.12*	0.29***	0.35	(0.48)	1,604
Adults hunger last week (=1)	0.17***	(0.06)	[0.01]	0.07	0.24***	0.38	(0.49)	1,610
Child hunger last week (=1)	0.24***	(0.07)	[0.00]	0.03	0.30***	0.23	(0.42)	1,469
Panel C: Household Recomposition								
Number of children under 18 years	0.30**	(0.15)	[0.03]	-0.14	0.56***	3.27	(1.97)	1,610
Number of boys aged 13-17 years	0.16**	(0.07)	[0.33]	-0.20***	0.24***	0.40	(0.66)	1,610
Panel D: COVID-19 Outcomes								
Weekly adult income (during lockdown)	-20.07***	(7.19)	[0.02]	-52.48***	-17.86***	30.22	(66.86)	1,609
Total COVID-19 symptoms	0.84***	(0.19)	[0.00]	-0.03	0.94***	0.47	(1.22)	1,610
Number of people who are symptomatic	0.46***	(0.09)	[0.00]	0.03	0.52***	0.29	(0.61)	1,581
Number of children who are symptomatic	0.17***	(0.05)	[0.01]	-0.14***	0.19***	0.06	(0.32)	1,472
Number of visits to healthcare institutions	-0.56***	(0.23)	[0.04]	-1.66***	-0.40*	1.14	(2.03)	1,608
Did not keep distance (=1)	0.16***	(0.06)	[0.08]	0.09	0.26***	0.51	(0.50)	1,609
Panel E: Loans								
Loans taken pre-first-lockdown (=1)	-0.20***	(0.07)	[0.01]	-0.27***	-0.12*	0.61	(0.49)	1,608
Loans taken during first lockdown (=1)	-0.34***	(0.07)	[0.00]	-0.38***	-0.24***	0.78	(0.41)	1,608
Loans taken after first lockdown (=1)	-0.33***	(0.07)	[0.00]	-0.45***	-0.28***	0.42	(0.49)	1,607
Panel F: Other effects								
Total of school days attended per child	-0.43*	(0.24)	[0.04]	-0.79***	-0.22	2.33	(2.49)	1,031
Child alertness (s.d. units)	-0.73***	(0.20)	[0.00]	-0.50***	-1.45***	0.00	(1.00)	975
Difficulty with self-care (z-score)	0.38***	(0.14)	[0.08]	-0.43***	0.43***	-0.00	(1.00)	1,610
Number of learning activities	0.19**	(0.09)	[0.08]	-0.21***	0.27***	0.22	(0.56)	1,385

Notes: The table shows the regression results on statistically significant results from the midline survey. Each row is its own dependent variable. Panel A outcomes are Clean Water, defined as an indicator for households having treated drinking water (such as by a filter). Panel B outcomes are food consumption, Respondent (and adult) food insecurity equals 1 if the respondent went to bed hungry on at least one day in the last week and 0 otherwise. Panel D outcomes are COVID symptoms, which is the sum of indicators that increase the likelihood of COVID contraction by the respondent including leaving the house, attending social gatherings, not keeping distance from others, going to mosque or other religious institutions, going to grocery store or market, and leaving village/neighborhood. The other outcomes are as explained in the labels. Panel E The periods of the loans referred to are before, during, and after first lockdown. Panel F outcomes are Total number of school days attended per child is the total number of school days (from none to 5 days) attended by all school-aged children in ages 9-17, divided by the total number of children who are school-aged; Difficulty with self-care with higher values indicating more disability. This is an item of the The Washington Group Short Set on Functioning; learning activities are done in the last 24 hours and include homework, e-learning, educational programs and videos, and reading; Respondent networks corresponds to people (excl. household members) that the respondent knew who were already in their current neighborhood when they moved to their current residence. Monetary values and health visits are in USD PPP and are winsorized at the top 1% of values in order to limit the influence of outliers. The regressions also have month-by-year fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to various families in the pre-analysis plan. The lower and upper bounds refer to the ones proposed in Lee (2009) to correct for attrition.

Table (B.3) Significant Impacts at Endline and Follow-up (with Lee bounds)

Outcome	TOT	(se)	FDR q-values	Lower Bound	Upper Bound	Control mean	(sd)	N
Panel A: Endline								
Respondent Childcare & Chores (Hours)	-6.91***	(2.02)	[0.01]	-13.27***	-5.98***	17.76	(18.39)	1,418
At least 30 JD (95 \$ PPP) in savings (=1)	0.14***	(0.05)	[0.02]	-0.07	0.18***	0.12	(0.32)	1,422
Applied but Did Not Take Loan (=1)	0.06***	(0.02)	[0.00]	-0.02***	0.06***	0.01	(0.10)	1,422
Total Labor and Chores Hours Last Week	-10.62***	(3.24)	[0.01]	-22.95***	-6.61**	25.54	(23.39)	1,422
IP Assistance in Last 12 Months (=1)	0.08**	(0.03)	[0.14]	-0.11***	0.09***	0.04	(0.19)	1,422
IP Assistance in Last 12 Months (\$ PPP)	128.25***	(46.50)	[0.10]	-88.67***	145.45***	28.74	(198.87)	1,422
General Happiness (Scale 1-3)	-0.22***	(0.08)	[0.02]	-0.51***	-0.04	1.76	(0.61)	1,422
Panel B: Follow-up								
Child SDQ score, std	-0.56***	(0.17)	[0.01]	-0.72***	-0.33*	-0.01	(1.00)	859

Notes: The table shows the regression results on statistically significant results from the endline survey. Each row is its own dependent variable. Monetary values are in USD PPP and are winsorized at the top 1% of values in order to limit the influence of outliers. The regressions also have month-by-year fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to various families in the pre-analysis plan. The lower and upper bounds refer to the ones proposed in Lee (2009) to correct for attrition. The lower bounds marked by NA are not possible to report because trimming the lower tail would result in no variation in the variable, making the regression impossible to run as it becomes a column of zeros. The outcome "Focus Respondent Childcare & Chores" was pre-specified in two outcome families; the table reports the lower of the two q-values; the higher q-value is 0.02

Table (B.4) Impacts on Neighbor's Attitudes Toward Refugees (with Lee Bounds)

Outcome	TOT	(se)	FDR q-values	Lower Bound	Upper Bound	Control mean	(sd)	N
Panel A: Primary Outcomes								
Social attitudes & perceptions (SD)	-0.33**	(0.14)	[0.08]	-0.75***	-0.01	-0.00	(1.00)	1,102
Economic attitudes & percep. (SD)	0.02	(0.16)	[0.56]	-0.55***	0.22	0.02	(1.02)	1,102
Policy preferences (SD)	0.17	(0.15)	[0.37]	-0.28*	0.49***	-0.02	(1.00)	1,102
Altruism to Syrians	0.21	(0.18)	[0.37]	-0.49***	0.47**	0.88	(1.18)	1,102
Panel B: Selected Social Attitudes Index Components								
Of the 3 people you exchange advice with, how many are Syrian refugees?	-0.24**	(0.11)	[0.32]	-0.56***	-0.19*	0.34	(0.72)	1,102
Do the children in this household have any Syrian refugee friends?	-0.19*	(0.11)	[0.37]	-0.33***	-0.11	0.40	(0.49)	704
Panel C: Other Related Outcomes								
Media consumption (days last week)	-1.08***	(0.40)	[0.08]	-1.69***	-0.04	4.78	(2.78)	1,102
Neighbor perceptions of average refugee aid receipt (PPP)	-347.17***	(109.69)	[0.04]	-636.19***	-239.58**	630.84	(657.41)	685
Life Satisfaction (SD)	-0.32**	(0.15)	[0.06]	-0.62***	0.16	0.01	(0.98)	1,097
Subjective Health (SD)	-0.30*	(0.16)	[0.06]	-0.51***	0.25*	0.02	(0.99)	1,102

Notes: This table reports impacts of treatment on Jordanian neighbors of Syrian refugees in the experimental sample. Each row is its own dependent variable. Panel A reports pre-specified treatment effects on primary neighbor outcomes. All indices are standardized so that a positive point estimate reflects more “pro-refugee” sentiments. The only statistically significant outcome is a negative treatment effect on social attitudes and perceptions. This outcome is a standardized index of the following outcomes: out of 3 closest friends, how many are Syrian refugees; out of 3 people you exchange advice with, how many are Syrian refugees; do the children in this household have Syrian refugee friends (=1); do the children in this household share recreational spaces with Syrian refugee children (=1); how comfortable would you be accepting the marriage of a friend or loved one to a Syrian refugee (1-5); how comfortable would you be accepting a Syrian refugee as a neighbor (1-5); what is the net effect of Syrian refugees on Jordanian society (1-3); are Syrian refugees hardworking or lazy (1-7). Economic attitudes and perceptions is a standardized index of the following outcomes: an indicator =1 if the respondent listed “hosting Syrian refugees” as one of the most important challenges facing Jordan (options not read); an outcome with higher values for responses associated with beliefs that Syrian refugees pay more in taxes than Jordanians; an outcome with higher values for responses associated with refugees having a positive effect on the economy. Policy preferences in a standardized index of the following outcomes: less belief that Syrian refugees should be forced to live in camps; more belief that refugees should have the right to work outside camps; more belief that refugees should be able to attain Jordanian citizenship; more support for unrestricted work permits for refugees; more support for integrated classrooms with Syrians and Jordanians; more support for refugee right to enter/exit camps freely; support for housing assistance for Syrian refugees; belief that the international community should spend more money supporting refugees. Altruism to Syrians reports how many JD out of 5 the respondent allocated to a charity supporting Syrian refugees. The other two options were allocating to a charity supporting low-income Jordanians or keeping for self. Panel B reports the two significant subcomponents of the social attitudes and perceptions index. Panel C reports other statistically significant treatment effects on neighbors. Subjective health is the response to “Would you describe your general health as good, fair, poor, or very poor?” Life satisfaction is the response to “All things considered, how satisfied are you with your life as a whole these days on a scale of 1 to 10?” Days of media consumption is the response to “In the past 7 days, how many days did you read or listen to the news from any source, including newspapers, online, WhatsApp, etc.?” Neighborhood perceptions of refugee aid receipt is the response to a question asking “Of the refugee households in your neighborhood who receive assistance, what do you think is the average value in Dinar of the assistance (in cash or in kind) that they receive from organizations in a typical month?” The sample size for this question is smaller due to the large number of “don’t know” responses. There is no treatment effect on the probability of reporting “don’t know” to this question. The regressions include month-by-year fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to various families in the pre-analysis plan. The lower and upper bounds refer to those proposed in Lee (2009) to correct for attrition.

B.2 Heterogeneous Effects Analysis

Table (B.5) Gender Heterogeneity at Endline

	Market Mobility	Relatives Freedom	Safety Perceptions	Gender Equality
Treat	-0.33* (0.19)	-0.22 (0.17)	-0.11 (0.21)	-0.37** (0.19)
Female	-1.11*** (0.07)	-0.89*** (0.06)	-0.15** (0.07)	0.30*** (0.06)
Treat*Female	1.02*** (0.28)	0.89*** (0.27)	-0.12 (0.31)	0.47* (0.24)
p-val: T + T*female=0	0.00	0.00	0.32	0.53
Control mean	2.65	2.79	-0.00	0.00
Control sd	1.04	0.95	1.00	1.00
N	1,415	1,414	1,420	1,415

Notes: The table shows the regression results on mobility and gender outcomes. Each column is its own dependent variable. Market Mobility is defined as a z-score from the FR's response to whether they can go to the local market without permission, after informing someone, after being granted permission, or cannot go alone, with higher values indicating better mobility. Relatives Freedom is defined in the same manner as the market mobility outcome, but it asks about visiting the home of relatives, friends, or neighbors. Safety Perceptions is defined as a z-score from the FR's responses about how safe (very safe, safe, neither safe nor unsafe, unsafe, very unsafe) he or she thinks it is to walk outside during the day in the area where you live for women, with higher values indicating more safety. Gender Equality is defined as a normalized index (z-score) constructed from four items in which the FR strongly agrees, agrees, disagrees, or strongly disagrees. The four items are "A married woman can work outside the home if she wishes", "Husbands should have final say in all decisions concerning the family", "A woman can be a president or prime minister of a Muslim country", and "Women and men should have equal rights in making the decision to divorce." Each item is normalized and then the resulting index is normalized again where higher values indicate views in accordance with more gender equality. The regressions also have month-by-year fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%).

Table (B.6) Neighbor Impacts on Social Attitudes: Heterogeneity

	Social Attitudes			
Treat	-0.53*** (0.22)	-0.39*** (0.14)	-0.39*** (0.14)	-0.43** (0.20)
Above Median Marlow-Crowne	-0.01 (0.08)			
Treat*Above Median Marlow-Crowne	0.41 (0.30)			
Palestinian Grandparents		-0.03 (0.10)		
Treat*Palestinian Grandparents		0.54 (0.45)		
Non-Jordanian Grandparents			0.16 (0.11)	
Treat*Non-Jordanian Grandparents			0.35 (0.45)	
Above Median Distance				-0.07 (0.08)
Treat*Above Median Distance				0.39 (0.28)
p-val: T + T*Het = 0	0.52	0.73	0.93	0.85
Control Mean	-0.00	-0.00	-0.00	-0.00
Control SD	1.00	1.00	1.00	1.00
N	1,102	1,102	1,102	1,044

Notes: Each column reports a TOT regression of the social attitudes index on treatment fully interacted with one of 4 aspects of heterogeneity. Palestinian is an indicator equal to one if the Jordanian respondent reported one or more of their grandparents was born in Palestine, or answered yes to “Are you or your spouse of Palestinian descent?” Non-Jordanian grandparent is equal to one if the respondent indicated at least one of their grandparents or their spouse’s grandparents was not born in Jordan. Above median Marlow-Crowne is equal to one if their Marlow-Crowne score was above median, indicating a higher propensity to give socially desirable responses. Above median distance is equal to one if the neighbor was above median distance away from the Syrian refugee in the experiment. Median distance was 63 meters. Distance is measured in meters between the refugee house and the neighbor house.

B.3 Additional Results

Table (B.7) Impacts of the Program on Earnings, Labor, & Occupational Choice

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
Total Earnings in Last 30 Days (PPP, IHS)	-0.51	(0.38)	[0.33]	1.59	(2.83)	1,422
Total Labor Hours Last Week	-3.29	(2.35)	[0.33]	7.72	(18.73)	1,422
Total Labor Hours [Monthly Average]	-0.65	(8.20)	[0.78]	24.19	(67.80)	1,422
Net Wage-Employment Income Last Month (PPP, IHS)	-0.46	(0.37)	[0.40]	1.56	(2.81)	1,417

Notes: The table shows the regression results on pre-specified labor market outcomes of the Focus Respondent using the 2021 in-person data. Each row is its own dependent variable. The outcomes that require definitions are: Total Earnings is defined as the total of business profits and net wage salary in all jobs and is transformed using the inverse hyperbolic sine transformation. Total Labor Hours [Monthly Average] is defined as average weekly wage-employment hours multiplied by the numbers of months worked in last 12 months multiplied by 52/12 to obtain a monthly average of all jobs. The independent variable of interest is the TOT treatment indicator, which is the predicted value from a first-stage regression of treatment implementation on treatment assignment. Monetary values are in USD PPP. Total Earnings in Last 30 Days, Taxes Paid in Last 30 Days, Net Wage-Employment Income Last Month, Self-Employment Profits, Expenses, and Revenues Last Month and Year, Self-Employment Profits Last Year, Total Employees Last Month, Self-Employment Rent Last Month, and Total Labor Hours [Average] are winsorized at the top 1% of values in order to limit the influence of outliers. The regressions also have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family 4 in the pre-analysis plan.

Table (B.8) Robustness of Baseline Balance

Outcomes	Treatment	(se)	FDR q-values	Control mean	(sd)	N
(1) Overall Housing Quality (Z-Score)	0.05	(0.12)	[1.00]	-0.01	(1.00)	4,218
(2) Total Monthly Housing Expenditures (USD PPP)	6.82	(15.45)	[1.00]	228.80	(208.80)	1,422
(3) Food Consumption (Log USD PPP)	-0.01	(0.03)	[1.00]	4.41	(0.72)	4,218
(4) Log Total Consumption (Log USD PPP)	-0.02	(0.02)	[1.00]	8.73	(0.55)	4,218
(5) CESD Score (Higher: Less Depression)	-0.01	(0.05)	[1.00]	0.01	(1.00)	4,218
(6) SDQ Score (Higher: Better Child Wellbeing)	0.01	(0.04)	[1.00]	-0.01	(1.00)	4,218

Notes: The table shows the regression results on baseline balance for our primary outcomes. Each row is its own dependent variable. Overall Housing Quality is defined as a normalized housing quality index that includes indicators for quality floors, roofs, and walls, indicators for access to grid electricity and piped water, and the number of people per room. Total monthly housing expenditures is only reported for endline in the pooled table due to variations in measurement of housing expenditures across rounds and due to when treatment ended. Housing expenditure was measured including assistance payments at midline, and excluding assistance payments at endline since most recipients had stopped receiving payments but were still subject to rent freezes at part of the program. Results on housing expenditures in each round are reported in the appendix. Food consumption is the log of the sum of food consumed in last seven days (Cereals and cereal products, Live animals, meat, and other parts of slaughtered land animals; Fish and other seafood; Milk, other dairy products, and eggs; Oils and fats; Fruits and nuts; Vegetables, tubers, pulses; Sugar and desserts; Ready-made food and other food products (baby food, spices)), plus home-produced foods produced in the last 7 days, plus assistance received in the last 30 days divided by 4.33. Total consumption, including food and non-food expenses, was only measured at endline due to the length of the module. Reported in this table is food consumption, which was measured at endline and followup. CES-D was measured in all rounds with exactly the same questionnaire. It includes 10 questions on depressive symptoms. Individuals who skipped more than 2 questions were marked as missing. The continuous scores were standardized with respect to the control group. The Child Strengths and Difficulties questionnaire (SDQ) was administered regarding a randomly selected child aged 3-8 at endline. At follow-up, the SDQ was administered regarding the same child. If not surveyed previously or if that child is no longer in the household, the survey was administered with respect to a randomly selected child age 4-9. The continuous scores were randomized with respect to the control group. The balance test consists of two equations. Equation 1: $\text{Outcome} = \text{Controls} + \text{FE} + \text{error}$. From Equation 1, we obtain fitted values. Then, we use them as the outcome variable in Equation 2: $\widehat{\text{Outcome}} = \beta \cdot \text{Treatment} + \text{FE} + \text{error}$. The test is $H_0 : \beta = 0$ vs. $H_a : \beta \neq 0$. The regressions have assessment month-by-year fixed effects, enumerator fixed effects, community-level controls (Irbid/Mafraq governorate and population quartile), and household-level controls (vulnerability-assessment quartile, shelter program, baseline number of children, baseline number of children plus adults, respondent gender, and respondent age). In the parentheses are robust standard errors clustered at the locality level. Regressions are weighted by the number of people interviewed in each household. Statistical significance represented by * (10%), ** (5%), and *** (1%). Q-values are calculated per Anderson (2008) and correspond to Family 1 in the pre-analysis plan.

B.4 Addressing Low Take-up Rates

Table (B.9) Baseline Averages Interacting Treatment with Compliance

Outcome	Assignment	Noncompliers (A=1, T=0)		Compliers (A=1, T=1)		Control group (A=0, T=0)		(3)-(1)	
		(1) Pt. Est.	(2) SE	(3) Pt. Est.	(4) SE	(5) Mean	(6) SD	(7) Pt. Est.	(8) SE
Governorate: Mafrqa	High	-0.00	(0.01)	0.00	(0.00)	0.48	(0.50)	0.01	(0.01)
Governorate: Irbid	High	0.00	(0.01)	-0.00	(0.00)	0.52	(0.50)	-0.01	(0.01)
District: Aghwar N.	High	0.00	(0.00)	0.00	(0.00)	0.00	(0.04)	0.00	(0.00)
District: Badiyah N.	High	-0.02	(0.05)	-0.01	(0.04)	0.14	(0.35)	0.01	(0.03)
District: Badiyah NW.	High	-0.04	(0.08)	-0.03	(0.08)	0.20	(0.40)	0.01	(0.02)
District: Bani Kenanah	High	-0.03	(0.02)	0.00	(0.01)	0.08	(0.28)	0.03*	(0.02)
District: Bani Obeid	High	0.07**	(0.03)	0.07**	(0.03)	0.02	(0.13)	-0.00	(0.01)
District: Irbid Qasabah	High	-0.16	(0.10)	-0.16*	(0.09)	0.30	(0.46)	-0.00	(0.02)
District: Jerash Qasabah	High	-0.00	(0.00)	-0.00	(0.00)	0.00	(0.04)	-0.00	(0.00)
District: Koorah	High	-0.05**	(0.03)	-0.04*	(0.02)	0.03	(0.17)	0.01	(0.01)
District: Mafrqa Qasabah	High	0.06	(0.07)	0.04	(0.08)	0.18	(0.39)	-0.02	(0.03)
District: Mazar N.	High	-0.00	(0.01)	-0.00	(0.01)	0.01	(0.10)	-0.00	(0.00)
District: Ramtha	High	0.15	(0.10)	0.14	(0.10)	0.02	(0.15)	-0.01	(0.02)
District: Rwaished	High	-0.00	(0.00)	-0.00	(0.00)	0.00	(0.04)	-0.00	(0.00)
District: Taybeh	High	0.03	(0.02)	-0.01	(0.01)	0.00	(0.06)	-0.03	(0.02)
District: Wastiyyah	High	-0.00	(0.01)	0.00	(0.01)	0.01	(0.07)	0.01	(0.01)
Team: ICLA Amman	High	0.01	(0.01)	-0.00	(0.00)	0.00	(0.05)	-0.01	(0.01)
Team: ICLA Irbid	High	-0.07**	(0.04)	-0.02	(0.06)	0.13	(0.34)	0.06*	(0.03)
Team: ICLA Jarash	High	-0.00	(0.00)	-0.00	(0.00)	0.00	(0.04)	-0.00	(0.00)
Team: ICLA Madaba	High	0.00	(.)	0.00	(.)	0.00	(0.00)	0.00	(0.00)
Team: ICLA Mafrqa	High	-0.03	(0.04)	-0.06	(0.04)	0.37	(0.48)	-0.03	(0.03)
Team: ICLA Ramtha	High	0.02	(0.03)	0.02	(0.04)	0.04	(0.21)	0.00	(0.02)
Team: Shelter Irbid	High	0.06*	(0.03)	0.01	(0.04)	0.29	(0.45)	-0.05	(0.03)
Team: Shelter Mafrqa	High	0.02	(0.04)	0.05	(0.04)	0.16	(0.36)	0.03	(0.03)
Team: Shelter Ramtha	High	0.00	(0.00)	-0.00	(0.00)	0.00	(0.04)	-0.01	(0.00)
Team: Eman Naba' fdn.	High	0.00	(.)	0.00	(.)	0.00	(0.00)	0.00	(0.00)
Month assessed (03/2019 - 06/2020)	High	0.44	(0.34)	-0.76**	(0.33)	7.92	(3.27)	-1.20***	(0.27)
Married {0,1}	High	0.01	(0.02)	-0.00	(0.03)	0.84	(0.37)	-0.01	(0.03)
Distability level = 3 {0,1}	High	0.01	(0.03)	0.01	(0.03)	0.24	(0.43)	-0.00	(0.04)
General vulnerability score	High	0.01	(0.01)	-0.00	(0.01)	0.51	(0.14)	-0.01	(0.01)
Dependency ratio score	High	0.00	(0.01)	-0.00	(0.01)	0.50	(0.24)	-0.01	(0.01)
Living score	High	0.00	(0.01)	-0.02**	(0.01)	0.53	(0.30)	-0.02**	(0.01)
Family score	High	-0.00	(0.01)	-0.01	(0.01)	0.42	(0.20)	-0.01	(0.01)
Single HH head score	High	-0.01	(0.01)	-0.01	(0.02)	0.43	(0.29)	-0.00	(0.03)
Debt score	High	0.04	(0.02)	0.06***	(0.02)	0.49	(0.26)	0.02	(0.02)
Disability score	High	0.00	(0.02)	0.01	(0.02)	0.48	(0.28)	0.00	(0.02)
High treatment score	High	0.00	(0.00)	-0.00**	(0.00)	0.69	(0.13)	-0.01**	(0.00)
Low treatment score	High	0.02*	(0.01)	0.01	(0.01)	0.26	(0.18)	-0.01	(0.01)
WASH score	High	0.06	(0.04)	0.05	(0.04)	0.01	(0.06)	-0.01	(0.01)
C4R score	High	-0.01*	(0.00)	0.00	(0.01)	0.01	(0.08)	0.01	(0.01)
WC score	High	0.00	(0.00)	0.01	(0.01)	0.00	(0.00)	0.01	(0.01)
IK score	High	0.01	(0.03)	0.01	(0.03)	0.23	(0.40)	-0.00	(0.03)
Shelter type = permanent {0,1}	High	0.04	(0.03)	0.06***	(0.02)	0.85	(0.35)	0.02	(0.03)
Times moved (12 months, wins.)	High	-0.03	(0.06)	0.18**	(0.09)	0.46	(0.91)	0.20**	(0.09)
Has UNHCR proof of res. {0,1}	High	0.01	(0.03)	-0.01	(0.04)	0.83	(0.37)	-0.02	(0.03)
Rent (monthly, wins.)	High	1.56	(5.15)	-1.62	(3.99)	95.56	(50.35)	-3.18	(4.35)
Has debt with landlord {0,1}	High	0.03	(0.04)	0.06*	(0.04)	0.71	(0.46)	0.03	(0.04)
Months owed to landlord (wins.)	High	0.19	(0.15)	0.34**	(0.15)	2.03	(1.84)	0.15	(0.16)

Note: Columns (1)-(4) display results from an OLS regression that interacts treatment assigned with treatment received. Columns (5)-(6) display the mean and standard deviation in the control (base) group. Columns (7)-(8) display results from a test comparing the averages in columns (3) and (1) such that $(7)=(3)-(1)$ and $(8)=\sqrt{(2)^2 + (4)^2 - 2 * cov((2), (4))}$ where $(.)$ is a symbol for column entries. SEs are clustered at the locality level. Regressions control for age, gender, HH size, enumerator fixed effects, and quartiles for subsidy score, MA, and, IR. The month assessed variable takes a value of 1 in March 2019, 2 in April, and so on. The dependent variables are observed in the 2021 phone survey. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table (B.10) Variables Selected for LASSO

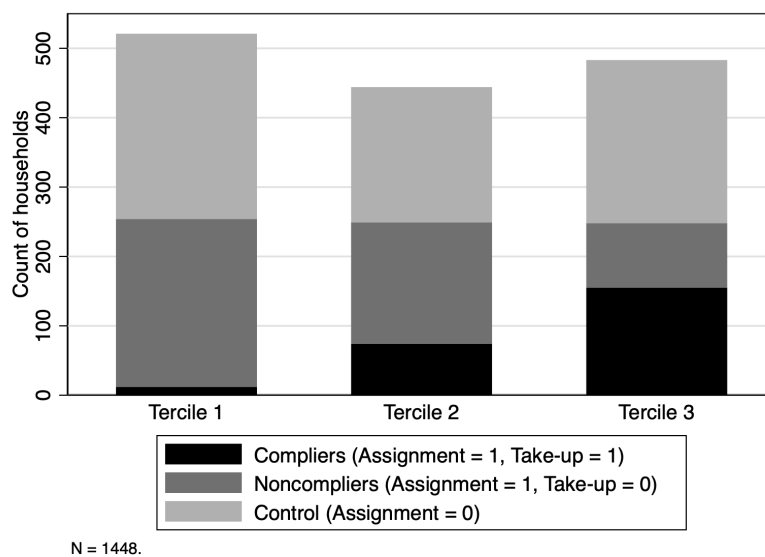
Covariate	Treatment group	
	All	High
district 1		x
district 3		x
district 4		x
district 6		x
district 9		x
district 10		x
district 13		x
district 14		x
gov 1 * team 7		x
gov 1 * changeshelter w		x
gov 1 * unhcr		x
gov 2 * Debtwiththelandlord		x
team 1 * PeopleLivingScore		x
team 1 * DebtScore		x
team 2 * PeopleLivingScore		x
team 2 * C4RScore2		x
team 2 * IKSore2		x
team 2 * changeshelter w		x
team 3 * genvul sc		x
team 5 * PeopleLivingScore		x
team 5 * SingleHoHScore		x
team 5 * flex sc		x
team 5 * unhcr		x
team 6 * month order		x
team 6 * married		x
team 6 * PeopleLivingScore		x
team 6 * FamilyScore		x
team 6 * DisabilityScore		x
team 7 * Dependent ratio score		x
team 7 * DisabilityScore		x
team 7 * REScore2		x
team 8 * SingleHoHScore		x
team 8 * flex sc		x
team 9 * DisabilityScore		x
married * FamilyScore		x
married * flex sc		x
married * changeshelter w		x
DisabilityLevel3 * Dependent ratio score		x
DisabilityLevel3 * WASHScore2		x
PeopleLivingScore * unhcr		x
PeopleLivingScore * Rentpermonth w		x
FamilyScore * changeshelter w		x
SingleHoHScore * flex sc		x
SingleHoHScore * REScore2		x
SingleHoHScore * WASHScore2		x

Continued on next page

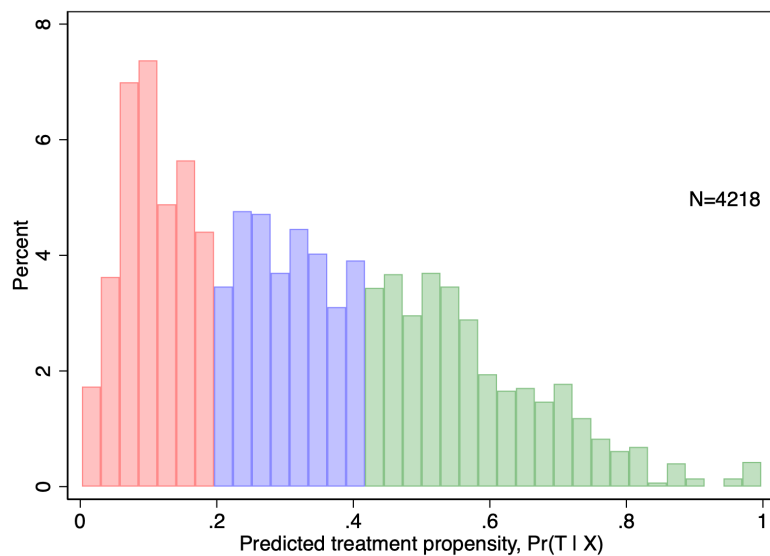
Table B.10 – continued from previous page

Covariate	Treatment group		
	All	High	Low
SingleHoHScore * IKSore2		x	
SingleHoHScore * changeshelter w		x	
DebtScore * REScore2		x	
flex sc * ShelterType		x	
WASHScore2 * C4RScore2		x	
WASHScore2 * changeshelter w		x	
WASHScore2 * Months owed to landlord		x	
C4RScore2 * changeshelter w		x	
C4RScore2 * unhcr		x	
IKScore2 * changeshelter w		x	
ShelterType * unhcr		x	
genvul sc sq		x	
flex sc sq		x	
month order sq		x	
Rentpermonth w sq		x	
cons		x	

Note: This table shows the variables selected in the lasso model.



(a) Actual Compliance by Predicted Tercile



(b) Predicted Tercile using LASSO

Figure (B.1) Heterogeneity Analysis: LASSO-Based Predicted Compliance

Table (B.11) Primary Outcomes

Outcome	A (se) [FDR]	Ter2	Ter3	A × Ter2	A × Ter3	Control	N
Overall Housing Quality (Z-Score)	0.06 (0.05) [0.54]	-0.04 (0.05)	-0.07 (0.08)	0.09 (0.06)	0.01 (0.07)	-0.01 (1.00)	4,218
Monthly Housing Expenditures (\$ PPP)	-25.64 (19.38) [0.54]	-19.83 (25.78)	-29.31 (33.74)	-7.72 (28.22)	11.53 (28.33)	201.06 (209.30)	1,421
Food Consumption (Log USD PPP)	0.12*** (0.05) [0.06]	0.10 (0.07)	0.14 (0.09)	-0.11 (0.08)	-0.15** (0.07)	4.41 (0.72)	2,753
Log Total Consumption (Log \$ PPP)	0.06 (0.06) [0.54]	0.12 (0.08)	0.13 (0.11)	-0.11 (0.08)	-0.04 (0.08)	8.73 (0.55)	1,422
CESD Score (Higher: Less Depression)	0.03 (0.08) [0.73]	0.14 (0.09)	0.17 (0.12)	-0.10 (0.10)	-0.02 (0.12)	0.01 (1.00)	4,207
SDQ Score (Higher: Better Wellbeing)	-0.08 (0.11) [0.73]	-0.00 (0.15)	-0.11 (0.20)	-0.08 (0.12)	0.02 (0.18)	0.00 (1.00)	1,782

Note: “A” denotes assignment to treatment and “Ter” is the tercile in discussion. This table reports impacts of treatment assignment on Syrian refugees in the full sample. Some outcomes are not available in all rounds. Interacted with treatment assignment are terciles of the LASSO-predicted treatment propensity described and shown above. Regressions include baseline covariates selected in the LASSO shrinkage step.

Table (B.12) Primary Outcomes and Selected Social Attitudes Index Components

Outcome	Assg. (se)	Tercile 2 (se)	Tercile 3 (se)	Assg. × Terc. 2 (se)	Assg. × Terc. 3 (se)	Control (sd)	N
Social attitudes & perceptions (SD)	-0.19** (0.09)	-0.17 (0.14)	-0.03 (0.19)	0.17 (0.16)	-0.00 (0.13)	0.01 (1.00)	1,102
Economic attitudes & percep (SD)	0.02 (0.10)	-0.24* (0.14)	0.07 (0.19)	0.18 (0.15)	-0.24* (0.14)	0.02 (1.02)	1,102
Policy preferences (SD)	0.15* (0.08)	-0.01 (0.12)	0.13 (0.18)	-0.09 (0.14)	-0.13 (0.13)	-0.02 (0.99)	1,102
Altruism to Syrians	0.13 (0.12)	0.09 (0.15)	-0.01 (0.20)	-0.18 (0.22)	-0.11 (0.20)	0.88 (1.17)	1,102
Syrians in advice network (1-3)	-0.15** (0.07)	-0.13** (0.07)	-0.05 (0.13)	0.11 (0.09)	-0.01 (0.11)	0.34 (0.72)	1,102
Any refugee friends for children?	-0.05 (0.05)	-0.01 (0.08)	0.07 (0.13)	0.05 (0.09)	0.00 (0.08)	0.41 (0.49)	704

Note: This table reports impacts of treatment assignment on Jordanian neighbors of Syrian refugees in the full sample. Interacted with treatment assignment are terciles of the LASSO-predicted treatment propensity described and shown above. Regressions include baseline covariates selected in the LASSO shrinkage step.

Intent-to-Treat Results**Table (B.13)** Pooled Treatment Effects on Primary Outcomes

Outcomes	ITT (se)	TOT (se)	FDR q-values	Control (sd)	N	Rounds
Overall Housing Quality (Z-Score)	0.11 (0.07)	0.31 (0.20)	[0.19]	-0.01 (1.00)	4,218	1, 2, 3
Monthly Housing Expenditures (\$ PPP)	-28.99*** (11.50)	-84.71*** (33.10)	[0.07]	228.80 (208.80)	1,421	2
Food Consumption (Log \$ PPP)	-0.01 (0.03)	-0.02 (0.07)	[0.63]	4.41 (0.72)	2,753	2, 3
Log Total Consumption (Log \$ PPP)	-0.01 (0.03)	-0.03 (0.08)	[0.63]	8.73 (0.55)	1,422	2
CESD Score (Higher: Less Depression)	-0.03 (0.05)	-0.10 (0.13)	[0.49]	0.01 (1.00)	4,207	1, 2, 3
SDQ Score (Higher: Better Wellbeing)	-0.12** (0.06)	-0.34** (0.15)	[0.07]	-0.01 (1.00)	1,782	2, 3

Note: This table replicates the Treatment-on-the-treated results, showing Intent-to-Treat effects for comparison.

Table (B.14) Significant Impacts of the Program at Midline

Outcomes	ITT	(se)	Control mean	(sd)	N
Panel A: Housing Improvements					
Quality Roof (=1)	0.08***	(0.03)	0.71	(0.45)	1,610
Clean water (=1)	0.07***	(0.02)	0.17	(0.38)	1,610
Occupied Rooms	0.18***	(0.06)	2.85	(1.05)	1,610
Panel B: Food Aid and Insecurity					
Food consumption (aid) USD PPP	-18.23**	(7.85)	218.51	(201.68)	1,569
Food consumption (aid) USD PPP (<i>Per capita</i>)	-3.34***	(1.17)	35.88	(29.19)	1,569
Respondent hunger last week (=1)	0.08***	(0.02)	0.35	(0.48)	1,604
Adults hunger last week (=1)	0.06***	(0.02)	0.38	(0.49)	1,610
Child hunger last week (=1)	0.09***	(0.03)	0.23	(0.42)	1,469
Panel C: Household Recomposition					
Number of children under 18 years	0.11**	(0.05)	3.27	(1.97)	1,610
Number of boys aged 13-17 years in household	0.06**	(0.03)	0.40	(0.66)	1,610
Panel D: COVID-19 Outcomes					
Weekly adult income (during first lockdown)	-7.05***	(2.58)	30.22	(66.86)	1,609
Total COVID-19 symptoms	0.29***	(0.06)	0.47	(1.22)	1,610
Number of people who are symptomatic	0.16***	(0.03)	0.29	(0.61)	1,581
Number of children who are symptomatic	0.06***	(0.02)	0.06	(0.32)	1,472
Number of visits to healthcare institutions	-0.20**	(0.08)	1.14	(2.03)	1,608
Did not keep distance (=1)	0.05**	(0.02)	0.51	(0.50)	1,609
Panel E: Loans					
Loans taken pre-first-lockdown (=1)	-0.07***	(0.02)	0.61	(0.49)	1,608
Loans taken during first lockdown (=1)	-0.12***	(0.02)	0.78	(0.41)	1,608
Loans taken after first lockdown (=1)	-0.12***	(0.02)	0.42	(0.49)	1,607
Panel F: Other effects					
Total number of school days attended per child	-0.16*	(0.10)	2.33	(2.49)	1,031
Child alertness (s.d. units)	-0.24***	(0.07)	0.00	(1.00)	975
Difficulty with self-care (z-score)	0.13***	(0.05)	-0.00	(1.00)	1,610
Number of learning activities	0.07**	(0.03)	0.22	(0.56)	1,385

Note: This table replicates the Treatment-on-the-treated results, showing Intent-to-Treat effects instead.

Table (B.15) Significant Impacts of the Program at Endline and Follow-up

Outcomes	ITT	(se)	Control mean	(sd)	N
Panel A: Endline					
Focus Respondent Childcare & Chores (Hours)	-2.34***	(0.64)	17.76	(18.39)	1,418
Total Labor and Chores Hours Last Week	-3.63***	(1.05)	25.54	(23.39)	1,422
General Happiness (Scale 1-3)	-0.08***	(0.03)	1.76	(0.61)	1,422
At least 30 JD (95 USD PPP) in savings (=1)	0.05***	(0.02)	0.12	(0.32)	1,422
Applied but Did Not Take Loan (=1)	0.02***	(0.01)	0.01	(0.10)	1,422
IP Assistance in Last 12 Months (=1)	0.03**	(0.01)	0.04	(0.19)	1,422
IP Assistance in Last 12 Months (USD PPP)	43.88***	(15.86)	28.74	(198.87)	1,422
Panel B: Follow-up					
Child SDQ score, std	-0.21***	(0.07)	-0.01	(1.00)	859

Note: This table replicates the Treatment-on-the-treated results, showing Intent-to-Treat effects instead.

Table (B.16) Impacts on Neighbor's Attitudes Toward Refugees

Outcomes	ITT (se)	TOT (se)	FDR q-values	Control mean (sd)	N
Panel A: Primary Outcomes					
Social attitudes & perceptions (SD)	-0.33** (0.14)	-0.11** (0.05)	[0.08]	-0.00 (1.00)	1,102
Economic attitudes & perceptions (SD)	0.02 (0.16)	0.01 (0.06)	[0.61]	0.02 (1.02)	1,102
Policy preferences (SD)	0.17 (0.15)	0.06 (0.05)	[0.40]	-0.02 (1.00)	1,102
Altruism to Syrians	0.21 (0.18)	0.07 (0.06)	[0.40]	0.88 (1.18)	1,102
Panel B: Selected Social Attitudes Index Components					
Of the 3 people you exchange advice with, how many are Syrian refugees?	-0.24** (0.11)	-0.08** (0.04)	[0.41]	0.34 (0.72)	1,102
Do the children in this household have any Syrian refugee friends?	-0.19* (0.11)	-0.06* (0.04)	[0.42]	0.40 (0.49)	704
Panel C: Other Related Outcomes					
Days of media consumption (last week)	-1.08*** (0.40)	-0.36*** (0.13)	[0.09]	4.78 (2.78)	1,102
Neighbor perceptions of average refugee aid receipt (PPP)	-347.17*** (109.69)	-111.60*** (37.76)	[0.09]	630.84 (657.41)	685
Life Satisfaction (SD)	-0.32** (0.15)	-0.11** (0.05)	[0.06]	0.01 (0.98)	1,097
Subjective Health (SD)	-0.30* (0.16)	-0.10* (0.05)	[0.06]	0.02 (0.99)	1,102

Note: This table replicates the TOT results, showing Intent-to-Treat effects for comparison.

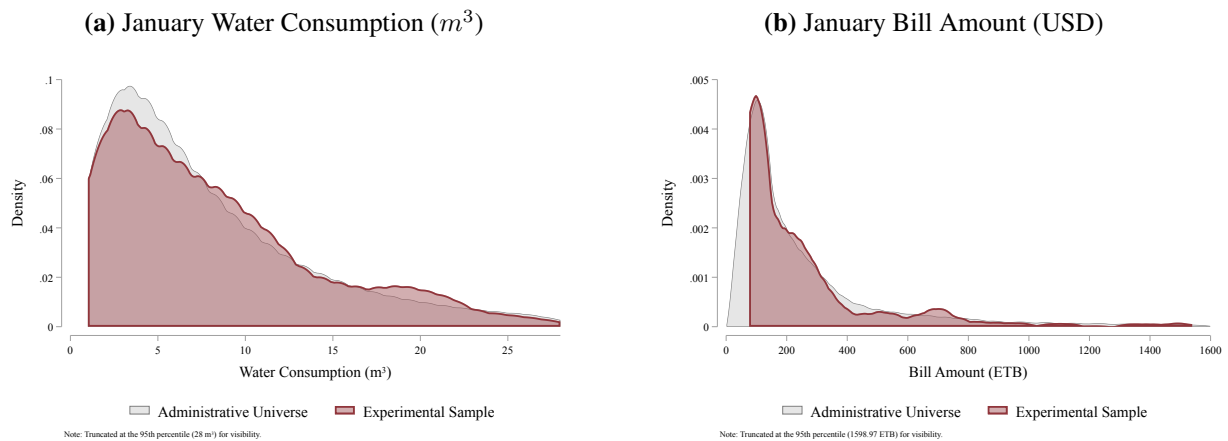
Appendix C

Appendix to Digital Payments Adoption

C.1 Sample Representativeness

Figure C.1 shows the densities. To improve visual comparability, we truncate at the 95th percentile of the universe distribution and apply a unified kernel smoothing bandwidth estimated from the experimental sample, ensuring that the vastly larger administrative dataset does not produce misleadingly granular density estimates. Two-sample t -tests fail to reject equality of means for consumption ($\bar{x}_{\text{universe}} = 9.6$, $\bar{x}_{\text{sample}} = 10.3$; $p = 0.185$) and bill amounts ($\bar{x}_{\text{universe}} = 484$, $\bar{x}_{\text{sample}} = 436$; $p = 0.446$), and KS tests, though rejecting distributional equality at conventional levels ($p = 0.002$ and $p = 0.008$ respectively), likely reflect the extreme statistical power from the large number of observations rather than economically meaningful divergence. These comparisons suggest that persistent in-person payers are not a systematically unusual subset in terms of consumption or billing, though by construction they differ sharply from the broader population in their payment channel behavior. Our results should therefore be interpreted as speaking to this specific margin of habitual in-person payers who have access to digital channels but have not yet used them rather than to the full customer population.

Figure (C.1) Administrative Universe vs. Experimental Sample



Notes: This figure compares the distribution of January water consumption (Panel a) and bill amounts (Panel b) between the experimental sample (maroon) and the administrative universe of utility customers (gray). Kernel density estimates are plotted using a unified bandwidth derived from the experimental sample to ensure visual comparability across vastly different sample sizes. Distributions are truncated at the 95th percentile of the universe to omit extreme right-tail outliers.

C.2 Additional Figures and Tables

Table (C.1) Balance of Baseline Covariates Across Treatment Arms

	Means			<i>p</i> -values of differences		
	Control	Information	Incentive	Control– Information	Control– Incentive	Information– Incentive
	(1)	(2)	(3)	(4)	(5)	(6)
Female respondent	0.66 (0.47)	0.60 (0.49)	0.63 (0.48)	0.20	0.53	0.56
Age	49.22 (14.43)	50.34 (14.61)	49.82 (15.46)	0.42	0.74	0.73
Household size	4.52 (2.02)	4.32 (1.85)	4.77 (1.94)	0.31	0.23	0.01
Has bank account	0.96 (0.19)	0.94 (0.24)	0.96 (0.20)	0.27	0.83	0.38
Has mobile money	0.65 (0.48)	0.61 (0.49)	0.63 (0.48)	0.40	0.60	0.69
Internet at home	0.34 (0.48)	0.32 (0.47)	0.36 (0.48)	0.59	0.79	0.39
Asset index	2.36 (1.02)	2.33 (0.98)	2.48 (1.01)	0.78	0.26	0.12
Avg. water bill (USD)	2.78 (5.14)	2.28 (3.64)	2.52 (3.96)	0.23	0.53	0.47
Has water tank	0.37 (0.48)	0.35 (0.48)	0.35 (0.48)	0.63	0.62	0.93
Shared meter	0.22 (0.41)	0.24 (0.43)	0.20 (0.40)	0.43	0.72	0.23
Late fee (ever)	0.02 (0.12)	0.01 (0.07)	0.01 (0.10)	0.27	0.65	0.57
Household income (USD)	124.22 (128.28)	119.52 (123.48)	124.55 (124.53)	0.64	0.95	0.65
Household expenditure (USD)	97.08 (80.96)	91.81 (77.70)	95.90 (71.66)	0.42	0.77	0.53
Observations	190	202	198			
<i>Joint F-test p-value</i>				0.999	1.000	0.999

Notes: Columns (1)–(3) report means with standard deviations in brackets. Columns (4)–(6) report *p*-values from two-sided tests of equality of means using OLS regressions with heteroskedasticity-robust standard errors. The joint *F*-test regresses the treatment indicator on all covariates and reports the *p*-value of the joint significance test.

Table (C.2) Attrition Rates

	Control (1)	Information (2)	Incentive (3)	Ctrl–Info (4)	Ctrl–Incent (5)	Info–Incent (6)
Attrition rate	0.079	0.079	0.096			
Observations	190	202	198			
Difference				-0.003	0.018	0.020
<i>p</i> -value				0.922	0.542	0.466
Joint <i>F</i> -test <i>p</i> -value				0.846	0.604	0.830

Notes: This table reports attrition rates by treatment arm and pairwise differences with *p*-values from OLS with robust standard errors. The joint *F*-test checks whether baseline covariates differentially predict attrition across arms.

Table (C.3) Baseline Means, Attriters vs. Stayers

	Attriters (1)	Stayers (2)	<i>p</i> -value (3)
Female respondent	0.560	0.635	0.287
Age	48.140	49.957	0.540
Household size	4.480	4.539	0.820
Has bank account	1.000	0.950	0.000
Has mobile money	0.620	0.633	0.631
Internet at home	0.280	0.344	0.222
Asset index	2.120	2.411	0.030
Avg. water bill	574.560	452.930	0.945
Has water tank	0.440	0.350	0.685
Shared meter	0.180	0.224	0.519
Late fee (ever)	0.000	0.011	0.096
Household income	16723.860	19476.254	0.129
Household expenditure	14391.880	14552.665	0.611

Notes: This table compares baseline means of attriters and stayers with *p*-values from regressions.

Table (C.4) Forecasts and Realized Treatment Effects

Outcome	Arm	Forecast [95% CI]	ATE (SE)	Diff	<i>p</i>
<i>Panel A: Digital adoption</i>					
Digital adoption (Dec)	Information	9.41 [7.47, 11.35]	6.77 (3.03)	2.64	0.407
	Incentive	21.97 [17.67, 26.27]	11.64 (3.31)	10.33***	0.009
Digital adoption (Jan)	Information	10.99 [8.98, 13.00]	9.28 (3.45)	1.71	0.635
	Incentive	23.36 [18.65, 28.08]	16.70 (3.77)	6.67	0.133
Digital adoption (Feb)	Information	10.92 [8.29, 13.55]	13.27 (3.99)	-2.35	0.575
	Incentive	17.56 [13.37, 21.74]	22.52 (4.37)	-4.96	0.306
<i>Panel B: Secondary outcomes</i>					
Plans to e-pay	Information	18.31 [14.82, 21.81]	19.31 (5.17)	-0.99	0.856
	Incentive	22.10 [18.15, 26.05]	31.36 (4.99)	-9.26*	0.084
Has mobile money	Information	7.27 [4.30, 10.23]	0.96 (5.29)	6.30	0.251
	Incentive	10.90 [7.75, 14.04]	7.37 (5.31)	3.52	0.524
Time to pay (min)	Information	-10.67 [-14.11, -7.23]	-1.33 (2.43)	-9.34***	0.002
	Incentive	-9.98 [-12.91, -7.04]	-4.97 (2.34)	-5.01*	0.069

Notes: Forecasts are mean predictions across N=36 expert forecasters with bootstrap-based 95% confidence intervals. ATEs are intent-to-treat estimates with strata fixed effects and robust standard errors, scaled to forecast units. *p*-values are from two-sided Wald tests of equality between forecast and realized ATE. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

C.3 Statistical Power

Table C.5 reports ex-post statistical power and minimum detectable effects (MDE) for each primary outcome. Power is computed using the normal approximation

$$\text{Power} = \Phi(|\hat{\beta}/\text{SE}| - z_{0.025}),$$

and the MDE is the effect size detectable with 80% power given the observed standard error: $\text{MDE} = 2.80 \times \text{SE}$. The study is well-powered for the primary outcome (electronic payment adoption) for both treatment arms. Ex-post power reaches 99.3% for the incentive arm (MDE of 10.6 pp) and 76.8% for the information arm (MDE of 9.7 pp). Both MDEs are economically meaningful relative to the 8.9% control-group adoption rate. For the remaining primary outcomes (bill payment, water consumption, and bill amount), the study is substantially underpowered. Power ranges from 5% to 18%, and the MDEs are large relative to control means (e.g., 6.7 pp for a 95.3% payment rate; 2.2 m³ for a mean of 9.4 m³). This lack of power is unsurprising: these outcomes exhibit high variance and the treatments were not designed to affect total consumption or payment levels, only the mode of payment. We caution against interpreting the null results on these outcomes as evidence that the treatments had no effect; the confidence intervals are wide enough to include economically meaningful effects in both directions.

Table (C.5) Ex-Post Statistical Power

Outcome	Treatment	$\hat{\beta}$	MDE (80%)	Power
E-Payment (Dec)	incentive	0.1164	0.0927	0.941
	info	0.0677	0.0849	0.608
E-Payment (Jan)	incentive	0.1670	0.1057	0.993
	info	0.0928	0.0966	0.768
E-Payment (Feb)	incentive	0.2252	0.1225	0.999
	info	0.1327	0.1117	0.914
Paid bill (Jan)	incentive	-0.0234	0.0671	0.164
	info	-0.0027	0.0621	0.052
Water consumption (Jan)	incentive	0.7753	2.2099	0.166
	info	0.2229	2.1633	0.060
Bill amount (Jan)	incentive	57.7210	154.2663	0.182
	info	5.6508	146.3463	0.051

Notes: Ex-post power is computed as $\Phi(|\hat{\beta}/\text{SE}| - 1.96)$. MDE is the minimum detectable effect at 80% power, calculated as $2.80 \times \text{SE}$.