

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Essays in Labor and Development Economics

Permalink

<https://escholarship.org/uc/item/5x19j9j0>

ISBN

9798189270406

Author

Palmer, Isabel Bailey

Publication Date

2026-05-01

Peer reviewed|Thesis/dissertation

Essays in Labor and Development Economics

by

Isabel Bailey Palmer

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 Christopher Walters, Co-Chair

Professor Supreet Kaur, Co-Chair

Professor Edward Miguel

Spring 2026

Essays in Labor and Development Economics

Copyright 2026
by
Isabel Bailey Palmer

Abstract

Essays in Labor and Development Economics

by

Isabel Bailey Palmer

Doctor of Philosophy in Economics

University of California, Berkeley

Professor Christopher Walters, Co-Chair

Professor Supreet Kaur, Co-Chair

This dissertation examines economic challenges faced by three marginalized groups in Middle Eastern labor markets — women, young adults, and refugees — and evaluates policies that expand access and promote economic empowerment. Chapter 1 examines how adverse selection on job amenities like transportation prevents firms from offering jobs which would increase women’s labor market access, and evaluates different policies to solve this market failure. Chapter 2 examines the long run impacts of entering the labor market during a recession on young adult’s labor market and marriage outcomes in Egypt. Chapter 3 evaluates a rental subsidy for Syrian refugees in Jordan and its impact on labor market participation, child welfare, and social integration.

The first chapter, titled “Job Amenities, Adverse Selection, and Women’s Labor Supply in Jordan,” examines how adverse selection on job amenities impacts women’s labor market access in Jordan. In many settings, women face logistical barriers to working, such as transportation and childcare. Firms could bridge these gaps by offering job amenities like firm-provided transportation, yet these amenities remain rare in developing countries. This paper asks whether adverse selection can help explain why. When workers’ willingness to pay (WTP) for job amenities is negatively correlated with unobserved productivity, offering amenities instead of higher wages can disproportionately attract less productive workers. This dynamic may lead firms to avoid providing amenities even when WTP exceeds cost. This issue is particularly relevant

in Jordan, where female employment is low and logistical barriers to work are significant. Working with a Jordanian research firm that exclusively hires female survey enumerators, I combine administrative firm records on worker productivity with a survey eliciting individual level WTP for job amenities such as childcare, transportation, and preferred pay schemes. For all amenities except childcare, the least productive workers have the highest willingness to pay. This negative correlation is particularly large for transportation, and cannot be fully explained by income, location, commuting costs, or any other observable. To assess the empirical magnitude of this adverse selection, I estimate a monopsony model that allows job amenities to increase labor supply and attract workers with different unobserved productivities. Estimates of the model demonstrate that mandating firm-provided transportation while allowing wages to adjust downward would increase employment, due to workers' large WTP for the amenity, but would lower average worker productivity via negative selection. The net effect is a reduction in firm profit, which may explain why the firm does not offer transportation. However, allowing workers to sort between a contract offering a higher wage without transportation and one offering a lower wage with transportation could simultaneously solve the adverse selection problem and increase worker welfare.

The second chapter, titled "Economic and Social Impacts of Entering the Labor Market During a Recession: Evidence from Egypt," examines labor market scarring among Egyptian youth. Relatively little is known about the effects of entering the labor market during a recession for young people in developing countries. Using novel, newly digitized historical unemployment rates from Egypt stretching back to 1968, I estimate the long run labor and marriage market effects of entering the labor market during a period of high unemployment. I find effects which reflect the same kind of economic scarring found in high income countries, but which manifest differently in this setting due to the differences in the job ladder. Men experience reductions in employment for 2 years after entry, but the long run effect on employment is positive. This unexpected increase is mostly driven by an increase in self-employment without any employees, often a form of disguised unemployment. Male recession entrants are also less likely to ever have a formal wage work job, an effect which is strongest for college graduates. In terms of family formation, both men and women turning 18 during a recession experience delays in entering marriage. However, there is some suggestive evidence that this benefits women: female recession entrants are more likely to complete secondary education and experience increased decision-making in their marriages.

The third chapter, titled "Housing Subsidies for Refugees: Experimental Evidence on Life Outcomes and Social Integration in Jordan" and coauthored with Abdulrazzak Tamim, Emma Smith, Edward Miguel, Samuel Leone, Sandra V. Roza and Sarah Stillman, evaluates the impact of a rental subsidy for Syrian refugees in Jordan. Refugees require assistance for basic needs like housing but local host communities may feel excluded from that assistance, potentially affecting community relations. This study experimentally evaluates the effect of a housing assistance program for Syrian refugees in Jordan on both the recipients and their neighbors. The program offered full rental

subsidies and landlord incentives for housing improvements, but saw only moderate uptake, in part due to landlord reluctance. The program improved short-run housing quality and lowered housing expenditures, but did not yield sustained economic benefits, partly due to redistribution of aid. The program unexpectedly led to a deterioration in child socio-emotional well-being, and also strained relations between Jordanian neighbors and refugees. In all, housing subsidies had limited measurable benefits for refugee well-being while worsening social cohesion, highlighting the possible need for alternative forms of aid.

To my family and friends

Contents

Dedication	i
Acknowledgments	viii
1 Job Amenities, Adverse Selection, and Women’s Labor Supply in Jordan	1
1.1 Introduction	1
1.2 Institutional Background	6
1.2.1 Jordan	6
1.2.2 Research Firm	6
1.2.3 Amenities	8
1.3 Data	8
1.3.1 Firm Data	8
1.3.2 Elicitation of Willingness to Pay for Job Amenities	11
1.3.3 Descriptive Statistics	13
1.4 Empirical Strategy	16
1.5 Results	17
1.5.1 Ruling Out Alternative Mechanisms	18
1.6 Counterfactuals with Firm-Provided Transportation	21
1.6.1 Model	21
1.6.2 Estimation	24
1.7 Counterfactual Results	28
1.7.1 Counterfactual: Amenity Mandate and Relaxed Minimum Wage	28
1.7.2 Counterfactual: Minimum Transportation Subsidy Required to Make Provision Profitable	29
1.7.3 Counterfactual: Firm Offers Workers Choice of Transportation or Higher Wage	29
1.8 Policy Implications	30
1.9 Conclusion	32
1.10 Figures	35
1.11 Tables	40
2 Economic and Social Impacts of Entering the Labor Market During	

a Recession: Evidence from Egypt	49
2.1 Introduction	49
2.2 Data	52
2.2.1 Unemployment Rates	52
2.2.2 Egyptian Labor Market Panel	53
2.3 Empirical Strategy	55
2.3.1 Labor Market Outcomes	55
2.3.2 Marriage Outcomes	56
2.4 Context and Descriptive Statistics	57
2.5 Results	59
2.5.1 Education	59
2.5.2 Employment	59
2.5.3 Wages and Earnings	61
2.5.4 Marriage Market Outcomes	62
2.6 Conclusion	63
2.7 Figures	64
2.8 Tables	67
 3 Housing Subsidies for Refugees: Experimental Evidence on Life Outcomes and Social Integration in Jordan	 77
3.1 Introduction	77
3.2 Background and Context	82
3.2.1 Syrian Refugees in Jordan	82
3.2.2 Humanitarian Shelter Program	83
3.3 Experimental Design and Data	84
3.3.1 Data on refugee outcomes	86
3.3.2 Measures of social cohesion among neighbors	88
3.3.3 Data sources	88
3.3.4 Descriptive statistics	89
3.3.5 Forecasts of Program Impacts	90
3.4 Empirical Strategy	91
3.4.1 First stage compliance with program assignment	91
3.4.2 Estimating program impacts on refugee well-being	91
3.4.3 Bounding estimated effects in the presence of non-random attrition	92
3.4.4 Estimating impacts of the program on neighbors' attitudes . . .	92
3.4.5 Estimating the accuracy of forecasts	93
3.5 Housing Subsidy Program Impacts on Refugee Life Outcomes	93
3.5.1 Average program impacts on refugees pooling survey rounds . .	93
3.5.2 Impacts in the midline round	95
3.5.3 Impacts in the endline survey round	97
3.5.4 Impacts in the follow-up survey round	99
3.6 Impacts on Neighbor Attitudes	99
3.7 Addressing Low Take-up Rates	101

3.8	Comparing Forecasts to Estimated Program Impacts	103
3.9	Conclusion	104
3.10	Tables and Figures	107
A	Appendix to Job Amenities, Adverse Selection, and Women’s Labor Supply in Jordan	126
A.1	Appendix Figures and Tables	127
A.2	Willingness-to-pay Elicitation Appendix	130
A.3	Mathematical Derivations	132
A.4	Full Text of Survey Instrument	133
B	Appendix to Economic and Social Impacts of Entering the Labor Market During a Recession: Evidence from Egypt	134
B.1	Appendix Tables and Figures	134
C	Appendix to Housing Subsidies for Refugees: Experimental Evidence on Life Outcomes and Social Integration in Jordan	139
C.1	Lee Bounds	140
C.2	Heterogeneous Effects Analysis	144
C.3	Additional Results	146
C.4	Addressing Low Take-up Rates	148
C.4.1	Intent-to-Treat Results	159

List of Figures

1.1	Distribution of Productivity Scores	35
1.2	WTP for Job Amenities	36
1.3	Distribution of WTP	37
1.4	Distribution of WTP and Productivity	38
1.5	Probability Density of Worker Productivity by WTP Quartiles	39
2.1	National Unemployment Rates	64
2.2	Residualized Governorate Unemployment Rates	65
2.3	Two Views of Youth Joblessness	66
3.1	Geographic Location of Treated and Control Communities	107
3.2	RCT, Data Collection, and Forecasts Timeline	108
3.3	Program Impacts on Primary Outcomes - Estimates vs. Predictions . .	116
A.1	Distribution of Uncensored WTP	127
C.1	Actual compliance by predicted tercile (unique households assigned to treatment across all rounds)	155
C.2	Predicted tercile using the LASSO on a logistic regression (all rounds) .	156

List of Tables

1.1	Evidence on Lack of Additional Firm Screening	40
1.2	Summary Statistics - Study Sample vs Population	41
1.3	Summary Statistics: Sample Subpopulations	42
1.4	Correlates of Willingness to Pay	43
1.5	Main Results	44
1.6	Transportation and Monthly Pay with administrative controls	45
1.7	Fixed Daily Wages and Childcare with administrative controls	46
1.8	Transport and Monthly Pay with Additional Survey-Based Controls	46
1.9	Labor Supply Parameters	47
1.10	Firm and Worker Outcomes Under Counterfactuals	48
2.1	ELMPS Observation by Round	67
2.2	Data quality	68
2.3	Descriptive Statistics from each ELMPS Wave	69
2.4	Effect of Entry Conditions on Educational Attainment	70
2.5	Long-run Effect of Entry Unemployment Rate on Male Employment	71
2.6	Long-run Effect of Entry Unemployment Rate on Female Employment	72
2.7	Effect of Entry Conditions on Earnings and Wages	73
2.8	Hazard into Marriage	74
2.9	Effect of Recession Entry on Marriage Outcomes	75
2.10	Marital Decisionmaking and Women's Mobility	76
3.1	Average Respondent, Household, and Shelter Characteristics	109
3.2	Forecasts of Treatment Effects	110
3.3	Panel Retention and Compliance	111
3.4	Pooled Treatment Effects on Primary Outcomes	112
3.5	Statistically Significant Impacts of the Program at Midline	113
3.6	Statistically Significant Impacts of the Program at Endline and Follow-up	114
3.7	Impacts of the Program on Neighbor's Attitudes Toward Refugees	115
A.1	Main Results with Selection-Correction Weights	128
A.2	Heterogeneity by Fixed vs Piece Rate Pay	129

B.1	Long-run Effect of Entry Unemployment Rate on Male Employment: Less than Secondary	135
B.2	Long-run Effect of Entry Unemployment Rate on Male Employment: Secondary Grads	136
B.3	Long-run Effect of Entry Unemployment Rate on Male Employment: University Grads	137
B.4	Long-run Effect of Entry Unemployment Rate on Male Employment: Post 1989	138
C.1	Pooled Treatment Effects on Primary Outcomes (with Lee Bounds) . .	140
C.2	Statistically Significant Impacts of the Program at Midline (with Lee Bounds)	141
C.3	Statistically Significant Impacts of the Program at Endline and Follow- up (with Lee bounds)	142
C.4	Impacts of the Program on Neighbor's Attitudes Toward Refugees (with Lee Bounds)	143
C.5	Gender Heterogeneity at Endline	144
C.6	Neighbor Impacts on Social Attitudes: Heterogeneity	145
C.7	Impacts of the Program on Earnings, Labor, & Occupational Choice .	146
C.8	Robustness of Baseline Balance	147
C.9	Baseline averages interacting treatment with compliance	148
C.10	Variables selected in the lasso step	151
C.11	T4, Panel A: Primary Outcomes	157
C.12	T7, Panels A & B: Primary Outcomes and Selected Social Attitudes Index Components	158
C.13	Intent-to-Treat - Pooled Treatment Effects on Primary Outcomes . . .	159
C.14	Intent to Treat - Statistically Significant Impacts of the Program at Midline	160
C.15	Intent to Treat - Statistically Significant Impacts of the Program at Endline and Follow-up	161
C.16	Intent to Treat Impacts of the Program on Neighbor's Attitudes Toward Refugees	162

Acknowledgments

This dissertation would not exist without the support of my (many) advisors, including Chris Walters, Ted Miguel, Pat Kline, Supreet Kaur, and Jesse Rothstein.

Chris, thank you so much for meeting with me so regularly and for so many hours. I have learned so much from you and am a better economist for it. In addition to all of your substantive contributions to my education and to this dissertation, your unflappability and calm was a critical input to my success.

Ted, literally none of this work would have existed without you and the research infrastructure you have created. I am incredibly grateful for the opportunity to have learned so much from you - economics, but also leadership, integrity, and what making a positive impact looks like.

Pat, my job market paper is dramatically better thanks to your feedback, and I am so grateful to have had the opportunity to learn from you for 6 years. Everything you do for the students of this department does not go unnoticed.

Supreet, thank you so much for all your feedback and support over the years. When designing experiments, I often think “what would Supreet do?”

Jesse, your office hours have been a bright spot of my PhD! Your honest feedback on both research and career considerations have been so important to my success.

I am also extremely grateful to the rest of the labor and development faculty groups for their comments and support over the years. Thank you to Fred Finan, Ben Faber, Jeremy Magruder, Marco Gonzalez-Navarro, Gautam Rao, David Card, Sydney Caldwell, Enrico Moretti, Benjamin Schoefer and Hillary Hoynes for all your contributions to both my work and my professional development as an economist.

Thank you to all the economists and educators who shaped my path here. I am so grateful to have learned from Leah Boustan, Alex Mas, Will Dobbie, and David Lee at Princeton and for all their support during my predoc. Thank you to Meghan Skira, Chris Cornwell, Nick Magnan, and Laura Zimmermann at the University of Georgia for introducing me to economics and for believing in me. Thank you to Mr. Elghotni at North Atlanta for giving me a YESAbroad flyer, and to Mr. Harris for submitting my recommendation to go to Oman in the middle of a snowstorm. Thank you to the American taxpayers and the YESAbroad staff for supporting my year abroad in Oman from 2011-2012 and to Ms. Amal, my Arabic teacher from Syria at Azzan Bin Qais who read Arabic picture books with me and who was also the first person to ever teach

me calculus.

My cohortmates and friends at Berkeley made this whole thing possible, and dare I say, enjoyable. Charoo, Matteo, Ale, Steve, Tere, Wouter, Sam, Alfredo, Chiara, Lucy, Dustin, Shreya, Margaret, Gaby, Caleb, Kristy, Sarah, Gwyneth, Tyler and so many others, I am so grateful for your friendship! I am also deeply grateful for my S-RLS teammates - Sandra, Emma, Sarah, Abed and Mansi, it was a honor working with and learning from you.

I would also like to thank my incredible colleagues in Jordan and Egypt for their contributions to much of the work contained in these chapters. Thank you to Mohammad Qardan, Majd Masannat and Feryal Ghazi for excellent project management and support; thank you to Rana Samara and the rest of the Mindset team for their continued support; and thank you to Ahmed El Sayed, who digitized the unemployment rate data which made chapter 2 possible.

I also gratefully acknowledge financial support from the CEGA Development Economics Challenge, the Weiss Fund for Research in Development Economics, Berkeley Economists for Equity, and the International Growth Centre. Without this funding, none of this research would have existed. Thank you to Samarth Bordia and the rest of the CEGA team who make most graduate development work at Berkeley possible, and to the IGC Jordan office whose support and facilitation have not only funded this work but also dramatically increased the probability of it having a policy impact.

Finally, the biggest thank you is owed to my family and friends who supported me every single day over the past 6 years. Thank you to my parents and sister who unconditionally love and believe in me, who remind me what matters, and who always call me back. Thank you to Will for being the world's best partner, I love you. I'm so glad I met you, twice.

Chapter 1

Job Amenities, Adverse Selection, and Women’s Labor Supply in Jordan

1.1 Introduction

Across much of the world, logistical barriers like transportation and childcare present obstacles to labor force participation for women who want to work (Martinez and Peticara, 2017; Clark et al., 2019; Christensen, Nino, and Osman, 2025; Martínez et al., 2020; Caria et al., 2025; Garlick, Field, and Vyborny, 2025). Firms could bridge these gaps by offering transportation or childcare as job amenities. This may be especially profitable in settings where secondary markets are missing or thin or where firms enjoy returns to scale and are able to offer these amenities more cheaply than workers can purchase them privately. Yet, female-friendly amenities are rare in places like Jordan, the setting of this paper (World Bank, 2024). Why don’t labor markets provide jobs that work better for women?

The lack of firm-provided amenities may be explained by adverse selection. Workers have heterogeneous willingness to pay (WTP) for job amenities like transportation and hold private information on their own productivity.¹ To the extent that preferences for job amenities are correlated with unobserved productivity, offering contracts with amenities may attract workers with different productivities. The theoretical potential for job amenities to induce worker selection has been highlighted in recent reviews by Mas (2025) and Kline (2025), and Emanuel and Harrington (2024) provide empirical evidence that introducing a work-from-home option induced negative selection on worker productivity among call center workers in the US.² Theory predicts that if

¹Many contributions have been made documenting variation in WTP for job amenities in a variety of settings (for example see Mas and Pallais, 2017; Feld, Nagy, and Osman, 2022; Mahmud et al., 2020; Bustelo et al., 2023; Dube, Naidu, and Reich, 2022)

²Brekke and Nyborg (2008) also develop a model of the labor market where morally minded workers who value working at a socially responsible firm also shirk less, making corporate responsibility

adverse selection is severe, meaning that the workers with the highest WTP are much less productive (unobservably), then the market for jobs with amenities can unravel completely (Nekoei, 2024). While Emanuel and Harrington (2024) have shown evidence of adverse selection, no study to date has established whether adverse selection can explain firms' failure to offer certain amenities. If it can, the resulting market failure would imply an equilibrium where women remain outside the labor force even when their marginal product exceeds their reservation compensation value and their willingness to pay for the amenity exceeds the firm's cost of providing it.

Why might worker productivity be negatively correlated with WTP for an amenity like transportation? One possibility is that workers with a higher cost of effort are less productive at work and also place a higher value on job amenities that increase convenience. Another possibility is that the type of woman who needs accommodative workplace amenities the most also disproportionately faces other challenges at home and elsewhere that make her less productive. For example, overlapping structural disadvantages could mean that the woman with the highest WTP for transportation simultaneously faces less supportive attitudes at home, worse health, or other issues that negatively impact her productivity that the firm cannot observe *ex ante*. Either of these scenarios could result in employers attracting a lower productivity pool of workers after increasing the amenity-share of worker compensation.

I study this question in Jordan where this issue is particularly salient. Jordan has one of the lowest female labor force participation rates in the world – lower even than Saudi Arabia – and over 45 percent of Jordanian women cite transportation or childcare as the biggest barrier to women's labor force entry, with just half that proportion citing a cultural barrier (Arab Barometer, 2022). In short, many women desire to work but are unable to find suitable work. Is it the case that low-productivity workers disproportionately value certain amenities? If this adverse selection does exist in the labor market, is it severe enough to make offering certain amenities unprofitable? What policy interventions can increase women's access to jobs and increase welfare?

This paper combines administrative records of female workers' productivity at a Jordanian firm with survey data on those workers' willingness to pay for real-life job amenities including transportation to and from home, childcare, and preferred pay schemes. I show that the least productive workers have the highest willingness to pay (WTP) for three out of the four job amenities studied. The negative correlation is statistically significant and, in the case of transportation, generates enough adverse selection to make offering these job amenities to new hires unprofitable.

A key contribution of this study is its empirical setting: a firm where worker productivity is directly measurable and where it is feasible to elicit workers' willingness to pay (WTP) for job amenities. The sample consists of survey enumerators, all of whom are women due to cultural norms, whose productivity can be precisely quantified given the nature of their work. Detailed records on the speed and quality of data collection across multiple years of projects provide rich measures of individual pro-

a job amenity which induces positive selection on worker productivity.

ductivity. I pair this productivity information with survey-based WTP data, elicited through incentive-compatible multiple price list experiments covering four amenities identified as desirable in qualitative interviews: childcare, transportation, fixed pay (versus performance pay), and regular monthly pay (versus a lump-sum payment at the end of data collection). The transportation amenity involved a trusted driver picking the worker up from their home instead of a central meeting point (the status quo) for field interview work, which saves women from lengthy commutes on buses or expensive taxis. The childcare amenity was a voucher that could be used at a childcare center of the worker’s choosing. I find the largest negative correlation for transportation: a 1-SD increase in WTP is associated with 0.056 SD lower productivity. WTP for fixed daily pay and monthly paychecks are also statistically significantly negatively correlated with productivity; WTP for childcare is negatively correlated as well, but not statistically significantly so.

Further analysis reveals that this negative correlation is indeed consistent with adverse selection, especially for transportation. Controlling for variables observable to the firm does not eliminate it, and in the case for transportation, even detailed covariates on income, attitudes, and information on commute cost, duration, and safety cannot fully explain the negative correlation. The point estimate also remains significant after controlling for tenure with the firm, ruling out the possibility that workers are equally productive at baseline but high WTP workers quit before gaining productivity-enhancing experience. This would generate a negative correlation in the cross section but would be inconsistent with the notion that offering transportation would attract less productive workers.

Still, the existence of adverse selection is not enough to conclude that offering transportation instead of higher wages would lower firm profit. While I show that offering transportation and lowering wages would induce adverse selection, I also document large WTP. The average worker in the sample is willing to forgo a 55 percent increase to their wages to work in a job that includes transportation, and 59 percent of workers have a WTP above the firm’s expected cost to provide transportation to them. This has two key implications. First, the firm may be able to lower wages by more than the cost of provision, potentially offsetting the reduced productivity. Second, offering job amenities might dramatically increase labor supply, especially in this context where female labor supply is constrained by lack of access to transportation. This mechanism would also increase profit in a monopsonistic setting. Therefore, a model is needed to disentangle the net impact of offering transportation. Nekoei (2024) develops a model where job amenities can induce selection, but the model is focused on sorting dynamics of workers between firms and has no labor supply margin, making it insufficient for studying the impact of adverse selection when job amenities can generate labor supply responses.

To determine the net impact of offering transportation on firm profit, I develop and estimate a monopsony model that allows job amenities to increase total labor supply and induce adverse selection. Workers accept the firm’s offer if their utility from the wage and amenity (if included) is greater than their outside option, and workers’

valuation of the amenity is allowed to covary with productivity. Firm profit is given by total worker productivity minus total labor costs (wages and amenities). A monopsony model is appropriate in this setting since women have relatively few alternative sources of employment outside of teaching, especially in rural areas. As in all monopsony models, the firm makes positive profit for each employed worker, so the firm can increase profit in three main ways: increase employment, attract more productive workers, or lower expenditure on compensation. As a result, adding amenities and lowering the wage to offset the cost of provision could increase profit if doing so attracts more productive workers. On the other hand, if offering amenities attracts less productive workers, the firm would need to lower the wage by more than the cost of provision to maintain per-worker profits.

Estimating the model requires knowledge of the unselected joint distribution of productivity and WTP from a pool of potential workers. The data used here allows me to identify the joint distribution of productivity and amenity valuation *among workers*. To account for selection into working at the firm, I leverage a rich set of covariates along with information on applicants who apply but do not accept job offers. Specifically, I estimate willingness to pay and the likelihood of accepting an offer as a function of applicant characteristics, then reweight estimates to adjust for differences in covariate distributions between workers and non-workers.

The second estimation challenge I address is identifying the outside option distribution. Directly measuring the utility value of the outside option is not possible in this setting, partially because the relevant outside option for many in the sample is being a housewife. Instead, I make parametric assumptions on the outside option distribution following Kline (2025). Outside options are assumed to be distributed according to a power function and are assumed to be independent from productivity and amenity WTP. This makes the model tractable without individual level outside option data and focuses on the adverse selection mechanism. I estimate the parameters of the outside option distribution by leveraging reservation wage data from similar Jordanian women in a representative labor market survey. Reassuringly, the estimated parameters generate a distribution similar to the distribution of reported wages in the enumerator sample.

To estimate counterfactual firm profit, employment, and wages if the firm offered transportation to all workers, I express two main identifying assumptions as moment conditions and solve for counterfactual outcomes using Generalized Method of Moments (GMM). The identifying assumptions are 1) earnings from piece rates are proportional to latent productivity and 2) the firm is currently profit maximizing. To account for estimation error in the outside option parameters which enter into the moment conditions, inference is based on a nonparametric bootstrap.

I find that offering amenities would reduce average productivity of employed workers but expand employment, leaving firm profits lower. Increased employment and more amenities that are highly valued by the worker results in higher total welfare for the population of potential workers. Because the difference in profits with and without amenities is relatively modest, I also find that a subsidy of only 6 percent of trans-

portation costs is needed to make it more profitable for the firm to offer transport, generating a 36 percent increase in worker welfare.

I next evaluate a second counterfactual where the firm allows workers to choose their preferred contract: high wage without transportation or lower wage with transportation. This is equivalent to allowing workers to buy transportation from the firm for a price. If workers could reveal themselves as having high WTP by selecting the low wage contract, the firm could potentially increase profit and worker utility. To determine whether this two contract counterfactual could strictly dominate the others, I estimate the profit maximizing pair of wages and associated profit, employment, productivity, and worker utility. I find that by allowing workers to sort into their preferred contract, the firm can attract those with high WTP with the transportation amenity and attract those with low WTP with a higher wage. The result is higher employment and worker utility than in the amenity mandate counterfactual and higher profit than in the status quo.

These findings have several policy implications. First, mandating that firms offer all workers transportation in a setting without strong enforcement is not likely to be effective since it results in lower firm profit, but a relatively small subsidy may be effective in settings where worker WTP is high and/or the amenity could have large labor supply impacts. Second, the finding that offering workers bundles with and without amenities can increase profit and worker welfare suggests that there may be returns to helping firms experiment with different types of compensation packages. Short-term policies that encourage exploration of different job amenity schemes may help firms find new ways to compensate workers that better fit workers' needs and increase firm profitability. Third, policymakers interested in increasing women's labor market access should consider prioritizing transportation investments. WTP for transport was larger than childcare even among women with children, and mechanically, all women (and men) must commute while childcare is only needed in a particular phase of life. This is especially true in safe and communally-oriented settings like Jordan where it is not uncommon for older children, extended family, or neighbors to look after school-age children.

Section 1.2 describes the setting, with context on Jordan, the firm, and the job amenities studied. Section 1.3 describes the administrative records on worker productivity and the methods for measuring willingness to pay for amenities, then presents descriptive statistics. Section 1.4 describes the empirical strategy for estimating the relationship between WTP and productivity, and section 1.5 reports the results. Section 1.6 develops the monopsony model that allows job amenities to induce selection to interpret the results found in the previous section, then fits the model to data on worker preferences and productivity to discern how different counterfactual policies might impact outcomes like firm profit, employment, wages, and worker welfare. Section 1.7 reports the results of this counterfactual analysis. Section 1.8 discusses policy implications and section 1.9 concludes.

1.2 Institutional Background

1.2.1 Jordan

Jordan is located in the Middle East and bordered by Saudi Arabia, Syria, Iraq, Israel, and the West Bank. It is politically stable and classified as lower-middle income by the World Bank (Metreau, Young, and Eapen, 2024). It experiences relatively high rates of unemployment (ranging from 10 to 20 percent), and 35 percent of employment is in the informal sector (World Bank, 2025b). Approximately half of the population has completed secondary school, and 20 percent have some kind of post-secondary education. The gender gap in education has shrunk dramatically in the past several generations, with women under 40 currently attaining more education than their male counterparts.³

Despite near-equal rates of education between the sexes and gender norms that are comparable to, if not more liberal than peer countries in the MENA region, Jordan ranks in the bottom third in the region with respect to the ratio of female/male labor force participation, and ranks 6th lowest in the world in female labor force participation (World Bank, 2025a; ILO, 2025). While cultural norms contribute to these low rates of employment (Assaad, Krafft, and Selwaness, 2022; Bursztyn, González, and Yanagizawa-Drott, 2020; Jayachandran, 2020), logistical barriers also play a large role. Approximately 40 percent of Jordanian respondents (male and female) reported in the 2020-2021 wave of the Arab Barometer that the biggest barrier to women entering the workplace in their country was lack of access to transportation or childcare (higher than the regional average of 34 percent), and an additional 24 percent cited low wages (Arab Barometer, 2022). Less than a quarter cited a cultural reason, such as working being socially unacceptable. In fact, female respondents were far more likely to cite structural instead of cultural reasons as barriers to women’s employment than men: 74 percent of female respondents in Jordan cited lack of transportation, childcare or low wages as the biggest barrier. For these reasons, Jordan is a particularly salient context in which to study how job amenities impact labor supply and worker selection, especially among women.

1.2.2 Research Firm

The research firm I partnered with is the largest research firm in Jordan, and handles many of the large surveys conducted by the government, UN agencies, and academics. The firm employs 30-40 full time formal staff who manage projects, data, finances, and human resources. The firm also hires hundreds of temporary informal workers to work as survey enumerators, drivers, and other staff who interact with respondents to schedule and conduct interviews. These survey enumerators comprise the research sample.

³Statistics on formality and educational attainment are sourced from the 2016 Jordanian Labor Market Panel Survey (OAMDI, 2018).

In most developing countries, using survey enumerators to collect data from respondents remains the norm. Survey enumerators typically have higher levels of education so that they are able to help respondents respond accurately to questionnaires. In this setting, nearly all enumerators are women due to both social and logistical considerations. Many female respondents would feel uncomfortable being alone with a male enumerator. Even when the respondent is male, the presence of a male enumerator may require women in the household to wear a hijab or otherwise adjust their dress to receive a male guest. Female enumerators avoid these issues.

Enumerators are hired on a project-by-project basis, so employment is temporary and informal. At the start of each project, all enumerators are trained on how to administer the questionnaire and other relevant tasks like using the tablet and gaining consent ethically. Phone surveys are either conducted at the enumerator's home or at the central office in Amman, depending on the project. For projects with in-person surveys, the firm recruits enumerators that live in the vicinity of respondents. Enumerators are responsible for traveling to a central meeting point, from which a logistical coordinator drives 2-3 enumerators to respondents' homes.⁴ Because addresses are often not precise, the coordinator usually communicates with the respondent to locate the home. The coordinator will also be in touch with the project manager throughout the workday as needed. The enumerator then knocks on the respondent's door, gains consent, and conducts the survey inside the respondent's home or nearby. Introducing oneself and the survey succinctly before gaining consent requires considerable social and communication skills. Administering the surveys quickly and accurately requires attention to detail, good communication skills, patience, and a strong working memory, in addition to sufficient understanding of the meaning of survey questions. Sensitive or long surveys also benefit from strong social skills on behalf of the enumerator.

Hiring

Individuals apply to the firm to work as a survey enumerator using a short online form, often accessed via a QR code posted in public places. The form elicits basic demographic information including name, gender, and age, as well as educational attainment and months/years of experience as an enumerator. The firm has a minimum education requirement for new enumerators of at least a post-secondary degree (i.e., associates or bachelors).

When hiring new applicants, the firm screens on geographic fit with the project's needs and more experience is favorable, but conducts no additional screening. There are no interviews, references, etc. Therefore, productivity conditional on location and experience is unobservable at the time of hiring. Table 1.1 provides supporting evidence of this claim. After conditioning on the information contained in the application (age, education, location, and past experience), there is no difference in the probability of being contacted by the firm by marital status, number of children, Big 5 conscien-

⁴In the case of random block sampling, the coordinator drives the enumerator to the correct location and the enumerator knocks on doors on foot to invite potential participants to the survey.

tiousness, perceived safety on public transport, or the extent to which transportation is a barrier to employment.⁵

Once the firm contacts applicants, applicants may accept or decline the job. This is the main source of selection which may differentiate workers from applicants, and therefore may bias estimates of the joint distribution of productivity and WTP. I correct for this potential source of bias using inverse probability weights, described in section 1.6.2.

Therefore, the firm’s database of applicants can be split into 3 categories: individuals who have applied and not (yet) been contacted, individuals who were contacted but declined, and individuals who have worked on at least one project at the firm. I collected survey data on individuals from each of these 3 categories.

1.2.3 Amenities

Willingness to pay was elicited for 4 amenities: transportation to and from home instead of a central meeting point for field projects, a childcare voucher for the duration of the project to be used at a center of the worker’s choosing, regular monthly paychecks instead of lump sum pay at the end of the project, and fixed daily pay instead of piece rate pay. These amenities were selected to be relevant to the specific job, desired by workers, and feasible to provide to be incentive compatible.

Transportation was identified as a major issue in qualitative interviews, especially for workers living in rural areas, and there was broad interest in childcare among mothers with young children. Workers also expressed that waiting for pay until a project ended is burdensome especially when the household had urgent financial needs, and there was agreement that monthly paychecks were preferred since this is the standard paycheck frequency in Jordan.⁶ Workers also disliked being paid per survey completed since it makes pay unpredictable. While performance is related to effort and skill, sometimes respondents simply do not want to consent to be surveyed, as is their right, or do not pick up the phone, resulting in lower output.

1.3 Data

1.3.1 Firm Data

The firm provided several datasets related to worker productivity as well as deidentified information from the application database. The productivity data fall primarily in two categories: supervisor evaluations and survey metadata.

⁵These outcomes were selected to be as exogenous to employment as possible, since being contacted increases the probability of being employed and the outcome data is measured ex post. For example, this test fails on labor market experience since being contacted increases the probability of employment which increases ex post labor market experience.

⁶The firm’s practice of paying enumerators after data collection is complete is likely related to the timelines with which grants are disbursed.

Supervisor evaluations

After each project, each enumerator receives a score on their performance on a scale of 0-5 from the Operations team and from the Data Processing team. The Operations team is in charge of day-to-day management of enumerators and provides a score based on the enumerator’s professionalism, perceived competence, the number of surveys conducted, and any other information the manager has on the enumerator’s performance. The Data Processing team assigns a score based on data quality assessed during the data cleaning process. The firm maintains a database of these scores and consults the enumerator’s average evaluation when making future hiring decisions.

The firm provided me with enumerator-by-year averages of the Operations evaluation score and the Data Processing evaluation score from 2020 onward, when they began keeping formal records of these scores.⁷ I standardize these enumerator-by-year averages into z-scores within year. For example, the data might report that a particular enumerator’s average Operations evaluation score was 0.2 SDs above the average in 2022. Standardizing allows me to combine various measures of productivity to generate a composite measure of an enumerator’s productivity.

Enumerators whose performance falls well below expectations or who commit fraud or other unethical behavior are added to a “blacklist” and will not be hired again. The firm also has a policy of adding enumerators who are extremely tardy to training sessions to the blacklist as a high-powered incentive to promote punctuality. Based on analysis triangulating the dates on which individuals were added to the blacklist against the dates of projects they worked on, it appears that at least half of individuals were added to the blacklist due to tardiness, not poor performance. For this reason, the blacklist outcome is included in robustness, not the primary specification.

Survey Metadata

The firm also provided a dataset of survey metadata which sheds light on enumerator quality by describing the data she collected directly. The firm’s standard practice is to code the client’s survey and host the survey on their own CTO server.⁸ The firm prepared metadata from each survey stored on their server dating back to 2020. Each observation from this dataset corresponds to a survey completed, and contains non-identifying information about that survey, including the enumerator who conducted it, the project it was associated with, the date and time the survey was conducted, the duration in seconds, and the callback accuracy where available. Callback accuracy measures data reliability. The firm’s standard practice is to call back approximately 10 percent of respondents within a week or so of the survey to ask them 5-10 short questions from the survey whose responses should not have changed (for example, date of birth or educational attainment). Low callback accuracy suggests the enumerator is

⁷It was not possible to access project-level evaluations.

⁸Clients may elect to host surveys on their own CTO server. For example, this project was hosted on a private server to ensure the firm did not have access to identified worker surveys.

asking questions incorrectly or fabricating data. From the date of survey completion, I construct measures of the total number of surveys conducted by the enumerator on a given project and the average number of surveys conducted by the enumerator per day on the project. Conducting many surveys per day at a steady pace is preferred. This dataset also includes the score the enumerator earns on the project’s training exam, which assesses whether or not she understands the correct way to enumerate survey questions, and includes administrative information about the project such as whether the survey was phone-based or in the field and whether it was paid per day or per survey.

I generate z-scores for each productivity outcome listed above (total number of surveys, surveys per day, callback accuracy, and exam score) reflecting the enumerator’s performance within a project on each outcome. Standardizing within project is necessary since surveys vary in length, difficulty, and nature from project to project. The resulting dataset reports multiple measures of each enumerator’s productivity on a given project where available. For example, the data may report that one enumerator scored 0.5 SDs above the mean in terms of survey quantity and 0.1 SDs above the mean in terms of exam score on her first project, and 0.25 SDs above the mean in terms of callback accuracy on a second project. Figure 1.1 reports histograms of each of the standardized measures of productivity, which will be stacked to form a composite measure of worker productivity.

Applicant Database

The firm also shared the deidentified applicant database, which includes information on age, governorate of residence, educational attainment, prior experience, and the job(s) they applied to. This information is available for applicants and workers, and comprises the universe of administrative data that the firm has about an applicant before hiring.

Cost of Amenity Provision

Finally, the firm provided estimates of the cost to provide each of the amenities as described in the survey. To estimate the cost of the transportation amenity, the firm estimated how much the daily wage paid to drivers would need to be increased to pick up enumerators from their home instead of a central meeting point. For driving in the same governorate (typical), the additional cost was 12 Jordanian Dinars (JD)/day. For driving longer distances, the additional cost was 17 JD/day. In the status quo, drivers pick up 2-3 enumerators from a central meeting point and transport them to respondent’s homes. To calculate the cost/worker/day of the transportation amenity, workers who live in an urban area are assumed to have a cost of 4.8 JD (12 JD divided by 2.5) and rural workers assumed to have a cost of 6.8 (17 JD divided by 2.5). These cost estimates will be used in the model estimation described in section 1.6.

When asked to estimate costs to run a daycare for 12 students, the firm included

rent, the wages of 2 teachers near the minimum wage and a higher paid principal/manager. Including snacks and supplies, the cost per student per day cost is between 10-12 JD. This is much higher than the average cost of 3.5 JD/day/student reported by daycares in Jordan in a 2015 survey of childcare facilities. Adjusted for inflation, 3.5 JD in 2015 is equivalent to 4.15 JD in 2024. The differences in cost likely partially reflect the fact that the survey was geographically representative and rent/wage costs are lower outside of Amman, and that private daycares are operating at a larger scale.

The marginal cost to pay a single worker’s paycheck monthly instead of as a lump sum was quoted to be 0 JD, however the firm estimated that it would need to hire one additional full time worker on the finance team to pay all workers monthly instead of in lump sums. Dividing a full-time worker’s salary by the number of enumerators working in a given month and the number of working days in a month yields an amenity cost around 0.5-1 JD/worker/day.

1.3.2 Elicitation of Willingness to Pay for Job Amenities

The survey which elicited WTP for job amenities and measured other relevant covariates was conducted over the phone during the summer of 2024. Workers and applicants were invited to participate in randomized order until the total number of surveys was completed. Approximately 50 percent of individuals who were invited to participate consented and completed the survey. Enumerators conducting the WTP survey were employed by the same firm but excluded from the sample.⁹

The first 4 sections of the survey measured willingness to pay for 4 job amenities separately in randomized order: transportation, childcare, regular monthly paychecks instead of lump sum payment at the end of the project, and daily wages instead of piece rates. Each section implemented a series of discrete choice questions akin to a multiple price list where the respondent is asked to choose between job A, which pays 15 JD per day and includes the amenity, or job B, which pays x JD per day and does not include the amenity, where $x \in \{15, 30\}$.¹⁰ The respondent is told that the jobs are otherwise identical. The wage of job B is varied until her indifference point is revealed, which identifies her WTP.

The survey design took care to avoid order effects, adaptive bias, and multiple/never switching issues while remaining minimally tedious over the phone (see Appendix A.2 for more details). For each amenity, the survey first asked 4 questions, increasing or decreasing the pay of job B in increments of 5 dinar with no adaptiveness. Avoiding adaptiveness early on in the elicitation avoids anchoring on a mistake as described by Jack, McDermott, and Sautmann (2022). Whether the list was increased or decreasing was randomized. For example:

⁹Enumerators conducting the survey were also first-time hires to minimize the probability that an enumerator had ever been the coworker of a respondent.

¹⁰The price list for transportation was expanded up to 40 JD since in-person piloting revealed very high WTP for transportation.

If job A pays **15** JD/day and includes transport, and job B pays **15** JD/day without transport, which do you prefer?
 If job A pays **15** JD/day and includes transport, and job B pays **20** JD/day without transport, which do you prefer?
 If job A pays **15** JD/day and includes transport, and job B pays **25** JD/day without transport, which do you prefer?
 If job A pays **15** JD/day and includes transport, and job B pays **30** JD/day without transport, which do you prefer?

Assuming there are no logical inconsistencies in her choices, the survey then asks questions adaptively in 1JD increments until her WTP is determined within a 1 JD range. For example:

If job A pays 15 JD/day and includes transport, and job B pays **15** JD/day without transport, which do you prefer? **A**
 If job A pays 15 JD/day and includes transport, and job B pays **20** JD/day without transport, which do you prefer? **A**
 If job A pays 15 JD/day and includes transport, and job B pays **25** JD/day without transport, which do you prefer? **B**
 If job A pays 15 JD/day and includes transport, and job B pays **30** JD/day without transport, which do you prefer? **B** Begin adaptive section:
 If job A pays 15 JD/day and includes transport, and job B pays **23** JD/day without transport, which do you prefer? **B** If job A pays 15 JD/day and includes transport, and job B pays **22** JD/day without transport, which do you prefer? **A**

At the end of the elicitation, the enumerator confirms with the respondent her choices to make sure it aligns with her preferences and that no mistakes were made. *“So, when job B pays between 15-22 JD you preferred job A (15 JD with transportation), but when job B pays 23 or more JD, you prefer job B. Is that correct?”* This implies that she is indifferent between job A and B when job B pays between 22-23 JD, or 7-8 JD more than job A. This implies her WTP for transportation is between 7-8 JD. The analysis that follows takes the midpoint, or 7.5 in this example.

The consent script explained that participants would receive a small incentive for participation (3 JD) and would also be entered into a random draw to receive the job (A or B) that they chose in a randomly selected question if they were selected by the firm to work again. The high take-up rate of a relatively long phone survey suggests that the incentivization was well understood.

At the start of each amenity section, several sentences of relevant details about the job and the amenity were provided. The transportation, childcare, and paychecks sections described the project as lasting for two months and having a daily target of 5 surveys. Two months was selected to be long enough to be worth adjusting one’s schedule for, and to be a reasonable comparison for lump sum vs monthly pay. A project with a target of 5 surveys per day connotes a survey that is of medium length and complexity. Childcare was described as a voucher to be used at a childcare center of their choosing for the duration of the project, and transportation was described as a trusted driver affiliated with the firm driving them to/from home instead of a meeting point. The monthly paycheck amenity was described as guaranteed regular monthly paychecks at the end of each month instead of a lump sum payment at the end of the project, the status quo. For the daily pay amenity, the project was described as lasting 2 months and having a target of 3 surveys per day with extra information like the

sampling method, expected range of survey duration, and the transport time between respondents to help the respondent gauge her expected productivity accurately.¹¹ The full text of the survey instrument is available in appendix A.4. Attention questions were included to ensure the respondent internalized the relevant details of the job. Between 85-90 percent of respondents answered these questions correctly the first time.

Following the WTP sections, the phone survey included the following modules: demographic information (including age, location, education, marital status and living arrangements, number and age of children, etc.), commuting arrangements, childcare arrangement and attitudes towards different childcare arrangements, basic time use in the past 7 days, employment history, attitudes and motivations towards employment, household income, savings and loans, and gender attitudes. Two additional modules were included to potentially predict productivity: a digit span test of working memory and an Arabic-validated Big 5 personality test (Hussein and Abdel-Khalek, 2021).¹²

1.3.3 Descriptive Statistics

Table 1.2 reports demographic information for the study sample, which includes all applicants and workers who participated in the survey, and reports the most recent representative labor force survey in Jordan for comparison. Demographic data in this table for the study sample is sourced from the survey. The population comparison data includes all Jordanian women aged 18-65 in the 2016 Jordanian Labor Market Panel Survey (JLMPS), and restricts to women employed or searching for work for comparisons to the labor force. The women in this study are of similar age and marital status as the average woman in the labor force in Jordan. The women in the study sample are more educated than the population due to the job requirements, but the difference in educational attainment between the study sample and the population of women in the labor force is much smaller than the population at large. This paper's sample is slightly more rural than the population, and has employment rates which are lower than the average woman in the labor force but higher than the average woman.

Table 1.3 reports summary statistics from the survey describing each of the three categories of individuals in the sample: workers, applicants who declined, and applicants who have not been contacted. Comparing workers to individuals who declined the job characterizes the type of selection into the productivity dataset that will be corrected for. In particular, individuals who decline are more likely to be from the capital city, more likely to be career oriented, and less likely to be financially insecure. These facts are suggestive that women who decline have better outside options.

Willingness to pay for all of the amenities is on average high. Figure 1.2 reports the

¹¹To be as realistic as possible, these facts were based on a real survey previously conducted by the firm with which I was affiliated.

¹²Enumeration requires listening to respondent's responses attentively and remembering/recording accurately, which is related to working memory. Successfully gaining consent from cold-call or door-knock respondents requires high levels of extroversion and agreeableness, and conscientiousness is likely correlated with careful data collection.

average willingness to pay reported in percentage terms of the base wage separately for workers, applicants who turned down a job, and those not yet contacted. Willingness to pay reported in percentage terms can be understood as follows: a respondent who preferred job A (pays 15 JD + amenity) when job B paid 19 JD but then switched to job B when job B paid 20 JD has a WTP between 4-5 JD. Taking the midpoint, their WTP is 4.5, which is 30% of the base wage. The average worker is willing to forgo a 57 percent increase to their wages to work in a job that includes transportation, 43 percent for childcare (among mothers), 35 percent for regular monthly paychecks instead of a lump sum, and 29 percent for daily wages. These means are high, but conceal a great deal of variation across person, displayed in the histograms in Figure 1.3. Childcare and regular, monthly paychecks have distributions which are bimodal. This bimodality is consistent with 10-20 percent of workers having no need for these amenities, while another 10-20 percent of the population are so constrained by childcare issues or liquidity constraints that they prefer the amenity to a doubling in wages. WTP for transportation has more variation across the distribution but also features a thick right tail. 8 percent of workers prefer transportation to an increase to wages by 150 percent. Finally, most workers prefer daily wages to piece rates, but approximately 5 percent are indifferent, and nearly 30 percent have some degree of preference for performance pay.

The other stylized fact revealed by the summary statistics on WTP is that applicants have higher WTP than workers for all amenities. The difference between WTP as a percent of base pay between workers and applicants ranges from 20 to 30 percentage points. Additionally, applicants who turned down a job have either higher or statistically indistinguishable WTP for all amenities than workers and those not yet contacted. Given that the current job does not offer these amenities, this pattern is consistent with higher-WTP individuals declining the position because it lacks the amenities they value.

Next, I characterize what predicts high WTP for each of the 4 amenities studied to understand what drives variation in WTP. Table 1.4 reports coefficients from separate regressions of WTP for each amenity on relevant covariates, controlling for whether or not the respondent is worked for the firm, whether or not they've been contacted, and governorate. Many of these relationships are sensible: for example, higher commute costs are associated with higher WTP for transportation, increased financial strain is associated with higher WTP for regular paychecks, and the child's grandmother being unavailable for childcare increases WTP for the childcare amenity. The following paragraphs describe predictors of WTP for each amenity in more detail.

Transportation. WTP for transportation is higher among rural women, women with higher commute costs, and those who feel unsafe on public transportation more often. The safety relationship is statistically and economically significant, though it should be noted that public transportation is relatively safe for women in Jordan. 63 percent of respondents reported always feeling safe on public transportation, so this

effect is driven by a minority of the population. WTP for transportation is also 1.3 JD (11 percent) higher among women who report that transportation is the largest barrier preventing women in Jordan from entering the labor force. The length of respondents' commutes to their job with the firm was uncorrelated with WTP. Finally, income is weakly negatively correlated with WTP for transportation, meaning that firm-provided transportation was more attractive to lower income women. This reflects the fact that low-income women's commutes are longer and/or more tiring. A negative correlation between income and WTP is the opposite of what one would expect if WTP was mostly driven by income effects.¹³

Childcare. WTP for childcare is uncorrelated with household income, how much women currently pay for daycare (if using), or even how close the nearest acceptable daycare is to their home. Being willing to use paid childcare, citing childcare as the primary barrier to women's labor force entry, and the respondent's mother or mother-in-law being unavailable for childcare are highly correlated with WTP. Demand is also higher among women with more conservative gender attitudes and women whose male family members disparage women who work, suggesting that firm-provided childcare could make working outside the home more palatable for more conservative households. This finding aligns with McKelway (2025) who finds that an intervention emphasizing the female-friendliness of a job to women's families increased women's employment. WTP is also higher among women who are currently working but want a different job in the next two years (either different employer or different job within the same employer), and women who have not worked in the past 3 months. Taken together, these facts show that 20–30 percent of mothers report little interest in childcare, but demand is high among the rest, for whom firm-provided childcare could ease frictions that limit their ability to work as much as they would like. Indeed, over 75 percent of all mothers with young children in the sample reported they would work longer hours, open a business, or begin working again if they had access to childcare even for a fee.¹⁴

Regular Monthly Paychecks. WTP for monthly paychecks instead of lump sum payments is higher among those with little savings and no one from whom they can borrow money easily. It is also higher among those who reported they had to reduce expenditure or take out a formal loan the last time their paycheck was late, and those who have a formal loan.¹⁵

¹³Low income women presumably have higher marginal utility of cash, which would lower WTP.

¹⁴The wording of this question follows a recent World Bank survey on childcare in Jordan, Iraq, and Lebanon (World Bank, 2023). The share of women in the representative sample in the World Bank survey reporting the same is extremely similar, at 73 percent. The question simply refers to "paid childcare" and does not specify a specific fee amount.

¹⁵Loan interest and delinquency was measured, but is difficult to measure in this setting and largely irrelevant since Islamic finance forbids time-accruing interest. Individuals who take out a loan pay back more than the principle, but missing an installment does not increase the amount of debt in the way it does if interest is time-accruing. Banks are not permitted to profit from late fees.

Fixed Wages. The main predictors of willingness to pay for fixed wages include measures of labor force attachment, experience with the firm, and the respondent’s motivations for work. Women with more labor force attachment and more attachment to the firm are less averse to performance pay, while those motivated primarily by earning money are *more* averse to performance pay, presumably because it increases variability in income.

All Amenities. Beyond demographic and household characteristics, personality traits also help explain variation in willingness to pay. In particular, higher scores on the Big Five neuroticism scale are associated with higher WTP for childcare and regular monthly paychecks. The coefficient for transportation is similarly large, though only marginally significant ($p = 0.106$). Neuroticism measures the tendency to feel negative emotions like anxiety, depression, and anger more often and more intensely, and was elicited using an Arabic-validated version that primarily includes items reflecting symptoms of depression and anxiety (e.g., “I feel sad,” “I feel anxious,” “I am a lonely person”). These patterns are not accounted for by observable characteristics such as income, education, location, or the age of children. Taken together, they suggest that individuals more prone to negative emotions or mental health challenges particularly value job amenities that reduce stress or uncertainty at work. However, none of the personality traits—including neuroticism—mediate the relationship between productivity and WTP.

1.4 Empirical Strategy

To examine the relationship between worker productivity and willingness to pay (WTP) for job amenities, I estimate

$$z_{ij} = \beta_0 + \beta_1 WTP_i + \varepsilon_{ij}, \quad (1.1)$$

where z_{ij} denotes standardized productivity of individual i on measure j (for example, her callback accuracy on the third project she worked on), and WTP_i is the individual’s stated willingness to pay for a specific amenity. Standard errors are clustered at the individual level.

Each productivity measure is standardized within its relevant reference group to have mean zero and standard deviation one. The preferred specification includes: (i) average surveys completed per day, (ii) total surveys completed per project, (iii) callback accuracy, (iv) training-exam score, (v) annual average of operations evaluations, (vi) annual average of data-processing evaluations, and (vii) the overall assessment score. These outcomes are chosen because higher values unambiguously indicate stronger performance. Robustness checks additionally include survey duration and inclusion on the firm’s “blacklist,” recoded so that higher values denote better performance.

Each worker contributes multiple productivity measures, drawn both across projects and across different indicators within projects. I pool all available measures to exploit this variation rather than averaging within individual. The stacked specification is algebraically equivalent to regressing the individual mean of productivity, $\bar{z}_i$, on WTP_i weighted by the number of productivity measures per worker. Under the assumption that each measure provides an equally noisy, independent signal of latent productivity, this approach yields a more precise estimate of the relationship between productivity and WTP than the unweighted regression of $\bar{z}_i$ on WTP_i .

WTP is measured in Jordanian dinars using the incentive-compatible multiple-price-list elicitations described earlier. The main specification uses the censored version of WTP, which top-codes individuals whose valuations exceed the maximum amount in the elicitation; results are similar using the uncensored measure that incorporates open-ended responses.

The coefficient β_1 summarizes the cross-sectional relationship between productivity and the value placed on job amenities. A negative estimate indicates that less productive workers report higher WTP for the amenity, consistent with adverse selection on productivity.

1.5 Results

Workers' willingness to pay (WTP) for key job amenities is negatively correlated with their productivity. Figure 1.4 illustrates this relationship for each amenity, plotting binned averages of productivity against WTP. In all panels, the slope of the best-fit line is negative, indicating that workers who value these amenities most highly tend to be less productive on the job.

Transportation. The strongest negative relationship appears for transportation. Table 1.5, column 1, shows that a one-Jordanian dinar increase in WTP for transportation is associated with a 0.007 standard deviation decrease in productivity. Given that the standard deviation of censored WTP for transportation is 7.34, a one-standard deviation increase in WTP corresponds to roughly 0.056 standard deviations lower productivity. To benchmark the magnitude, 0.05 SDs of productivity is equivalent to the difference between completing 63 and 65 total surveys on a particular project, or the difference between an average Operations score of 4.05 and 4.1 out of 5. This relationship is both statistically and economically significant: workers most willing to forgo wages for firm-provided transportation are, on average, those who perform less well across a wide range of productivity metrics. As discussed below, this pattern remains robust after controlling for all the characteristics the firm can observe and even commute characteristics it cannot observe.

Monthly paychecks. A similar, though slightly smaller, negative correlation emerges for preferences over pay frequency. Column 2 of Table 1.5 shows that a one-Jordanian

dinar increase in WTP for monthly paychecks is associated with a 0.006 standard deviation decline in productivity. Scaled by the variation in WTP, this implies that a one-standard deviation increase in WTP corresponds to approximately 0.03 standard deviations lower productivity. Workers with stronger preferences for regular pay appear to be somewhat less productive, and as discussed below, this correlation is only partially explained by financial precarity and liquidity constraints and remains significant after adding controls for these outcomes.

Fixed wages versus performance pay. Preferences over pay structure show a similar pattern. Workers who prefer fixed daily wages to piece rates tend to be less productive, with an effect size similar to that observed for monthly pay frequency (about 0.03 standard deviations per standard deviation of WTP). This finding aligns with the idea that less productive workers may prefer to avoid performance-based pay schemes that expose them to greater income risk or higher effort requirements.

Childcare. The negative relationship between WTP for childcare and productivity is smaller and statistically insignificant. The estimates are imprecise because this regression is limited to the subsample of mothers with young children, but the point estimate is also close to zero. This suggests that childcare demand primarily reflects life-stage needs rather than underlying personality or background differences that correlate with productivity.

Taken together, these results indicate that for most amenities—particularly transportation and regular pay—workers who value them most are those who are less productive. This pattern is consistent with adverse selection on unobserved worker productivity: amenities that reduce stress, uncertainty, or commuting effort appear especially attractive to workers whose productivity is lower on average. The next section examines whether these correlations are indeed consistent with the notion that offering amenities instead of higher wages would attract less productive workers.

1.5.1 Ruling Out Alternative Mechanisms

This section tests whether the negative association in the data is consistent with adverse selection. For this to hold, four alternative explanations must be ruled out. The first is that the negative correlation simply reflects selection into employment. If the correlation between WTP and productivity would not be negative in an unselected sample, then offering amenities would not induce adverse selection. Table A.1 reports results from regression equation (1.1) with the addition of inverse probability weights, based on the predicted probability of working at the firm conditional on being contacted using the detailed survey covariates, to correct for any selection into accepting work. Section 1.6.2 contains more information on the construction of these weights. Reassuringly, the weighted results are almost identical to the unweighted estimates. Under a selection-on-observables assumption, this rules out the concern that

the negative correlation is only the result of selective entry into the worker sample.

The second scenario to rule out is that the negative correlation is driven by a characteristic that the firm can observe and screen on, such as location. This would not be consistent with adverse selection because the firm could offer the amenity but avoid hiring less productive workers by screening on the observable characteristic. For example, if the correlation observed in Table 1.5 were entirely explained by rural women being less productive and having higher WTP for transportation, this would generate a negative correlation in the cross section but would not necessarily mean that offering transportation would result in less productive workers because the firm can observe and screen on location. Columns 1 and 2 in table 1.6 add controls to the regression on WTP for transportation for all of the characteristics that the firm observes about workers *ex ante*: location up to sub-district, educational attainment, year of birth, and industry experience. Because administrative experience data was not provided by the firm for all workers, results with and without this control are provided since it limits the sample size. The negative correlation between productivity and transportation is robust to all these controls and the magnitude is remarkably stable. Columns 4 and 5 report the same regressions with WTP for monthly paychecks. There also remains a negative correlation between WTP for monthly pay and productivity after controlling for all the information the firm collects about applicants *ex ante*. The negative correlation between WTP to avoid performance pay and productivity is not robust to these controls, and appears to be mediated by education and experience. This simply means that any correlation between productivity and WTP to avoid productivity pay is fully explained by education and experience, so the firm could screen on these characteristics to avoid adverse selection. Interestingly, the relationship between WTP for childcare and productivity flips sign and becomes positive with the addition of the administrative controls for industry experience. This suggests that between two women of the same age, location, education and experience, the woman with higher WTP for childcare (conditional on having children) may be more productive on average.

A third test rules out a scenario where experience effects and differential attrition combine to generate a negative correlation in the cross section not consistent with adverse selection. If enumerators become more productive with experience but workers with high WTP quit earlier and therefore do not become more productive, but would not quit earlier if the amenity were provided, this would generate a negative correlation in the cross section between WTP and productivity that is not consistent with adverse selection. To rule this out, I add fixed effects for whether the productivity outcome was measured on the enumerator's first project, second project, and so forth. This answers the question of whether high-WTP enumerators are less productive than low-WTP enumerators with the same amount of experience. Column 3 of Table 1.6 reports results with *n*'th project fixed effects for transportation, and the negative correlation survives this test, as does the negative correlation for monthly paychecks. This implies that the negative correlation between productivity and WTP for transport and regular paychecks is present even on workers' first project.

The final possibility to rule out is whether the amenity, if provided, could equalize

productivity. This is equivalent to ruling out that the amenity has a heterogeneous treatment effect that would disproportionately benefit those who value it most (and are currently less productive). This cannot be ruled out directly without experimental evidence, but I do the next best thing by adding controls for the characteristics that would most likely mediate the heterogeneous treatment effect. For example, if transportation would have a larger productivity-enhancing treatment effect on high-WTP workers, then logically, it must operate through differences in commutes. If controlling for commute difficulty does not eliminate the negative correlation with WTP to transportation, it seems unlikely that making commutes easier would make women with high WTP disproportionately more productive. Columns 1 and 2 of table 1.8 report the results from this exercise for transportation. Column 1 adds comprehensive demographic controls, household income controls, gender and worker attitudes, and personality characteristics, and column 2 adds further controls for her commute to the firm: length of commute, cost of commute, mode of transportation and perceived safety. The negative correlation is robust to these controls. This makes it implausible that worse commutes are the source of the negative correlation and therefore implausible that providing transportation would equalize productivity.

I conduct a similar exercise for monthly paychecks, since it is possible that financial stress makes workers less productive, but I do not find evidence that regular, monthly paychecks would equalize productivity (Kaur et al., 2024). While WTP for monthly paychecks is predicted by high levels of financial precarity, controlling for comprehensive household finance information, including financial precarity and liquidity constraints, does not attenuate the correlation between WTP and productivity. This finding makes it unlikely that financial stress is the cause of the existing productivity gap, and makes it unlikely that providing regular paychecks would close these gaps.

Having ruled out explanations inconsistent with adverse selection, the question remains: what explains the negative relationship between productivity and WTP for transportation and regular monthly pay? The previous section included all available controls, so the remaining explanations are necessarily speculative. One possibility is that an unobserved disadvantage lowers productivity and increases WTP for convenience-related amenities. Both transportation and regular pay reduce stress and logistical burdens, so poor physical health could plausibly mediate the relationship. The survey did not include detailed measures of physical and mental health, apart from the Big Five personality index, so this mechanism cannot be tested directly. Another potential explanation is heterogeneity in disutility of effort. Workers who find exertion more costly may both produce less and place higher value on stress-reducing amenities. Supporting evidence appears in Table A.2: while all workers become more productive under piece-rate pay, those with higher WTP for transportation or regular paychecks show the largest productivity gains. In other words, workers with high WTP respond more strongly to performance incentives, consistent with lower intrinsic motivation.

In conclusion, less productive workers have stronger preferences for transportation

and monthly paychecks for reasons that are almost entirely unobservable to the firm. If the firm were to lower wages in order to provide these amenities the average productivity of workers would likely fall. However, Section 1.3.3 also shows that women have extremely high willingness to pay for amenities, especially transportation. To the extent that women’s utility from working is continuous and that there are women on the margin of working and not working, these large WTP estimates imply that offering amenities like transportation could dramatically increase labor supply available to the firm. Depending on how large the labor supply response is, offering amenities could increase profit enough to outweigh the adverse selection mechanism. To estimate the net effect offering amenities would have on firm profit, total employment, and worker utility, I next turn to a monopsony model. I focus on transportation in the structural estimation exercise since transportation displays both the highest WTP and the most evidence of adverse selection, and is highly policy relevant. While childcare is also policy relevant, structural estimation is not needed to understand whether the firm might profitably offer it: average WTP for childcare is so much smaller than the firm’s stated cost of provision that it is implausible that any labor supply response might outweigh the impact of the direct cost of provision on firm profit in this setting.

1.6 Counterfactuals with Firm-Provided Transportation

The foregoing sections have shown that women in this setting have large WTP for firm-provided transportation, and also that there is considerable evidence that offering this amenity would induce adverse selection on worker productivity. To estimate the net effect of offering transportation to workers, I develop a monopsony model where the firm may use wages and/or amenities to grow large, and amenities can induce selection. I use the model to examine three main counterfactual policies, including i) mandating transportation, ii) subsidizing transportation and iii) allowing workers to purchase transportation from the firm.

1.6.1 Model

Workers

Let an individual’s utility from accepting an offer from the firm be given by

$$u_i(w, a) = w + v_i a - b_i$$

where w is the log of the offered wage, a is an indicator for the amenity, v_i is individual i ’s taste for the amenity, and b_i measures the value of the worker’s outside option in log wage units. This outside option may reflect either a different job or the utility value of being a homemaker. Individual i accepts the offer if $u_i(w, a) > 0$. Using this functional

form, v_i can be expressed as a function of the worker's WTP for the amenity and the reference wage against which WTP was elicited. For example, if a worker is indifferent between earning 15 dinars with transportation and 20 dinars without transportation, then the indifference implies $\ln(15) + v_i = \ln(20)$.

Define $\beta_i := \frac{b_i - \underline{b}}{\bar{b} - \underline{b}}$, where $\underline{b}$ and $\bar{b}$ are the lower and upper bounds of the outside option distribution, respectively. This transformation rescales each individual's outside option to the unit interval $[0, 1]$, simplifying the subsequent analysis. Assume $\beta_i \sim \text{Beta}(\phi, 1)$.¹⁶ Then the share of individuals accepting a job offering wage w and amenity level a — that is, the labor supply curve $F(w, a)$ — is given by equation (1.2), with full derivation provided in Appendix A.3:

$$\begin{aligned} F(w, a) &= \text{Prob}(u_i(w, a) > 0) \\ &= \left[\frac{w + v_i a - \underline{b}}{\bar{b} - \underline{b}} \right]^\phi \end{aligned} \quad (1.2)$$

Total utility to workers generated by this firm is given by the integral of utility above the outside option for those who chose to work. This can be understood intuitively as the worker surplus generated by hiring all workers willing to work at wage amenity pair (w, a) . This quantity is useful for assessment of the welfare impacts of different policies since it accounts for both employment effects and direct utility from wages and amenities. Expression (1.3) below allows for the calculation of the total utility generated by any wage and amenity combination. The detailed derivation is shown in Appendix A.3. Let $f_v(v)$ denote the probability density function (pdf) of v ; $f_{v,b}(v, b)$ the joint pdf of v and b ; and so on for other combinations of variables.

$$\begin{aligned} U &= \int_{-\infty}^{\infty} \int_{-\infty}^{w+va} (w + va - b) f_{v,b}(v, b) dv db \\ &= \int_{-\infty}^{\infty} \left[(w + va - \underline{b}) \cdot \left[\frac{w + va - \underline{b}}{\bar{b} - \underline{b}} \right]^\phi - (\bar{b} - \underline{b}) \frac{\phi}{\phi + 1} \left[\frac{w + va - \underline{b}}{\bar{b} - \underline{b}} \right]^{\phi+1} \right] f_v(v) dv \end{aligned} \quad (1.3)$$

Finally, workers in the model have fixed productivity θ_i , which is their revenue productivity. Fixed productivity rules out efficiency wages or productivity responses to amenities. For tractability, assume outside options are independent of the joint distribution of v_i and θ_i , or that $f_{v,\theta,\beta}(v, \theta, \beta) = f_{v,\theta}(v, \theta) f_\beta(\beta)$.¹⁷ Note that θ_i is productivity measured in currency units of worker product, but only productivity in

¹⁶This assumption is similar to the assumption made in Kline (2025), section 4.2. The main deviation from Kline (2025) is that wages enter utility in logarithmic rather than linear form. A linear specification would imply nontrivial employment at zero wage, which is not plausible in this setting.

¹⁷Because b_i is not observed in the study sample, some simplifying restriction must be imposed. This assumption is more plausible than it might be in other settings since many women's outside option is homemaking, and there is no strong reason to believe that utility from homemaking is correlated with productivity.

standard deviation units (z_{ik}) is observed. Section 1.6.2 explains how θ_i is estimated from the available data.

Profit Maximization

The firm chooses the dichotomous amenity a and the wage w to maximize its profits. The firm knows the joint distribution of worker productivity θ_i , amenity valuation v_i , and outside options b_i , but cannot observe this information for a given worker ex ante. The firm also knows ψ , the cost of providing the amenity. Recall that w is the log wage, so the wage in levels is $\exp w$. Then firm profit is given by:

$$\begin{aligned}\pi(w, a) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_0^{\frac{w+va-\underline{b}}{\bar{b}-\underline{b}}} (\theta - \exp w - \psi a) f_{v,\theta,\beta}(v, \theta, \beta) d\beta dv d\theta \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\theta - \exp w - \psi a) f_{v,\theta}(v, \theta) \int_0^{\frac{w+va-\underline{b}}{\bar{b}-\underline{b}}} f_{\beta}(\beta) d\beta dv d\theta \quad \text{by } v, \theta \perp\!\!\!\perp \beta \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\theta - \exp w - \psi a) \left[\frac{w+va-\underline{b}}{\bar{b}-\underline{b}} \right]^{\phi} f_{v,\theta}(v, \theta) dv d\theta\end{aligned}$$

Intuitively, firm profit can be increased by hiring more productive workers, lowering per worker labor costs, or increasing total employment (since the firm makes positive profit on each worker).

Since a is an indicator, the firm determines the profit maximizing wage with and without amenities and offers amenities if doing so generates more profit. Specifically, the profit maximizing wage for amenity level a (w_a^*) satisfies the first-order condition:¹⁸

$$\begin{aligned}\frac{d\pi}{d\exp w} &= 0 \\ \text{iff } \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} &-(w_a^* + va - \underline{b})^{\phi} + \phi(\theta - \exp w_a^* - \psi a) \frac{1}{\exp w_a^*} \cdot \\ &(w_a^* + va - \underline{b})^{\phi-1} f_{v,\theta}(v, \theta) dv d\theta = 0\end{aligned} \quad (1.4)$$

This wage first order condition will be used to solve the model.

Offering Workers Contract Choice

In this section, I relax the requirement that the firm set a single wage and amenity level. Suppose the firm offered workers a choice between a contract that paid a high wage without the amenity and a contract that paid a low wage with the amenity. This

¹⁸Following Kline et al (2025), the profit-maximizing wage can be shown to be unique since $\frac{d^2}{dw^2} \frac{1}{F(w,a)} > 0$

is equivalent to offering workers the choice to “buy” firm-provided transportation for a certain price. What wages would it set, knowing these wages would attract different types of workers? Could this generate higher utility or profit than offering only one contract?

Let w_h^* be the high log wage offered without the amenity and let w_l^* be the low log wage offered with the amenity. Then there are 5 mutually exclusive types of people: those who would work under either contract but prefer the high wage contract; those who would work under either but prefer the low wage with amenity contract; those who would only work under the high wage contract; those who would only work under the low wage with amenity contract; and those who would not work under either contract.

- Prefers high wage contract: $b_i < w^l + v_i < w^h$
- Prefers low wage contract: $b_i < w^h < w^l + v_i$
- Only works under high wage contract: $w^l + v_i < b_i < w^h$
- Only works under low wage/amenity contract: $w^h < b_i < w^l + v_i$

This can be simplified to the following:

- Works under high wage contract: $b_i < w^h \cap w^l + v_i < w^h$
- Works low wage contract: $b_i < w^l + v_i \cap w^h < w^l + v_i$

Then firm profit is the sum of profit generated by workers on each contract:

$$\begin{aligned}
\pi &= \int_{-\infty}^{\infty} \int_{w^h - w^l}^{\infty} \int_0^{\frac{w^l + v - \underline{b}}{\bar{b} - \underline{b}}} (\theta - \exp w^l - \psi) f_{v,\theta,\beta}(v, \theta, \beta) d\beta dv d\theta && \text{low wage contract} \\
&\quad + \int_{-\infty}^{\infty} \int_{-\infty}^{w^h - w^l} \int_0^{\frac{w^h - \underline{b}}{\bar{b} - \underline{b}}} (\theta - \exp w^h) f_{v,\theta,\beta}(v, \theta, \beta) d\beta dv d\theta && \text{high wage contract} \\
&= \int_{-\infty}^{\infty} \int_{w^h - w^l}^{\infty} (\theta - \exp w^l - \psi) \cdot \left[\frac{w^l + v - \underline{b}}{\bar{b} - \underline{b}} \right]^\phi f_{v,\theta}(v, \theta) dv d\theta \\
&\quad + \int_{-\infty}^{\infty} \int_{-\infty}^{w^h - w^l} (\theta - \exp w^h) \left[\frac{w^h - \underline{b}}{\bar{b} - \underline{b}} \right]^\phi f_{v,\theta}(v, \theta) dv d\theta
\end{aligned}$$

The firm then finds the pair of wages w^h and w^l which maximize profit.

1.6.2 Estimation

To estimate the model, I proceed in two steps. First, I estimate the parameters of the outside option distribution using a nonlinear least squares regression on reservation

wage data from a representative labor force survey, the Jordan Labor Market Panel Survey. Then, I express the main identifying assumptions of the model as moment conditions and solve for the unknown parameters using Generalized Method of Moments (GMM). I weight all moment conditions with inverse probability weights to address any selection into the productivity sample. Finally, because the outside option parameters enter into the moment conditions, I conduct inference with a non-parametric bootstrap of the entire estimation procedure.

Estimating the Outside Option Distribution

I use data on stated reservation wages among a sample of unemployed women from the representative Jordan Labor Market Panel Survey (JLMPS) to fit these parameters. To ensure comparability, the sample is restricted to Jordanian women who reported being open to data entry, telemarketing or teaching jobs (92% of respondents).¹⁹

Taking logs on both sides, equation (1.2) becomes:

$$\ln(F(w)) = \phi \ln(w - \underline{b}) - \phi \ln(\bar{b} - \underline{b}) \quad (1.5)$$

Using the reservation wage data, I construct the empirical CDF $\hat{F}(w)$ which reports what share of JLMPS respondents would work at each wage w . The empirical estimating equation associated with equation (1.5), estimated by nonlinear least squares, is given by equation (1.6):

$$\min_{\phi, \underline{b}, \bar{b}} \sum_w \left\{ \ln \hat{F}(w) - \phi [\ln(w - \underline{b}) - \ln(\bar{b} - \underline{b})] \right\}^2 \quad (1.6)$$

The results of this estimation are reported in Table 1.9. The estimated minimum outside option (transformed to levels) is 0.8 JD per day, or approximately 2.66 USD PPP per day. This wage is very low, but this should be expected of the lower bound. The lowest adjusted full time daily wage reported by women in the study sample at any job is 1JD, which adds support that this lower bound is accurate. The estimated maximum outside option is 24.3 JD per day. This is equivalent to the 75th percentile of adjusted full time wages or the 96th percentile of actual monthly wages in the survey sample (many work less than full time). The implied extensive margin labor supply elasticity with respect to wages at the benchmark wage 15 is approximately 2, which is reasonable for this setting. It asymptotes towards 1.75 as the wage approaches the maximum outside option.

Identification

To identify the profit-maximizing wage(s) under different counterfactuals and the employment, profit, average worker productivity, and worker welfare associated with

¹⁹The question was only asked of unemployed women.

these wages, I make two key identifying assumptions. The first assumption is that piece rate earnings are proportional to true worker product θ_i . The second assumption is that the firm is currently paying the profit-maximizing wage conditional on not offering transportation. I discuss each assumption in turn.

Assumption 1. *Piece rate earnings are proportional to true worker productivity.*

While the model pays workers fixed daily wages, in practice the firm pays fixed daily wages on some projects and piece rates (pay per survey) on others.²⁰ I collected earnings from current workers separately by whether it was paid a fixed wage or piece rate, and this data proves useful for identification. Let e_i be earnings from piece rates, and assume that piece rate earnings are proportional to true worker product, where λ is the markdown: $e_i = \lambda\theta_i$. Since workers' only job is to produce surveys, this is a plausible restriction.²¹ A consequence of this assumption is that $\mu_e \frac{1}{\lambda} = \mu_\theta$ and $\sigma_e \frac{1}{\lambda} = \sigma_\theta$, where μ_e and σ_e is the mean and standard deviation of the distribution of piece rate earnings. Similar notation applies for θ . Then, since z_i is the z-score of θ_i , θ_i can be recovered from z_i using μ_θ and σ_θ as follows:

$$\begin{aligned}\theta_i &= (z_i \cdot \sigma_\theta) + \mu_\theta \\ &= (z_i \cdot \frac{\sigma_e}{\lambda}) + \frac{\mu_e}{\lambda}\end{aligned}$$

Assumption 2. *The firm is currently paying the profit-maximizing wage conditional on not offering transportation and $w_{a=0}^* = \mu_e$.*

With assumptions 1 and 2, λ is identified using the wage first order condition. To see this, observe that the sample equivalent of the wage FOC (equation (1.4)) with $a = 0$ is given by:

$$\frac{1}{N} \sum_{i=1}^N \left[-(w - \underline{b})^\phi + (\theta_i - \exp w) \cdot \frac{\phi}{\exp w} \cdot (w - \underline{b})^{\phi-1} \right] = 0 \quad (1.7)$$

With the assumptions above, I plug in μ_e for $\exp w$ and $(z_i \cdot \frac{\sigma_e}{\lambda} + \frac{\mu_e}{\lambda})$ for θ_i . Combined with the outside option parameters estimated as described in the previous section, all parameters in equation (1.7) are observed except λ :

²⁰The decision to model wages as fixed was driven by 1) the fact that the firm does currently pay some projects daily wages and 2) the necessity of eliciting WTP with respect to a fixed wage. Eliciting WTP for a job amenity with respect to a piece rate would have been conceptually confusing for respondents and would not have generated policy-relevant estimates. Additionally, because θ_i encapsulates quantity and quality of data, the key findings of the paper go through even with piece rates and unobserved ex ante data quality.

²¹Implicit in this assumption is also the assumption that quantity and quality of data covary perfectly, but this assumption is already incorporated in the model by assuming that workers have uni-dimensional, fixed productivity.

$$\frac{1}{N} \sum_{i=1}^N \left[-(\ln(\mu_e) - \underline{b})^\phi + \left(\left(z_i \cdot \frac{\sigma_e}{\lambda} \right) + \frac{\mu_e}{\lambda} \right) - \mu_e \cdot \frac{\phi}{\mu_e} \cdot (\ln(\mu_e) - \underline{b})^{\phi-1} \right] = 0$$

This expression is included as a moment condition in a GMM framework to estimate λ , which reveals the rescaling of z_i into θ_i consistent with the model.

To estimate how wages would be impacted by a mandate that the firm must provide transportation to workers, the profit-maximizing wage with transportation $w_{a=1}^*$ is estimated using the FOC with $a = 1$ as an additional moment condition:

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N \left[- (w_{a=1}^* + v_i - \underline{b})^\phi \right. \\ \left. + \left(\left(z_i \cdot \frac{\sigma_e}{\lambda} \right) + \frac{\mu_e}{\lambda} - \exp w_{a=1}^* - \psi_i a \right) \cdot \frac{\phi}{\exp w_{a=1}^*} \cdot (w_{a=1}^* - v_i a - \underline{b})^{\phi-1} \right] = 0 \end{aligned}$$

The employment level, average worker productivity, firm profit, and total worker utility associated with $w_{a=0}^*$ and $w_{a=1}^*$ are estimated by including moment conditions equal to the sample equivalent of the corresponding model-based expression. Appendix A.3 reports each of these moment conditions explicitly.

Selection Correction

To adjust for the fact that θ_i is only observed for those who chose to work, all moment conditions are weighted to correct for any selection. These weights are inverse probability weights from a lasso probit regression predicting who chose to work for the firm conditional on being contacted.²² Covariates are rich, including demographics, household income and finances, location, education, gender and work attitudes, future work plans, commute information, etc.

Offering Workers Contract Choice

I estimate the profit-maximizing wages associated with the amenity and no-amenity contracts by maximizing the sample analogue of the firm's expected profit:

$$\pi(w_h, w_l) = \frac{1}{N} \sum_{i=1}^N (\theta_i - \exp w^l - \psi) \left[\frac{w^l + v - \underline{b}}{\bar{b} - \underline{b}} \right]^\phi \cdot \mathbf{1}\{w^h < w^l + v_i\} +$$

²²Lasso probit is used since probit drops observations with predicted probability of working = 1. Perfect separation occurs because of the number of detailed covariates. As a robustness check, weights are also constructed using a linear probability model. The results are very similar.

$$(\theta_i - \exp w^h) \left[\frac{w^h - \underline{b}}{\bar{b} - \underline{b}} \right]^\phi \cdot \mathbb{1}\{w^l + v_i < w^h\} \quad (1.8)$$

This expression is not smooth so I optimize using the Nelder-Mead method. Results using grid search are similar.

1.7 Counterfactual Results

I first report the distribution of θ_i recovered by the estimation procedure described above. Figure 1.5 displays the probability density function of θ_i , separately by quartiles of WTP. First, the distribution is reasonable: few people have a marginal product more than twice the wage, and very few people have a negative product.²³ The negative correlation between WTP for transportation and productivity is also shown in the figure. The yellow density function corresponds to the workers in the top quartile of WTP for transportation, and their productivity is shifted further to the left, while the quartile of workers with the lowest WTP shown in blue are shifted to the right.

1.7.1 Counterfactual: Amenity Mandate and Relaxed Minimum Wage

The first counterfactual considers a policy requiring the firm to provide transportation in conjunction with a relaxation of the minimum wage. This counterfactual allows me to evaluate whether a firm’s profits would increase or decrease if it were to introduce transportation benefits and to determine the new profit-maximizing wage conditional on offering transportation. It is also a policy relevant counterfactual since Jordanian law includes other similar mandates. For example, Jordanian Labor Law Article 72 requires all firms whose employees have at least 15 children under 5 cumulatively to offer a daycare, though it is poorly enforced (World Bank, 2024).

Table 1.10 reports statistics on wages, employment, average worker productivity, firm profit, and worker utility under the status quo and the amenity mandate counterfactuals. The column labelled “Status Quo” reports the wage, employment, average worker productivity, firm profit, and worker utility according to the model when $a = 0$ and transportation is not provided. The column labeled “Amenity Mandate” reports these outcomes when transportation is provided to all workers ($a = 1$). When the firm offers transportation, it is profit-maximizing to lower the wage by 4.13 JD. It should be noted that this new wage is below the minimum wage in Jordan. The third row reports the total labor cost, which includes wages and the person-specific cost of providing transportation. Total labor costs increase under the mandate by 8.7 percent, meaning that the firm maximizes profits by not lowering the wage by the full cost of

²³That a small fraction of enumerators have negative product is not unreasonable. Cases of severe fraud or unchecked data collection errors can result in a scenario where an enumerator generated no useable data and created costly problems that would not have existed if they weren’t hired.

the amenity. Employment, reported as a share of the relevant population who accept the job offer, increases by 27 percent. This large increase in employment is reflective of the fact that many women have much higher WTP for transportation than the cost of providing it. As a result of the large increase in employment and the increase in utility on the job for women who work in both counterfactuals and have high WTP for transportation, total utility to workers increases by 36 percent.

Due to adverse selection, average productivity of workers employed at the firm under the mandate falls by 2.5 percent. The net impact of a modest increase in labor costs, a large increase in employment, and small reduction in average productivity is a small and statistically insignificant reduction in firm profit. To determine whether adverse selection is driving the reduction in profit, the row titled “Total profit without adverse selection” reports the profit the firm would make if it offered the mandate wage and employment responded the same as it does in the mandate counterfactual, but average productivity of workers did not change. This back of the envelope exercise shows that profit without adverse selection would be higher than the status quo. This finding implies that adverse selection is driving the reduction in profit due to offering transportation. Therefore, the fact that it is not profitable to provide transportation is indeed a market failure, caused by asymmetric information and adverse selection.

1.7.2 Counterfactual: Minimum Transportation Subsidy Required to Make Provision Profitable

The relatively modest reduction in firm profit from mandating amenities suggests that a small subsidy could encourage firms to offer amenities and increase employment. What is the minimum subsidy for transportation that would induce the firm to offer the amenity, assuming the firm can lower wages? To estimate this counterfactual, I introduce a subsidy parameter that reduces ψ_i (the cost of providing transport) by x percent, and estimate the subsidy amount that equalizes profit with and without the amenity. The minimum subsidy is 6.2 percent (s.e.: 5.5), and the upper bound of the 95 percent confidence interval is 16.9. The point estimate is relatively small, consistent with the small differences in profit between the amenity mandate and the no amenity status quo, though the confidence intervals are wide. This finding implies that a 6 percent subsidy on firm-provided transportation could increase employment by 30 percent and increase worker utility by 37 percent.

1.7.3 Counterfactual: Firm Offers Workers Choice of Transportation or Higher Wage

Under the no amenity status quo, the firm profits by offering high wages and avoiding adverse selection. With the amenity, the firm profits by attracting more workers to the firm who are willing to accept low wages if the amenity is provided. What if the firm did both, by offering workers a menu of contracts?

Column 3 of Table 1.10 reports the results of this counterfactual. In the two contract counterfactual, the amenity wage is lower than the wage in the amenity mandate counterfactual and the no-amenity wage is higher than the wage in the status quo counterfactual. This is reflective of the sorting taking place: the average productivity of workers choosing the no amenity contract is higher than the status quo counterfactual and the average productivity of workers choosing the amenity contract is lower than in the amenity mandate counterfactual. In other words, due to the sorting on WTP, the firm can pay high wages to the high productivity workers and lower wages with transportation to the low productivity workers.

Additionally, by allowing workers to chose their preferred compensation type, total employment is higher under the two contract counterfactual than either of the alternatives. This large increase in total employment (35 percent) explains the large increase in profit compared to the status quo. Employment increased dramatically while average productivity remains similar and wage + amenity cost per person only increases by 2.8 percent. Increased employment and increased utility at work also result in a large (44 percent) increase in worker welfare, which is the sum of utility above the outside option for all workers.

In summary, offering workers their choice of contract can increase both profit and worker welfare. This finding has important policy implications. In settings where productivity is unobservable at hiring and adverse selection on labor supply-increasing amenities operates, allowing workers the choice of purchasing the amenity from the firm could have large welfare enhancing effects if the firm is able to provide the amenity at cost lower than enough worker’s willingness to pay.

Why has the firm not already begun offering workers the option to purchase firm-provided transportation if it would increase profitability? One possibility is rational inattention: the firm is actively pursuing a several expansion plans, and these other plans may be more profitable than introducing transportation. To the extent that management has limited attention, focusing on geographic expansion, for example, may be rational. It may also be the case there are additional contracting or management costs to offering transportation for a fee to workers which exist outside the model. Finally, the firm may underestimate the magnitude of worker’s WTP for transportation, or perhaps has not considered the possibility of offering firm-provided transportation. This is not completely implausible since Jordan is a setting where amenities like healthcare, retirement, and maternity leave are run publicly and outside the purview of firms.

1.8 Policy Implications

This research has several important policy implications. The first insight is that mandating that firms providing transportation to all workers is unlikely to be an effective means of increasing women’s labor market access in an environment with less than perfect enforcement since it would likely result in lower firm profits, giving firms strong incentive to evade the policy. Similar evasion exists surrounding Jordan’s man-

date that firms with female employees who have at least 15 children under 5 in total provide a workplace nursery.

Given the finding that the menu of contracts with and without transportation could increase worker welfare and profit more than the transportation mandate, a policy which encourages firms to explore offering workers the option to purchase transportation may be more sustainable and effective. Short-term subsidies on firm-provided transportation to speed learning or information campaigns directed at employers could be effective.

Another conclusion arising from the two-contract counterfactual is that a binding minimum wage could prevent employers from offering high-amenity jobs to lower skilled women. Specifically, the cash compensation component in the contract including transportation is below the minimum wage in Jordan. While allowing workers to purchase transportation from the firm could likely avoid this issue, the fact remains that insufficiently flexible enforcement of the minimum wage could result in lower provision of high-amenity jobs and ultimately lower employment among women whose marginal product is near the minimum wage. Allowing employers to meet the minimum wage with a combination of cash wages and tangible benefits ensures that employers are not prevented by policy from offering high-amenity contracts that increase labor market access.

A key finding of this paper - that women in Jordan have high WTP for non-wage benefits - aligns indirectly with the findings of several papers concerning labor market interventions in the Middle East designed to match workers to jobs. A common theme in each of these papers is that matching interventions often fail because the jobs available do not meet workers expectations with respect to non-wage job quality. This issue is particularly salient for women, but appears to impact men's employment as well. For example, approximately 40 percent of Jordanian women in the capital city quit jobs facilitated by wage subsidies because they didn't like the job despite wanting work at baseline, and a job matching program for Jordanian men and women failed to generate almost any matches because job seekers declined or quickly quit jobs 83 percent of the time (Groh, Krishnan, et al., 2016; Groh, McKenzie, et al., 2015). Finally, a study which cross randomized a childcare subsidy and job matching for women in Egypt also had little success (Caria et al., 2025). As explanation, they cite a large mismatch between the expectations women have for non-wage job quality and the jobs that are available. While it is possible that these studies simply reflect high levels of reservation job utility and rationally low levels of employment, my findings suggest that an emphasis on increasing non-wage job quality in the private sector could have larger than expected impacts on employment.

Finally, the high estimates of women's WTP for transportation suggests that investments in public transportation could have large benefits to women's labor market access. Average WTP for transportation was higher than WTP for childcare even among women with children under 5, showing that even the women most likely to value childcare, on average, value transportation more. To the extent that all women (and men) require transportation to work for their whole career, while childcare is only

needed for a specific phase of life, investing in a more affordable and more connected public transportation network could yield large dividends. Saudi Arabia is an excellent example of this in action. Between lifting bans on women driving and *Wusool*, an Uber subsidy program for women commuting to private sector jobs, Saudi Arabia has seen its female labor force participation rate increase by nearly 10 percentage points (Arab News, 2022). Experimental evidence shows that Saudi women who gained access to driver’s license training in the immediate wake of the lifting of the driving ban became 41 percent more likely to be employed (Abou Daher et al., 2025). Taken together, these findings emphasize the enormous impact that improved access to transportation can have on women’s labor market participation.

1.9 Conclusion

Around the world, women face a host of different barriers to employment. Many of these issues are not within the purview of employers to address, but some are. Firms interested in hiring women could offer transport as job amenities to address some of the missing markets which reduce women’s employment. Despite policymaker enthusiasm surrounding firm-provided job amenities, firms remain empirically unlikely to offer these benefits. Using individual-level productivity data and willingness to pay (WTP) data for several relevant job amenities among a sample of female survey enumerators in Jordan, this paper explores one understudied mechanism which makes offering female-friendly job amenities privately undesirable to firms: adverse selection. I document a negative correlation between productivity and WTP for three out of four amenities studied which increase access to work or the predictability of pay. For transportation in particular, I rule out all possible mechanisms which might generate a negative correlation in the cross section but not result in adverse selection, such as selective attrition, selection into employment, and heterogeneous treatment effects of the amenity itself.

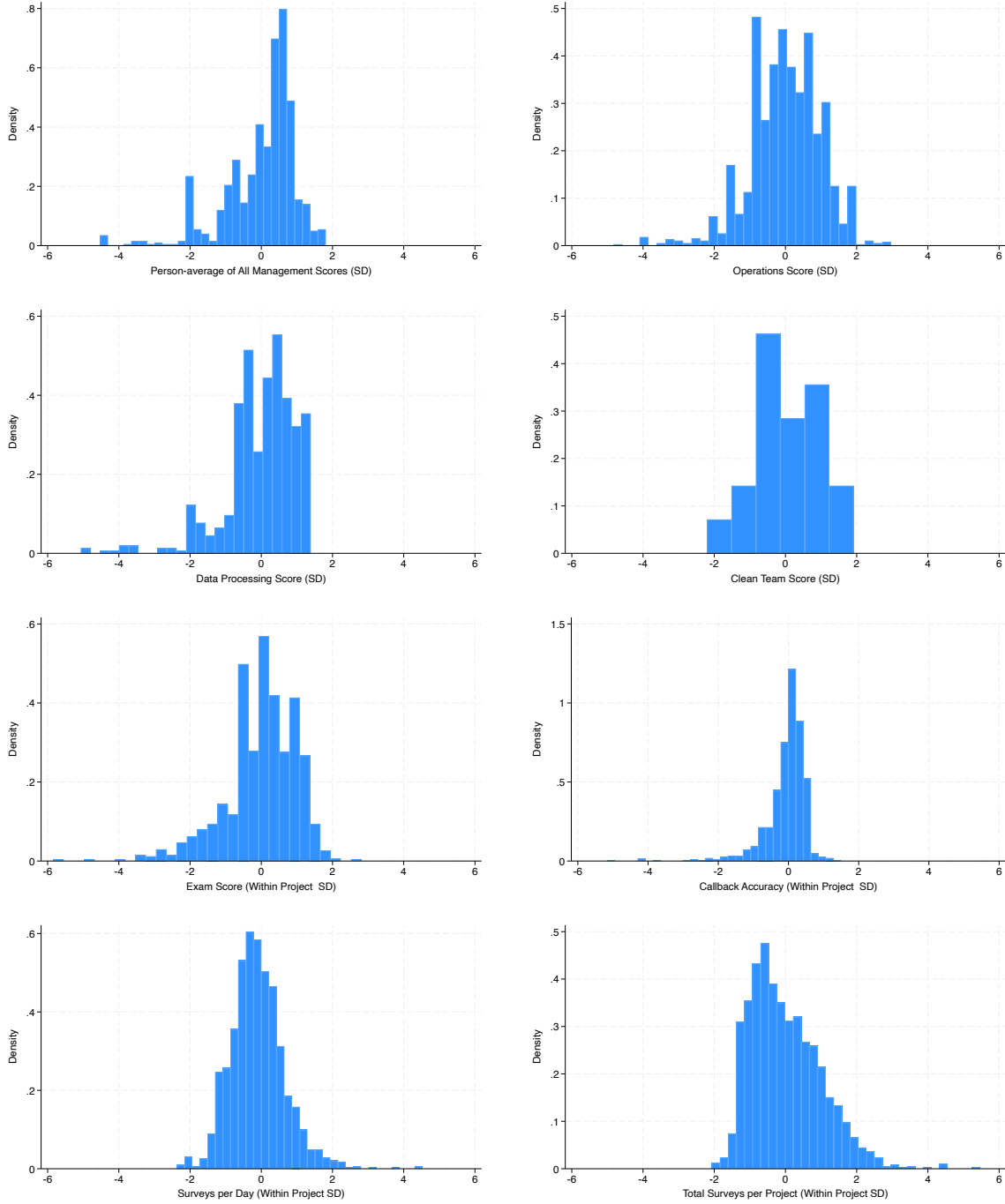
To evaluate how this adverse selection would impact the profitability of offering transportation to workers, I develop and estimate a monopsony model where amenities can increase labor supply and induce selection on workers. The results show that offering transportation to all workers and lowering the wage accordingly lowers firm profit (though a small, and statistically insignificant amount) due entirely to the adverse selection margin. Therefore, it is indeed due to a market failure that the firm does not offer transportation to its workers.

I also show that according to the model, the market failure can be corrected by allowing workers to opt into purchasing transportation from the firm. By allowing workers to sort into their preferred contract, the firm can continue to attract higher-productivity workers who are more likely to prefer wages, and set the price of firm-provided transportation high enough to keep it profitable. The net effect is the result of small changes on average compensation per worker and average worker productivity outweighed by large increases in labor supply.

These findings constitute an important contribution to our understanding of how labor market failures can perpetuate the exclusion of marginalized groups from the labor market. In any context in which productivity is not fully observable ex ante and workers who value accommodative amenities the most are unobservably less productive, the absence of full information can result in an equilibrium in which accommodative amenities are not provided, even to workers whose marginal product is above their reservation compensation level. This issue plausibly affects many labor markets beyond this context, since firms often lack full information prior to hiring and it is often the case that the type of worker who has the largest demand for a flexible or accommodative amenity might also disproportionately faces other challenges in their life which negatively impact their productivity. Future research can establish whether these patterns generalize to other contexts and clarify the precise mechanisms driving the negative correlation between WTP and productivity shown here. Determining whether the negative correlation reflects unobserved disadvantages such as physical health, the disutility of effort, or other underlying preferences remains an important task for future research.

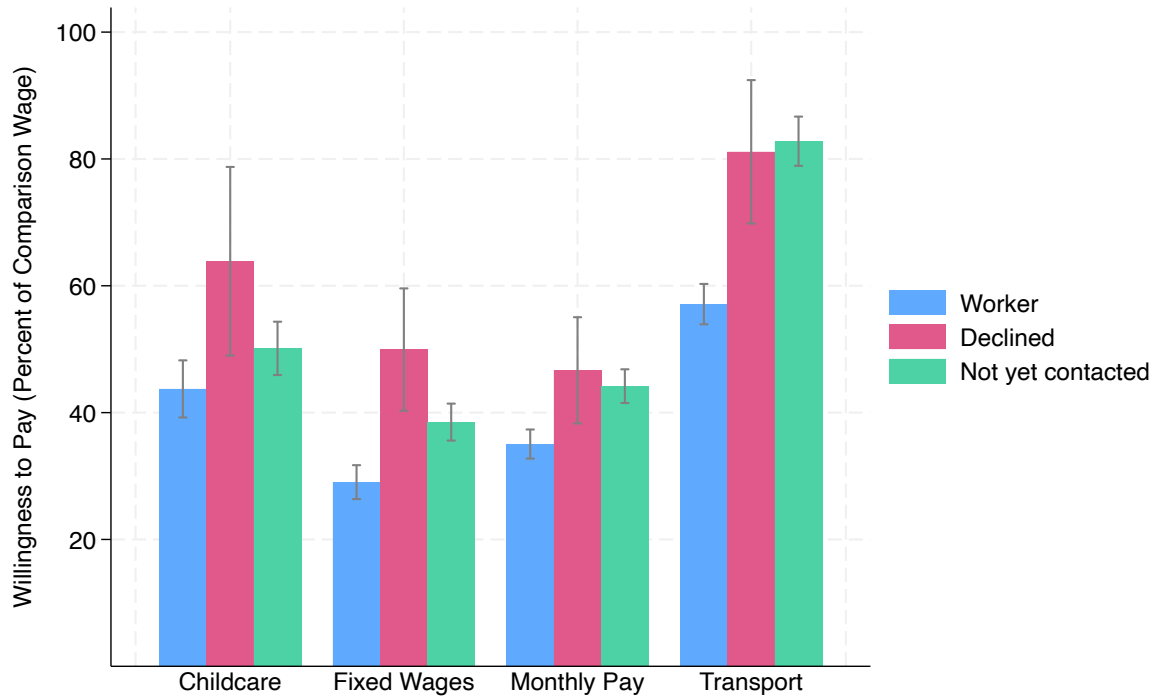
1.10 Figures

Figure 1.1: Distribution of Productivity Scores



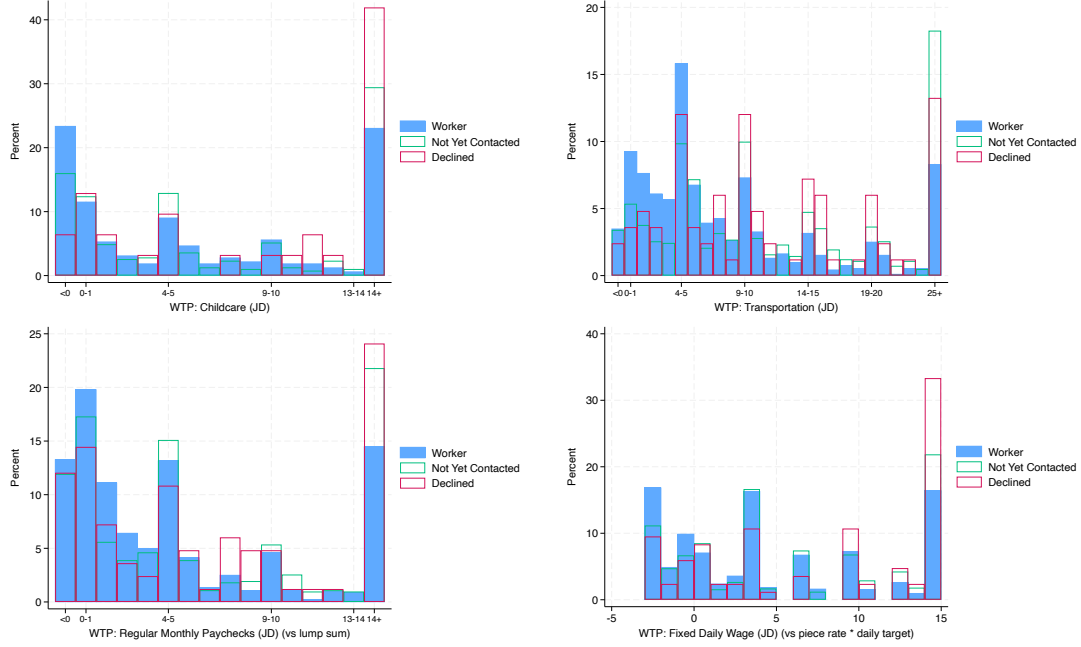
Note: This figure displays histograms of each of the different measures of productivity that enter into the primary specification. Each outcome has been standardized as described in section 1.3. The unit of observation is the enumerator-project for outcomes labeled as within-project SD or enumerator-year for projects labelled as SD.

Figure 1.2: WTP for Job Amenities



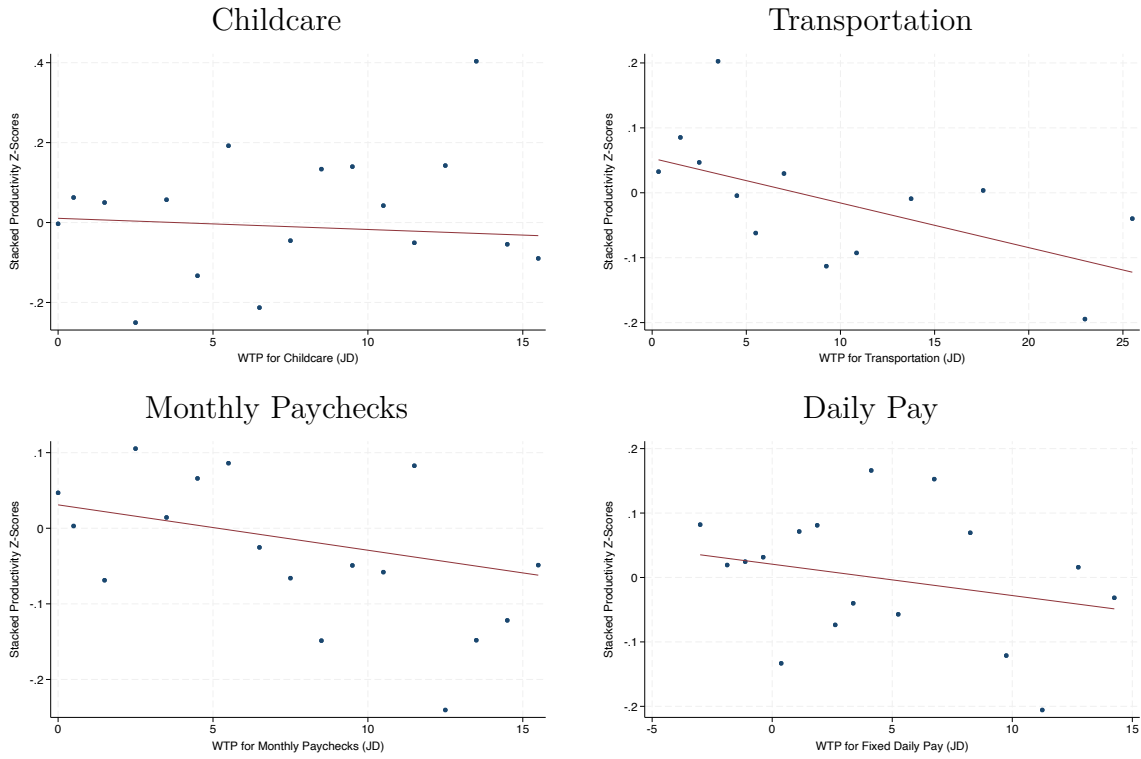
Note: This figure reports the WTP that Jordanian women in the sample assign to the four job amenities, reportedly separately by whether the individual was a worker, had been offered a job but declined, or had applied but the firm had not yet contacted them. The childcare amenity is a voucher for childcare at a daycare center of the worker's choosing for the duration of her contract. The transportation amenity involves a trusted driver from the firm providing transportation to and from the worker's home. The monthly pay amenity is a monthly paycheck instead of being paid in one lump sum at the end of a 2-month project. The fixed wage amenity involves a fixed daily wage instead of being paid per survey completed (i.e., a piece rate). The willingness to pay (WTP) elicitation was compared to a 15 Jordanian Dinar daily wage, and the y-axis reports WTP as a percent of the comparison wage (15 JD/day). So a woman who was indifferent between 15 JD/day with childcare and 25 JD/day without childcare has a WTP for childcare of 10 dinar, which is 66 percent of the comparison wage.

Figure 1.3: Distribution of WTP



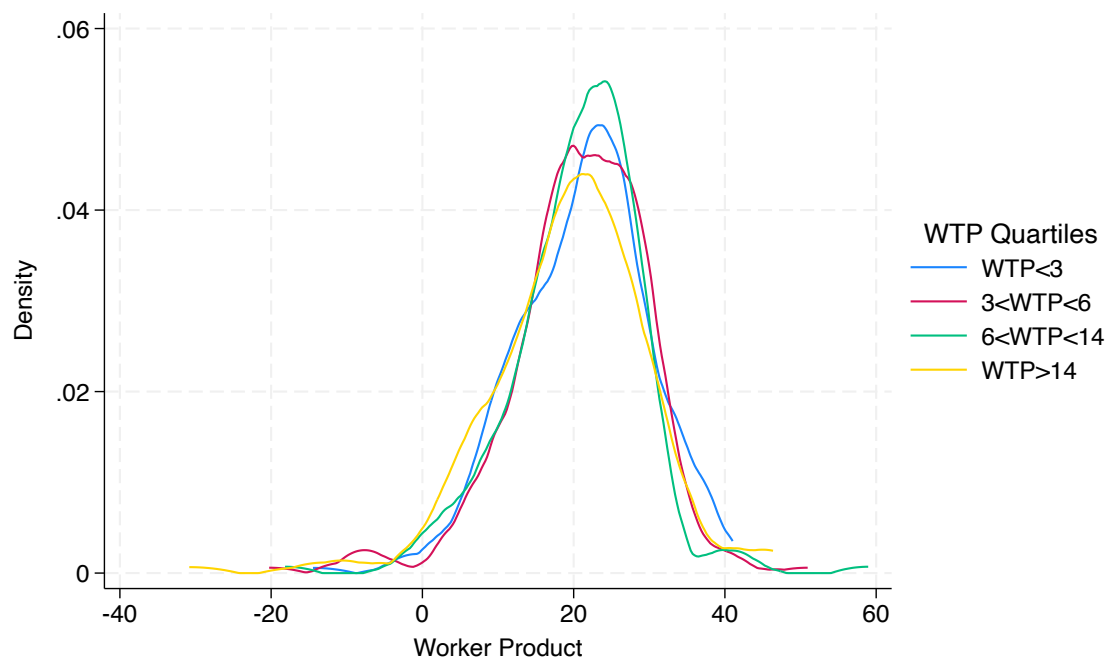
Note: This figure includes histograms of the WTP that Jordanian women in the sample assign to the four job amenities, reportedly separately by whether the individual was a worker, had been offered a job but declined, or had applied but the firm had not yet contacted them. Each histogram is top-coded by the top of the price list, meaning that individuals in the far right bin reported that they always prefer job A (the job with the amenity). These individuals were then asked an open-ended question about the lowest job B wage that would induce them to prefer the higher wage to the amenity. Appendix Figure A.1 reports the uncensored histograms that use the open-ended responses. The transportation price list included higher job B wages since piloting revealed higher WTP for transportation than the other amenities.

Figure 1.4: Distribution of WTP and Productivity



Note: This figure displays binscatters of WTP against productivity measures z_{ij} for all survey respondents with nonmissing WTP and at least one measure of productivity. The underlying data includes WTP with stacked measures of productivity, so there are multiple observations per respondent.

Figure 1.5: Probability Density of Worker Productivity by WTP Quartiles



Note: This figure displays the empirical probability density of θ by quartiles of WTP. θ_i is given by mean z_{ij} rescaled by the parameters of the θ distribution as estimated in the structural model.

1.11 Tables

Table 1.1: Evidence on Lack of Additional Firm Screening

	Point Estimate	N
Married	-0.028 (0.024)	1022
Has children	-0.035 (0.026)	1022
Number of children	-0.012 (0.018)	440
Conscientious (SD)	-0.008 (0.016)	1022
Feels safe on public transport (SD)	0.001 (0.011)	1022
Transport is a barrier for me (SD)	0.014 (0.012)	632

Note: This table provides supporting evidence that the firm does not conduct additional screening of applicants beyond the information requested in the application, which includes age, location, education, and prior experience in the industry. Each row reports the coefficient from a regression of an indicator for being contacted on the outcome listed along with fixed effects for governorate, industry experience interacted with birth year, education level, and the specific job they applied for. Because being contacted increases the probability of being employed, outcomes that are unlikely to be impacted endogenously by employment were selected, such as marital status. Conscientious (SD) reports the respondents score on the Conscientiousness scale from the Big 5 personality test. Feels safe on public transportation is a standardized index that reports the degree to which the respondent agrees that she feels safe on public transportation. Transport is a barrier for me is an index that reports the degree to which the respondent agrees with the statement: “Difficulties with transportation prevent me from working as much as I’d like”. The sample sizes for these regressions are smaller than the full sample because the applicant database shared by the firm contains some missing data on prior experience.

Table 1.2: Summary Statistics - Study Sample vs Population

	<i>Survey Sample</i>		<i>Pop, female 18-65 (2016)</i>	
	Mean	N	Mean (all)	Mean (in labor force)
Age	32	1,826	35.69	32.51
Married (=1)	0.49	1,828	0.65	0.54
Divorced/Widow (=1)	0.07	1,828	0.06	0.05
Any children (=1)	0.45	1,828	0.58	0.49
Number of children	2.56	827	3.88	3.13
Bachelor's degree (=1)	0.79	1,828	0.19	0.54
Associates degree (=1)	0.95	1,828	0.30	0.72
Location: Amman (=1)	0.19	1,828	0.38	0.32
Location: urban (=1)	0.63	1,828	0.87	0.85
Worked in last 3 months (=1)	0.48	1,828	0.12	0.61
Worked in last 7 days (=1)	0.29	1,828	0.12	0.60

Note: This table reports summary statistics from the survey sample and the Jordan Labor Market Panel Survey from 2016, the most recent publicly available representative labor survey.

Any children is an indicator for the respondent having any children. Number of children reports the number of children conditional on having any children. Bachelors and associates are indicators if the respondent has attained at least that level of education. Amman is an indicator if the respondent lives in Amman, the capital city. Urban is an indicator for living in a self-reported urban area. This includes urban areas in cities smaller than Amman. Worked in the last 3 months/7 days are indicators for having done any amount of wage work or self employment.

Table 1.3: Summary Statistics: Sample Subpopulations

	<i>Worked (W)</i>		<i>Declined (D)</i>		<i>Not contacted (NC)</i>		<i>W-D</i>	<i>D-NC</i>
	Mean	N	Mean	N	Mean	N	Diff	Diff
Age	33	919	32	84	31	823	1.77 * *	0.38
Married (=1)	0.45	921	0.48	84	0.53	823	-0.03	-0.06
Any children (=1)	0.42	921	0.42	84	0.49	823	0.00	-0.08
Number of children	2.66	387	2.46	35	2.47	405	0.20	-0.01
Bachelor's degree (=1)	0.72	921	0.87	84	0.86	823	-0.15***	0.01
Associates degree (=1)	0.91	921	1.00	84	0.99	823	-0.09***	0.01
Location: Amman (=1)	0.26	921	0.49	84	0.08	823	-0.23***	0.41***
Location: urban (=1)	0.72	921	0.77	84	0.52	823	-0.06	0.26***
Worked in last 3 months (=1)	0.62	921	0.38	84	0.35	823	0.24***	0.03
Worked in last 7 days (=1)	0.35	921	0.30	84	0.23	823	0.05	0.07
Main current job w/ firm (=1)	0.34	921	0.00	84	0.00	823	0.34***	-0.00
Monthly earnings (own, USD PPP)	368	849	291	76	228	740	76.20	63.15
Monthly hh income (USD PPP)	1,631	732	1,622	62	1,381	621	9.05	241.11
Weekly hours of domestic work	20.20	921	22.45	84	23.81	823	-2.25	-1.36
Months worked last 5 years	23.36	921	21.54	84	19.44	823	1.82	2.10
Wants a different job (=1)	0.86	534	0.92	25	0.86	227	-0.06	0.06
Career oriented (=1)	0.12	913	0.19	83	0.17	818	-0.08 * *	0.02
Financially insecure (=1)	0.34	919	0.25	84	0.33	822	0.09	-0.08

Note: This table reports summary statistics from three different groups within the survey sample: women who have worked for the firm, women who have been contacted by the firm but declined to work, and women who have not been contacted. A respondent is marked as having worked if the firm has productivity data for them or if they self-reported to have worked for the firm before. A respondent is determined to likely have declined if they have never worked before and are on the firm's "inactive" list, which is a list that individuals enter by stating that they are not interested, or by not replying to 3 separate attempts to contact them. Individuals fall into the not contacted category by having applied, never worked, and not being on the inactive list (meaning they never indicated that they were not interested). The column marked 1-2 reports the difference between workers and those who declined, and can be interpreted as the selection effects into employment conditional on a job offer. The column marked 2-3 reports the difference between individuals who declined and those not yet contacted. Consistent with information from the firm about how they contact potential workers, the main differences are geographic. Stars indicate statistical significance at the 1 percent (***), 5 percent (**), and 10 percent (*) levels.

Any children is an indicator for the respondent having any children. Number of children reports the number of children conditional on having any children. Bachelors and associates are indicators if the respondent has attained at least that level of education. Amman is an indicator if the respondent lives in Amman, the capital city. Urban is an indicator for living in a self-reported urban area. This includes urban areas in cities smaller than Amman. Worked in the last 3 months/7 days are indicators for having done any amount of wage work or self employment.

Main current job with firm is an indicator for the respondent's main job being with the partner firm. Monthly earnings (own) are the respondents earnings in the past month, scaled from JD to USD PPP by the USD PPP adjustment factor (divide JD by 0.33, see World Bank Data PA.NUS.PRVT.PP). Monthly household income includes all earned and unearned income from any household member in the past month. Hours of domestic work is the number of hours spend on domestic chores, childcare and eldercare in the past 7 days. Months worked last 5 years reports the number of months out of the last 5 years (60 months) that the respondent was employed. Want a dif job is an indicator equal to one if women currently working for a wage selected one of the following options: Two years from now, I'd prefer to 1) be employed in the same job, but with some things changed, or 2) employed in a different job or 3) be self-employed. Career oriented is an indicator equal to one if the respondent reported that building a career was her primary motivation for working besides earning money. (The other options were 'To gain independence', 'To spend my time productively', 'None of the above, I only work to earn money'.) Financially insecure is an indicator equal to one if the respondent's household does not have access to 100 JD / 300 USD PPP in savings and could not easily borrow this amount in an emergency.

Table 1.4: Correlates of Willingness to Pay

	Transport	Childcare	Regular Monthly Paycheck	Fixed Daily Pay
Urban (=1)	-0.58 (0.44)			
Commute for firm (hours)	0.07 (0.38)			
Commutes to firm job via bus (=1)	-0.39 (0.54)			
Cost of commute to firm job	0.43*** (0.11)			
Feels safe on public transport (SD)	-0.47*** (0.19)			
How often ride is late (SD)	-0.22 (0.26)			
Transport is main barrier (=1)	1.28** (0.66)			
Would use paid childcare (=1)		6.00*** (0.49)		
Mom available for childcare (=1)		-2.05*** (0.57)		
Men in family disparage women who work (SD)		0.71*** (0.22)		
Childcare is biggest barrier (=1)		1.66*** (0.52)		
Man should have final say (SD)		0.39* (0.22)		
Wants a different job 2 yrs from now (=1)		1.84** (0.92)		
Motivation for working = career		-0.99 (0.69)		
Minutes to nearest daycare		0.04 (0.03)		
Cost of current paid childcare (JD)		0.01 (0.01)		
<100 savings and can't borrow			0.70*** (0.28)	
Late paycheck: reduced expenditure or took loan			0.97** (0.42)	
Has formal loan (0/1)			0.76* (0.46)	
Only motivation for working is income				0.58* (0.30)
Worked for firm in last 3 months				-1.83*** (0.42)
Worked in last 3 months	-0.72* (0.39)	-1.16** (0.47)	-0.72*** (0.27)	-0.94*** (0.30)
HH income per capita (minus own) (100s USD PPP)	-0.15* (0.08)	0.07 (0.13)	0.03 (0.06)	-0.09 (0.06)
Big 5 neuroticism scale (SD)	0.43 (0.27)	0.82** (0.34)	0.30* (0.19)	0.15 (0.20)

Note: This table reports covariates which predict WTP for different amenities. Each coefficient is from a separate regression of WTP for an amenity on the covariate, with controls for ever working for the firm, being contacted by the firm, and governorate.

Table 1.5: Main Results

	Productivity Measures (z_{ij})			
	(1)	(2)	(3)	(4)
Transport	-0.00688*** (0.00230)			
Monthly Pay		-0.00600** (0.00274)		
Fixed Wage			-0.00487* (0.00269)	
Childcare				-0.00280 (0.00462)
WTP SD	7.34	5.24	6.08	5.85
Obs	5520	5530	5540	1769

Note: This table reports results from a regression of productivity (SDs) on WTP for the amenity in dinars: $z_{ij} = WTP_i + \epsilon_{ij}$. The data is stacked so that the left hand side variable includes all measures of productivity observed for an individual, and the right hand side is WTP. Standard errors clustered at the individual level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Column 1 reports that an individual with WTP 1 dinar higher is expected to be 0.00688 SDs less productive. Scaled by the standard deviation of WTP for transportation, this implies that an individual with a 1SD higher WTP for transportation will be on average 0.05 SDs less productive.

Table 1.6: Transportation and Monthly Pay with administrative controls

	Productivity Measures (z_{ij})					
	(1)	(2)	(3)	(4)	(5)	(6)
Transport	-0.00436*	-0.00487*	-0.00496**			
	(0.00228)	(0.00280)	(0.00250)			
Monthly Pay				-0.00442	-0.00861***	-0.00568*
				(0.00278)	(0.00327)	(0.00304)
Birth Year	Yes	Yes	Yes	Yes	Yes	Yes
Location	Yes	Yes	Yes	Yes	Yes	Yes
Education	Yes	Yes	Yes	Yes	Yes	Yes
Industry Experience		Yes			Yes	
Nth Project			Yes			Yes
Obs	5497	3863	4907	5507	3864	4914

Note: This table reports results from a regression of productivity (SDs) on WTP in dinars for transportation or monthly paychecks: $z_{ij} = WTP_i + \epsilon_{ij}$ with the addition of controls from the firm's administrative records. Standard errors clustered at the individual level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Location includes fixed effects for the respondent's subdistrict of residence, which is the 3rd administrative division in Jordan. Education is a fixed effect of educational attainment in bins. Industry experience is a fixed effect for the enumerator's self reported amount of experience in the enumeration industry as measured in the application. The sample size falls since the firm could not locate records of this outcome for all individuals. N'th project is a fixed effect for whether the productivity outcome corresponded to the enumerator's first, second, third, etc project with the firm. This information is only available for outcomes from survey metadata, including exam score, number of surveys, surveys per day, and callback accuracy. Results are not reported with industry experience and n'th project fixed effects because doing so limits the sample size considerably.

Table 1.7: Fixed Daily Wages and Childcare with administrative controls

	Productivity Measures (z_{ij})					
	(1)	(2)	(3)	(4)	(5)	(6)
Fixed Wage	-0.00355 (0.00254)	-0.00250 (0.00319)	-0.00341 (0.00262)			
Childcare				0.00338 (0.00506)	0.0164** (0.00671)	0.00913* (0.00502)
Birth Year	Yes	Yes	Yes	Yes	Yes	Yes
Location	Yes	Yes	Yes	Yes	Yes	Yes
Education	Yes	Yes	Yes	Yes	Yes	Yes
Industry Experience		Yes			Yes	
Nth Project			Yes			Yes
Obs	5517	3874	4924	1760	1165	1521

Note: Please refer to the table notes for Table 1.6.

Table 1.8: Transport and Monthly Pay with Additional Survey-Based Controls

	Productivity Measures (z_{ij})			
	(1)	(2)	(3)	(4)
Transport	-0.00596* (0.00319)	-0.00780** (0.00358)		
Monthly Pay			-0.0102** (0.00430)	-0.00983* (0.00535)
Fixed Effects	Location Education Birth Year Experience Surveyor	Location Education Birth Year Experience Surveyor	Location Education Birth Year Experience Surveyor	Location Education Birth Year Experience Surveyor
Controls	Demo, income, attitude	Demo, income, attitude + commute info	Demo, income, attitude	Demo, income, attitude + financial precarity
Obs	3091	2655	3112	2680

Note: This table reports results from a regression of productivity (SDs) on WTP in dinars for transportation or monthly paychecks: $z_{ij} = WTP_i + \epsilon_{ij}$ with the addition of controls from the firm's administrative records and information from the survey which the firm does not have access to. Standard errors clustered at the individual level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Location refers to subdistrict, the third administrative division in Jordan. Surveyor refers to the enumerator who conducted the WTP elicitation. Demographic controls include quadratic age, urban, marital status, number and age of children, household composition, household size, and months of employment in the past 5 years. Income refers to a control for the respondents own income and the rest of the household's income. Attitudes refer to gender attitudes, how much of her income she saves for herself, motivation for employment, and each of the respondent's scores on the Big 5 personality traits. Commute info refers to commute time in hours, mode of transportation interacted with cost of commute, and perceived safety on the commute. Financial precarity is an indicator equal to one if the respondent has less than 100 JD (333 USD PPP) in savings and has no friends or family from which to borrow money in an emergency.

Table 1.9: Labor Supply Parameters

	Parameter	Estimate
Elasticity parameter	ϕ	5.94 (1.79)
Outside option lower bound (JD/day)	$\exp \underline{b}$	0.84 (0.51)
Outside option upper bound (JD/day)	$\exp \bar{b}$	24.29 (2.17)
Implied labor supply elasticity	$\frac{d \ln F(w)}{dw}$	2.06 (0.22)

Note: This table reports estimates from a nonlinear least squares regression of the log-transformed empirical CDF of reservation wages on the log-transformed mathematical expression of the CDF with the shifted power function function form, of $\ln F(w) = \ln(\frac{\ln(w) - \ln(\underline{b})}{\ln(\bar{b}) - \ln(\underline{b})})^\phi$. The underlying reservation wage information comes from a sample of unemployed women in the representative Jordanian Labor Market Panel Survey. The last row reports the labor supply elasticity implied by the parameters estimated above at $w=15$.

Table 1.10: Firm and Worker Outcomes Under Counterfactuals

	Counterfactuals			Differences		
	Status Quo	Amenity Mandate	Contract Choice	2-1	3-1	3-2
Wage (dinar), no amenity	13.99 (0.41)		14.31 (0.46)		0.32*** (0.12)	
Wage (dinar), has amenity		9.84 (0.42)	9.14 (0.48)			-0.71*** (0.24)
Wage + amenity cost (dinar)	13.99 (0.41)	15.21 (0.42)	14.41 (0.43)	1.22*** (0.16)	0.42*** (0.15)	-0.80*** (0.13)
Employment (share), no amenity	0.34 (0.04)		0.36 (0.04)		0.02*** (0.01)	
Employment (share), has amenity		0.43 (0.05)	0.37 (0.04)			-0.06*** (0.02)
Employment (share), total	0.34 (0.04)	0.43 (0.05)	0.47 (0.04)	0.09*** (0.02)	0.12*** (0.02)	0.03*** (0.01)
Productivity (dinar), no amenity	20.60 (0.86)		21.19 (0.98)		0.59** (0.25)	
Productivity (dinar), has amenity		20.08 (0.89)	19.44 (0.99)			-0.64* (0.38)
Productivity (dinar), all	20.60 (0.86)	20.08 (0.89)	20.31 (0.87)	-0.52** (0.26)	-0.29** (0.14)	0.23* (0.14)
Total Profit	1989.45 (370.61)	1849.96 (372.69)	2388.32 (423.54)	-139.49 (122.40)	398.87*** (95.67)	538.36*** (80.24)
Total Profit w/o adv. selec.		2047.24 (368.74)				
Total Utility	121.74 (14.60)	166.08 (21.51)	176.21 (20.95)	44.34*** (8.93)	54.48*** (8.98)	10.14*** (3.64)

Note: This table reports wages, employment, profit, worker productivity, and total worker utility from the status quo and two key counterfactuals. The amenity mandate requires that the firm provide transportation to all workers, and the contract choice counterfactual allows the firm to offer two contracts, one with a high wage and one with a lower wage that includes transportation, and allows workers to chose the contract they prefer.

The amenity mandates increases employment and increases worker utility, but lowers average worker productivity and lowers total profit. The reduced profit effect is the net effect of increased employment (good for profit), reduced productivity (bad for profit), and a reduction in wages by less than the amenity cost (bad for profit).

By allowing workers to self-sort, the contract choice counterfactual provides workers with more total utility than the amenity mandate and results in considerably more profit for the firm. This is because the firm can efficiently use cash to attract workers with low valuations of the amenity, and attract workers with high valuations with transportation while lowering the amenity wage enough to offset the cost of provision.

The first two rows list the amenity and no amenity wages. The next 3 rows report the share of the population who are employed with and without amenities and the total employment. The 3 rows below report productivity of workers employed with and without amenities. Total profit reports the total profit the firm earns under the counterfactual. Total utility reports the total utility in excess of their outside option enjoyed by workers. It encapsulates the effects of both increased employment and increased job quality.

Total profit without adverse selection uses the same wage + amenity cost and employment as the amenity mandate counterfactual but uses the status quo level of productivity. This back of the envelope calculation reports what profit would be with the same labor supply response to amenities, without adverse selection. The fact that profit in this counterfactual is higher than the status quo implies that amenities reduce profit because of adverse selection.

Chapter 2

Economic and Social Impacts of Entering the Labor Market During a Recession: Evidence from Egypt

2.1 Introduction

Entering the labor market during a recession has the potential to disrupt or delay the transition to adulthood in a variety of ways. Long unemployment spells or a series of dead-end jobs post-graduation can harm long-run labor market outcomes and may have implications for family formation and health outcomes. Indeed, there is a large literature studying the long-run effects of graduating a recession in high income countries on both labor market and social outcomes. These papers often find that unlucky cohorts have depressed earnings for 5-10 years after graduation, and work in lower-paying firms and in lower paying or poorly matched industries and occupations (Kahn, 2010; Oreopoulos, T. von Wachter, and Heisz, 2012; Altonji, Kahn, and Speer, 2016; Liu, Salvanes, and Sorensen, 2016; van den Berge, 2018; Rothstein, 2023; Huckfeldt, 2022; Arellano-Bover, 2024). Recession entrants in high income countries also fare worse in the marriage market, and have even been shown to die younger (Schwandt and T. M. von Wachter, 2020). However, despite the fact that searching for one's first job during a recession is challenging regardless of context, we have very limited evidence on recession scarring for labor market entrants in low and middle-income countries. The evidence that does exist is limited in scope and methodological depth due to the challenging data environment, which is a key challenge that I overcome in this paper.¹

¹Kuchibhotla, Orazem, and Ravi (2020) and Naidoo, Packard, and Auwalin (2015) examine the effect of unemployment spells at the beginning of an individual's career, but use propensity score matching instead of unemployment rate at entry to address endogeneity. Cruces, Ham, and Viollaz (2012) estimates the effect of entering the labor market in Brazil and Argentina during periods of higher unemployment or higher informality. They only have access to unemployment and informality statistics at the national level, so they use a 2-step process where they estimate cohort effects then

There are numerous contextual differences in low- and middle-income labor markets that make extrapolation of existing scarring results from high income settings difficult. Youth unemployment is more common in developing countries, and these youth unemployment spells are often much longer. The consequences of this fact for labor market scarring are unclear: duration effects suggest that scarring may be worse in these settings, but it may also be easier to recover from a recession when the existing equilibrium already features high rates of joblessness. Additionally, characteristics of developing labor markets—such as large informal sectors, sizable public sectors, and high rates of self-employment—further suggest that the effect of entering the labor market during a recession in a developing country remains an open empirical question.

Evidence from high-income countries on how recession entry affects family formation also offers limited insight into what these impacts might look like in developing contexts. In many low- and middle-income countries, marriage is a stronger social norm and is an economic event often accompanied by substantial payments at the time of marriage. As a result, extrapolating from high-income settings is ambiguous. Would the effects be larger or smaller in a setting where marriage is more ubiquitous? Does the existence of marriage payments help clear the market or add an additional hurdle to scarred cohorts? The importance of marriage in the transition to adulthood and the large cash flows involved make the effects of graduating during a recession on marriage outcomes highly economically salient in a developing setting.

A key barrier to studying recession scarring in low- and middle-income countries is the data requirements, which this paper overcomes. First, subnational unemployment rate records are needed to identify the effect of unemployment conditions at labor market entry while controlling for cohort effects. Second, a sufficiently long time series of labor market entry is required to ensure statistical power. Few low- and middle-income countries began measuring and publishing subnational unemployment rates early enough to meet these criteria. Egypt, however, has national statistical records dating back to 1968 that include subnational unemployment rates. A major contribution of this paper is the digitization of these records.

Using publicly available labor market panel data alongside the newly hand-digitized historical unemployment rates, I estimate the effect of economic conditions at entry on long-run economic and social outcomes. I find effects consistent with the economic scarring documented in high-income countries, but which manifest differently in this setting due to the structure of the job ladder. Men experience reductions in employment for 2 years after entry, but the long run effect on employment is actually positive. However, this increase is mostly driven by an increase in self-employment, and in particular, self employment without any employees. This type of employment is associated with lower earnings, more precarity, and is often understood as disguised unemploy-

regress the cohort effects on the conditions at entry. They find that one percentage point higher unemployment rate lowers the probability of employment by .005 in Argentina and .009 in Brazil over the first 10 years. The data limit them to only 2 potential experience bins (0-10 and 11-20) and lacks information on job characteristics, so the paper is unable to comment on potential mechanisms.

ment (Majid, 2012).² Male recession entrants are also less likely to have a formal wage work job, which is generally perceived to be the most desirable form of employment due to its stability, access to labor protections, and higher average wages.³ Consistent with these findings, results on earnings based on a smaller subsample indicate that recession-scarred cohorts are more likely to have below-median earnings.

Suggestive evidence examining heterogeneous effects by educational attainment shows that scarring effects are largest among college graduates. This group is the most likely to hold formal jobs and, according to the results, experiences the largest declines when entering the labor market during a recession. This is consistent with the broader social science literature evaluating the labor market conditions which led to the Egyptian Revolution in 2011, but notably this pattern is also true of prior business cycles in Egypt.

Recession entrance also delays the marriage market significantly, for both men and women. For each additional point in the unemployment rate at entry, the hazard rate into marriage is lower by 1.5 percent for men and 1.1 percent for women. In settings where young people live in their family home until marriage, and where marriage is costly due to brideprice/dowry, high youth unemployment effectively impedes youth people’s transition to adulthood. This stagnation has been dubbed “waithood”, as young people wait to find a job, find a spouse, and for their adult life to begin (Dhillon and Yousef, 2009; Singerman, 2007). This delay is perceived quite negatively, evidenced by sad anecdotes about strings of broken engagements due to lack of resources (Arraf, 2016) and recent public information campaigns from Alazhar, a prominent religious institution, promoting modesty and practicality in marriage negotiations instead of excessive displays of wealth (Ramadan, 2020; Abdullah, 2022).

Yet despite the undesirable delay to the marriage market in general, there appear to be several silver linings benefiting women. Women turning 18 during a downturn are more likely to complete secondary school, and the age gap between husbands and wives is slightly reduced for women whose likely marriage cohort—the cohort of men six years older than them—is scarred. Perhaps related to more schooling and later marriage, women report higher levels of decision-making in the marital home. These results suggest that while policy proposals emphasizing reducing the cost of marriage are important to reducing the barriers young people face to starting families, a balanced approach that does not promote teen marriage may be best.

These results are an important contribution to our understanding of how recessions can affect young people’s long-run outcomes in low and middle income countries. Even (or perhaps especially) in a context marked by high youth unemployment, coming of age during an economic downturn results in permanently lower odds of men reaching the highest rungs of the job ladder. This economic scarring also slows the transition

²For example, consider a young person who takes up street vending in order to bring in some income while unable to find wage work. This would be counted as self employment without any employees.

³In the 2018 ELMPS round, formal workers earned on average 29 percent higher wages than informal workers.

into married life, leaving unlucky cohorts on the fringes of a society where marriage is central. Other work has shown that economic conditions which block men from marriage are a considerable risk factor for civil unrest, emphasizing the importance of taking these scarring effects seriously (Hudson and Matfess, 2017). While Egypt and other countries where groom-to-bride payments are practiced have taken steps to lower the cost of marriage such as facilitating mass weddings, addressing the root issue of protracted joblessness and long term labor market scarring should remain a priority.

The rest of the paper proceeds as follows. Section 2 presents the newly digitized historical unemployment rates and the labor market panel dataset used to measure life outcomes. Section 3 describes the empirical strategy used to estimate the long run effect of recession entrance on labor and marriage outcomes, and Section 4 presents summary statistics before reporting results in Section 5. Section 6 discusses and concludes.

2.2 Data

To estimate the long-run impacts of entering the labor market during periods of high unemployment, I combine two main data sources. The first is a newly digitized series of sub-national unemployment rates, used to measure local economic conditions at labor market entry. National unemployment rates would be colinear with cohort effects, so sub-national rates are required. The second is individual-level labor and marriage market outcomes from the Egyptian Labor Market Panel Survey (ELMPS), which provides information on respondents' life outcomes along with age and location at entry.

2.2.1 Unemployment Rates

The Egyptian statistical agency, CAPMAS, reports national unemployment rates in a machine-readable format from 1991 to the present. Governorate-level (the largest administrative division in Egypt) unemployment rates can be calculated from the Labor Force Survey microdata, which is available beginning in 2006. Using cohorts from 2006 and later would significantly limit both the sample size and the extent to which long-run effects may be measured.

CAPMAS also publishes PDF-format statistical bulletins dating back to 1968 which contain the number of men and women, by governorate, who are employed, unemployed, or out of the labor force.⁴ With the help of an excellent research assistant, I have digitized these records. The quality of the scan varies from year to year (and page to page), so there are some years in which only 80 percent of records can be transcribed, while some years can be fully transcribed. This scanning difficulty introduces classical measurement error since the quality of a PDF scan is plausibly uncorrelated

⁴Records actually go back to 1957, although reports were not conducted annually during this time period, and since the earlier outcome data was collected in 1998, cohorts who entered the labor market prior to 1968 would not contribute meaningfully to the analysis.

to the unemployment rate. Observations where the digitized employment status statistics do not add up properly to match the reported total (beyond rounding error) due to transcription or other errors are excluded. Several years between 1968-2005 are unavailable on CAPMAS’s website, including 1975, 1976, 1985-1988, 1996, and 2003. (Years ending in 6 were census years.) Because the gap in the mid 1980s coincided with a recession, I conduct sensitivity analysis in Appendix Table B.4 where I drop all entrants before 1989. The results are largely similar.

To assess the reliability of the statistical bulletins and their transcription, figure 2.1 plots the national unemployment rate as reported by the PDF statistical bulletin alongside the unemployment rate as published in a machine readable format. For years where both statistics are available (1991-2004), they are nearly identical. The figure also depicts an upward trend in unemployment rates in the pre-1990 period. This could reflect changes in the survey definition of unemployment over time, although it is also consistent with structural transformation since movement away from agriculture and towards wage work results in higher rates of unemployment. Because the main specification includes year fixed effects, this is not a threat to identification.

Another important data prerequisite is that the unemployment rate varies sufficiently across governorate. Figure 2.2 reports residualized gendered unemployment rates, including fixed effects for year, governorate and gender. It is not the case that governorate unemployment rates are level shifts around the national rate, which ensures the data has sufficient variation to identify the parameters of interest.

2.2.2 Egyptian Labor Market Panel

Outcome data is drawn from the Egyptian Labor Market Panel, a longitudinal labor-focused survey with 4 waves (1998, 2006, 2012 and 2018). The survey is a nationally representative panel that tracks households as well as the daughter households formed when members establish new households (Krafft, Assad, and Rahman, 2019). Several thousand households are added to each wave as refresher samples. Table 2.1 reports the number of individuals in each combination of panel rounds, including all ages. For example, 6,636 individuals appear in the 1998 wave only, 4,143 appear in 1998 and 2006, 3,073 appear in 1998, 2006 and 2012, and 10,145 appear in all 4 waves. Once a household cannot be tracked they are not attempted in the following rounds, so individuals only appear in consecutive waves.

All waves have labor market modules with detailed questions on current labor market participation and job characteristics, plus modules on education, migration, assets, marriage and marriage payments, fertility, and women’s empowerment. Each wave also includes a detailed work history module which asks respondents about their history of jobs and non-wage characteristics of those jobs. These job histories allows for the creation of an annual job panel, which will constitute the main sample of this paper’s labor market analysis. The job history module includes questions about the following characteristics: employment status (wage work/self employment without employees/self employment with employees), sector of employment, permanent/temporary

job, whether the job had a contract, whether the job had social security, self-reported firm size, job tenure, industry, and occupation. Because the job history does not have recalled wages or earnings (presumably due to poor recall potential), these outcomes are only measured at the time of the survey.

Job spells are technically reported at the month level, but I aggregate to the year level to focus on longer-term dynamics and because start and end months of jobs are likely subject to more recall error than the years of a job spell. Indeed, over half of job spells reported in the 2018 wave include a year but are missing the month (respondent replied “don’t know”). This rate is even higher in previous waves. When assigning the employment status for a given year, the most chronologically recent status in that year prevails, which will downwardly bias scarring. For example, if an individual was unemployed from 2014-2015, had their first job spell from 2015-2017, followed by a spell of nonemployment from 2017-2018, 2014 would be marked as nonemployed, 2015-2016 employed, and 2017-2018 nonemployed.

Job characteristics are aggregated to the year level following a similar principle. For the 1998, 2006, and 2012 rounds, job histories were collected in a consecutive manner so there are no overlapping jobs in the data. Job characteristics in a particular year are given by the job characteristic of the most chronologically recent job held that year, conditional on being marked as employed that year according to the decision rule described above. The 2018 round allows for holding multiple jobs at once, and approximately 50 percent of job spells have months. Years where spells contain months are assigned according to whichever job was held longer, deferring to the most recent one in the event of ties. Years with job spells without months are assigned the characteristics of most chronologically recent spell. In cases where multiple job spells overlap, the year-level job characteristic is assigned based on the “best” concurrent spell, where best is defined by associations with higher earnings: public over private, formal over informal, wage work over self-employment, permanent over temporary, and large firm over small firm.

Detailed marriage outcomes are measured in the 2006, 2012 and 2018 ELMPS waves, and the earliest wave with marriage payment data is kept for each individual, following the rationale that recall quality declines with time. Each of these rounds measure the amount in Egyptian Pounds of the mahr and deferred mahr (financially similar to “brideprice” and a “deferred brideprice”/alimony, described in section 2.4), as well as the total value of jewelry presented to the bride from the groom/groom’s family. Also measured is the total value of furniture, appliances and housing purchased for the marital home and the contribution of the bride, bride’s family, groom, and groom’s family to those expenses. All outcomes are inflation adjusted.

An advantage of using recalled job histories from several waves of a panel is that recall accuracy can be evaluated by assessing the concordance between recalled job histories in different waves. For instance, one can examine whether respondents reported the same employment status in 2005 in both the 2006 and 2012 waves. Table 2.2 reports these measures of data quality in the ELMPS and the reliability of the

retrospective data.⁵

Reported demographic characteristics are highly consistent across waves: 99 percent of respondents report the same sex in each survey round, and 70 percent report the same year of birth. The average within-person range in reported birth years is 0.76, and the median respondent reports the same birth year in all rounds. Minor discrepancies in reported birth year may arise when parents misreport a child’s age in an early round, or among elderly respondents who are uncertain of their exact birth date (not uncommon among individuals born in this region in the early-to-mid 20th century). Respondents are less accurate in their recall of their age at first marriage. Among married respondents, 37 percent report the same age at first marriage across rounds, and the average within-person range in reported ages is 1.86 years. Misreporting of age at first marriage may reflect memory errors or social desirability bias.

To assess the reliability of recalled employment status, the row labeled “Share conflicting employment status” reports the proportion of person–year observations (restricted to ages 16 and above) for which at least two survey rounds provide inconsistent information on whether the respondent was employed. For example, a conflict arises when an individual reports being employed in 2005 in the 2006 wave but reports being nonemployed in 2005 in the 2012 wave. To be conservative, *any* conflicting report counts as a conflict, even if the other 2-3 survey rounds report the same status. When including a status conflict using any wave of data, the share of years with a conflict is 21 percent. When restricting to conflicting statuses between adjacent rounds, the conflict rate drops to 16 percent. The adjacent rate is the most relevant, since the sample is constructed in such a way that information is always taken from the chronologically closest survey. While not perfect, these statistics lend support to the basic reliability of retrospective data generated from the ELMPs.

2.3 Empirical Strategy

2.3.1 Labor Market Outcomes

Following the literature, I estimate regressions of individual labor market outcomes on the unemployment rate at labor market entry, controlling for year, cohort, potential experience, and governorate-of-birth fixed effects.⁶ To allow for cohort fixed effects, the specification uses governorate-level unemployment rates. Therefore, the effect is identified on variation in unemployment rate within governorate. There are 27 governorates, and as I show in section 2.4, there is significant variation in governorate level unemployment rate. The specification uses governorate of birth to avoid endogeneity

⁵The sample sizes in this table are smaller than the total sample since it restricts to individuals who appear in at least two rounds.

⁶One year fixed effect is omitted to avoid perfect multicollinearity among the age, time, and cohort effects. As a result, the fixed effects themselves are not interpretable, but the coefficient of interest remains identified.

from migration, and the unemployment rate at age 18 to abstract from decisions on when to enter the labor market.⁷ Potential experience refers to years since age 18.

The baseline specification estimates a single, total effect of entering the labor market during a period of higher unemployment. UR_{cg}^{entry} denotes the unemployment rate at age 18 in governorate of birth g for cohort c . Π_e , Ω_t , Δ_c , and Λ_g are fixed effects for potential experience, calendar time, cohort, and governorate of birth, respectively. Because these sets of fixed effects are collinear, one year dummy is omitted to ensure identification. The coefficient β_1 in equation (2.1) captures the average effect of entering the labor market when the unemployment rate at entry is one percentage point higher. Since contemporaneous unemployment rates are not included in this specification, β_1 reflects the effect of entry conditions given the typical time path of unemployment. This specification is used to examine educational attainment and other outcomes that vary little over time.

$$y_{icetg} = \beta_0 + \beta_1 UR_{cg}^{entry} + \Pi_e + \Omega_t + \Delta_c + \Lambda_g + \epsilon_{icetg} \quad (2.1)$$

To understand how scarring effects evolve over an individual's career, specification (2.2) estimates the effect of entry conditions at each year of potential experience and serves as the preferred specification. exp refers to years of potential experience since age 18. β_s estimates the effect of one percentage point higher unemployment on outcome y , s years after age 18. The effect at years 10+ of potential experience is aggregated, which allows for identification.

$$y_{icetg} = \beta_0 + \sum_{s=0}^9 \beta_s UR_{cg}^{entry} \cdot \mathbb{1}\{exp = s\} + \beta_{10+} UR_{cg}^{entry} \cdot \mathbb{1}\{exp \geq 10\} + \Pi_e + \Omega_t + \Delta_c + \Lambda_g + \epsilon_{icetg} \quad (2.2)$$

As a robustness check, I also estimate specification (2.3), which additionally controls for contemporaneous unemployment rate in the governorate of birth.

$$y_{icetr} = \beta_0 + \sum_{s=0}^9 \beta_s UR_{cr}^{entry} \cdot \mathbb{1}\{exp = s\} + \beta_{10+} UR_{cr}^{entry} \cdot \mathbb{1}\{exp \geq 10\} + \beta_c UR_{rt}^{contmp} + \Pi_e + \Omega_t + \Delta_c + \Lambda_r + \epsilon_{icet} \quad (2.3)$$

2.3.2 Marriage Outcomes

Because marriage is an absorbing state for most individuals in the sample, a different empirical strategy is needed. It is possible to test whether recession entry delays or accelerates marriage by regressing age at first marriage on the unemployment rate at entry using specification (2.1). However, this outcome is missing for individuals who are still single, which introduces bias. Instead of restricting the sample to older cohorts who have had sufficient time to get married in the data, I use a cox proportional haz-

⁷18 is the median age at which men and women complete secondary school in the 2018 ELMPs.

ards model to estimate how recession entrance effects the hazard rate into marriage. The standard assumption is that the hazard rate, or the probability of an unmarried person at a given age getting married that year, is affected proportionally over time. This would be violated if the marriage rate was lowered at younger ages by recession entrance but then accelerated later in order to catch up. When proportionality is violated, the estimate can be interpreted as the time weighted average of the effect of recession entrance on marriage rates (Stensrud and Hernán, 2020). Therefore, the bias from any plausible violation of proportional hazards would bias the estimates towards zero.

The fact that grooms are on average six years older than their brides necessitates careful consideration of *when* exactly recessions matter for women vs men with regards to family formation. To the extent that men are typically the primary breadwinner and pay the largest financial marriage costs, it is possible that women might be most affected by the recession scarring of their “likely marriage cohort,” not themselves. To account for this, analysis of recession scarring on women’s marriage outcomes will include specifications using the unemployment rate at 18 and at age 12, when their likely marriage cohort was turning 18.

Other outcomes related to spousal characteristics, marriage cost amounts, and marital decisionmaking will be analyzed using specification 2.1, restricting to individuals above the 95th percentile of age at first marriage (29 for women and 33 for men).

2.4 Context and Descriptive Statistics

The Egyptian labor market is marked by moderate rates of educational attainment, high employment rates for men, and low employment rates for women. Table 2.3 reports descriptive statistics from each wave of the ELMPS by gender. There have been persistent increases in educational attainment across time, and the gender gap in educational attainment shrunk considerably between 1998 and 2018. In 1998, 54 percent of men and 34 percent of women aged 25-65 had completed grade 9. 47 percent of men and 30 percent of women had completed secondary, and 18 percent of men and 9 percent of women had completed university. By 2018, approximately half of women 25-65 had completed secondary and the gap in college attendance between men and women had nearly closed.

Across all waves, approximately of 90 percent of men aged 25-65 are employed, while only 20-28 percent of women are employed. However, this statistic conceals considerable heterogeneity by age. High youth unemployment is a key characteristics of the Egyptian labor market. Panel A of figure 2.3 reports the overall national unemployment rate and the national youth unemployment rate (ages 15-24). The overall unemployment rate ranges between 7 and 13 percent, while the youth unemployment rate ranges between 17 and 34. Panel B displays this high youth unemployment at the individual level, as the gap between the year individuals in the 2018 ELMPS exited school and the year they started their first job. The figure shows the long period of

waitthood quite strikingly. Of ever-working men with at least a secondary school education and who last exited schooling between the ages of 14 and 30, 20 percent held a job before they exited school, 40 percent found their first job within 1 year of exiting school, and approximately 35 percent begin their first job between 2-10 years after exiting school. Ever working women with the same qualification have ever longer gaps between school and work; 35 percent start their first job within 1 year of graduation, and approximately 30 percent begin their first job 2-10 years after exiting school. An additional 10 percent don't begin working for more than 10 years after exiting school.

There has been considerably less change in family formation timelines than educational attainment during this time. The average age at first marriage for married men aged 25-65 has remained fairly constant around 27, while the age at first marriage for women has increased slightly over time. In 1998, the average age at first marriage was 20, while in 2018 it had risen to 21.4

Egyptian marriages are accompanied by various payments and gifts, some of which are customary (such as gold jewelry) and others which are dictated by prenuptial religious marriage contracts. Except for foreigners, all marriages in Egypt are governed by religious family law pertaining to the couple's religion. I focus on Islamic marriages as only 5 percent of individuals in the ELPMS are Christian. Islamic jurisprudence requires a payment from the groom to the bride known as *mahr*, which is the sole property of the bride. Islamic marriage contracts also stipulate a *mahr muakhr* (deferred mahr) which is owed to the bride in the event of her husband's death, no-fault divorce initiated by the husband, or fault-divorce proven in court. Prior to 2000, no-fault divorce only existed for men, and women could only initiate a divorce in the event of abuse, abandonment, unfaithfulness, undisclosed impotence, and other related reasons. No-fault divorce for women (*khul'*) was introduced in 2000, but the wife forfeits the deferred mahr and must repay the mahr if she chooses this type of divorce (Corradini and Buccione, 2023). As such, the deferred mahr is similar but not identical to alimony negotiated in a prenup.

Another customary marital contract unique to Egypt is the *ayma*, or inventory of goods and wealth brought into the marriage by the wife, plus any wedding gifts bestowed by the groom, all of which is agreed to be the property of the wife (El-Kholy, 2000; Salem, 2018). Its value is negotiated up front between the bride and groom and their families, and enumerates what the wife is owed upon death or divorce. Unlike the mahr muakhr, which is enforceable under personal status laws, the ayma falls under civil contracts related to safekeeping of property. As a result, these contracts could be easier to enforce in practice. According to El-Kholy (2000), negotiating an ayma is more common among working class couples and is seen as an important form of economic protection for women.

The ELPMS measures these contractual amounts as well as costs associated with jewelry, furniture, gihaz (*preparations*), housing, and the wedding itself, plus the portion of these costs borne by the bride or groom's side. Summary statistics of these costs are reported in Panel B of table 2.3. Total costs of marriage declined from 2006 to 2012, then rose again when measured in 2018. Ayma amounts has risen, while

inflation-adjusted mahr amounts have fallen, consistent with the practice becoming more ceremonial than economically meaningful.

Panel B of table 2.3 also reports how other marital characteristics have changed over time. The average age gap between husbands and wives is approximately 6 years and has fallen slightly over time. In 2006, approximately half of couples lived alone (i.e., without in-laws) in the beginning of their marriage, which has risen to 63 percent in 2018. 30 percent of couples are related by blood in some capacity (first cousins or more distant relatives), and this has remained stable over time. Finally, polygamy is rare, with only 1 percent of married households in the ELMPS including more than 1 wife.

2.5 Results

2.5.1 Education

Table 2.4 reports the effect of entry conditions on educational attainment using specification (2.1). Columns 1 and 2 report the effect of an additional percentage point of unemployment at age 18 on the probability of completing secondary or post-secondary, restricting to individuals age least age 25. Column 3 reports the effect on completing university, restricting to those aged 28+. There are no statistically significant effects for men at any level of schooling, suggesting that young men do not respond to poor labor market conditions by extending their education (“sheltering”) or by leaving school early to supplement household income.

In contrast to men, women attain higher levels of education during recessions. Women turning 18 during a recession are 0.81 percentage points more likely to complete secondary school for each additional point in the unemployment rate at entry. The peak to trough of the average business cycle in Egypt is approximately 4 percentage points, meaning that the average recession increased the probability of a woman completing secondary school by 3.24 percentage points. Results described below point to a delayed marriage market, which may explain the increase in educational attainment. As the age at first marriage for women is around 20, a delayed marriage market may have allowed women to stay in school longer. It is also possible that the recession encouraged women to pursue education with higher expected labor market returns, and that they were less constrained by immediate household financial needs than similarly aged men.

2.5.2 Employment

With null effects on educational attainment, the effects of a recession on men’s employment outcomes can be straightforwardly interpreted. Table (2.5) reports estimates from specification 2.2, which regresses annual job outcomes on the governorate of birth unemployment rate at entry interacted with years of potential experience dummies, plus fixed effects for cohort, year, and governorate of birth. The coefficient on each

potential experience dummy interacted with the unemployment rate at entry can be interpreted as the impact of an additional percentage point of unemployment at entry, at s years of potential experience. These coefficients allow us to trace the long run effect over time. Column 1 reports the effects on an indicator for having any job, including self employment. There are statistically significant negative effects on the probability of employment for two years following entry. An additional percentage point of unemployment at entry lowers the probability of employment that year by 0.9 percentage points (almost 1:1), the following year by 0.6 percentage points, and the year after by 0.3 percentage points. By six year, there is a statistically significant and *positive* effect on the probability of employment that grows over time. In the long run (10+ years after entry), recession entrants are 0.5 percentage points *more* likely to be employed for each additional percentage point of unemployment.

This positive long-run result is somewhat counterintuitive, and no recent papers on recession scarring in high-income countries find positive long-run effects. However, this result masks important heterogeneity on the type of employment. Column 2 reports impacts on the probability of being employed in wage work (with anything else, including self employment or non-employment marked as zero). Recession entrants are less likely to be engaged in wage work for 3 years after entry, and there is no long-run positive impact. Instead, recession entrants shift towards self employment. By four years after entry, there is a statistically significant increase in the probability of being self employed, which grows over time and remains significant in the long run. At 10+ years after entry, recession entrants are more likely to be self-employed by 0.3 percentage points for each additional percentage point in the unemployment rate. Multiplied by four, this is associated with a total effect of 1.4 percentage points more likely to be self employed in the long run.

Some types of self-employment are associated with more income than others. Specifically, owning a small business with employees is very different from the casual self-employment that many turn to to make ends meet, such as street vending, odd jobs, etc. Column 4 estimates recession entry effects over time on the probability of being self-employed without any employees. Almost all of the long-run effect on self employment is concentrated in this more casual form of self-employment, associated with more precarity and lower earnings. These effects are consistent with long-term discouragement from the market for wage work.

Male recession entrants are also less likely to hold formal wage work in the medium-to long-term, which provides additional evidence of slowed job ladder growth. The negative effect becomes statistically significant four years after entry, and only converges to zero in the 10+ year long-run. The fact that this effect only emerges 4 years after age 18 suggests it may be concentrated among the college educated population (who are more likely on average to hold formal jobs). This question is investigated in suggestive heterogeneity analysis below.

There are almost no employment recession scarring effects for women, which is perhaps not unexpected given the low rates of labor force participation. With the increase in educational attainment, it is also possible that the null employment im-

pacts are the net result of recession scarring plus a boost from increased educational attainment. The only statistically significant impact reported in Table 2.6 is a long-run positive impact on the probability of holding a formal waged job. 10+ years after entry, women entering during a recession are 1.7 percentage points more likely to hold a formal waged job. The two most common formal jobs for women in the 2018 ELMPS are administrative assistant and elementary school teacher. A plausible explanation consistent with this finding is that, as educational attainment rises, women in their 30s with school-aged children may be more likely to take up teaching or administrative positions. Nonetheless, this is a single significant estimate among many null results, and should therefore be interpreted with caution.

Heterogeneity by Educational Attainment

This section examines heterogeneity by educational attainment in the scarring effects on men. Because recession entrance had no statistically significant impacts on men’s educational attainment at any level, recession-induced changes in who selects into higher educational are likely to be minimal. Nonetheless, these results should be viewed as suggestive.

Tables B.2 and B.3 report scarring impacts on men attaining secondary school and university, respectively. Both tables report largely null impacts on employment, but large negative impacts on the probability of having a formal job that remain significant in the long run. 10+ years after age 22, college grads remain 0.9 percentage points less likely to have a formal job for each percentage point increase in the unemployment rate at entry. Effects are slightly smaller but still large for secondary grads - 10+ years after age 18, secondary grads are 0.6 percentage points less likely to have a formal job.

Table B.1 examines the impact of recession entry on men who did not complete secondary school. There are no statistically significant impacts on employment or wage work vs self employment, but column 4 reports statistically significant increases in the probability of having a formal wage-work job. This is somewhat unexpected, although column 5 clarifies that much of this impact is concentrated among public enterprises (which is a separate sector code from “government” work). This sector categorization likely includes state enterprises and active labor market programs like the Public Works Programme (Semlali and Angel-Urdinola, 2012; Egyptian Social Fund for Development, 1999). Only 4 percent of male, formal wage workers with less than a secondary education report this sector of employment in the 2018 wave, while 47 percent of these workers are in the government sector. This result suggests that public enterprises play a non-trivial role in bolstering access to social security and the safety net for lower skill workers.

2.5.3 Wages and Earnings

The forgoing results suggest large earnings losses, especially for workers with at least a secondary degree. Unfortunately, there are much fewer observations with wage

or earnings information since these outcomes are only measured at the time of the survey and are excluded from the survey’s job history section. To maximize power and focus on aggregate impacts, table 2.7 reports results from regressions of earnings or wages, measured in one of the 4 survey rounds, on the unemployment rate at entry using (2.1). These specifications reveal the aggregate impact across the whole career. Column 1 examines the probability of having earnings below the (male/female) median. Columns 2 and 3 measure scarring effects on earnings below the 75th and 90th percentile, and column 4 reports effects on log hourly wages. Column 1, row 1 shows a statistically significant increase the probability of having below median earnings for men who enter during recessions. Scaling by the average 4 point increase in the unemployment rate during a recession, a recession entrant is, on average across the entire career, 1.3 percentage points more likely to have below median earnings. This effect appears to be stronger for college grads, for whom the effect is larger and statistically significant at the 75 and 90th percentiles as well. The average recession college grad is less likely by 2.6 percentage points to have earnings above the 90th percentile. These results reinforce the earlier evidence on employment and job type, indicating that recession entrants are less likely to obtain formal or wage employment and experience persistent earnings penalties over the long run.

2.5.4 Marriage Market Outcomes

Table 2.8 shows that recessions significantly delay marriage for both men and women. The table reports hazard ratios which reflect the impact of treatment on the relative hazard rates. Column 1 reports the effect of unemployment rate at age 18 on the time to marriage for men. The point estimate implies that each additional percentage point in the unemployment rate at age 18 lowers the probability of marriage in a given year by 1.5 percent for men. The implied survivor functions from this model help illustrate the magnitude of these effects. Among men aged 25, 30.8 percent of those entering at the 25th percentile of entry unemployment rates are married, compared to 28.7 percent at the 75th percentile. At 30 years old, the gap widens slightly: 70.8 percent of men are married at the 25th percentile of entry conditions vs 67.6 percent of men at the 75th percentile. These magnitudes are meaningful, as they represent thousands of young men for whom the entry into socially sanctioned intimate relationships and the cultural signifier of adulthood is delayed.

The impact of recession entrance has a similar magnitude of impact among women. Columns 3 and 4 reports the effect of the unemployment rate at age 18 and age 12 (the year that the “likely marriage cohort” turned 18). Both point estimates are statistically significant, and correspond to a 1.1-1.3 percent decrease in the probability of getting married each year. The fact that both estimates are significant suggests that women are affected by recession scarring on the cohort of men they will marry and by recessions at the time of their own labor market/marriage market entry. In other words, recessions can delay marriage in two ways: first, by lowering men’s early career labor market earnings, and by creating financial strain (for everyone) around the time of marriage.

Column 2 reports the effect of the unemployment rate at age 12 for men as a placebo test. Reassuringly, there is no impact.

Table 2.9 reports estimates on the effect of recession entrance on spousal characteristics and marriage costs. Each row/column corresponds to a separate regression. For individuals in multiple rounds, the earliest observation is kept to reduce recall bias. There are no statistically significant effects for men, although the sample sizes for these regressions are smaller since men were not surveyed on the costs of marriage until the 2012 round.

For women, increased unemployment rates at age 18 do not impact spousal characteristics but do increase total groom spending, with much of the increase coming from increased spending on home furnishings, appliances etc. As the price of gold often increases dramatically during times of inflation, there is anecdotal evidence that grooms shift spending toward home furnishings and appliances in lieu of unaffordably expensive gold (Kholaif and El-Ghobashy, 2016). It is also possible families demand more appliances up front instead of buying later to protect their daughters, in light of tougher economic circumstances (Arraf, 2016).

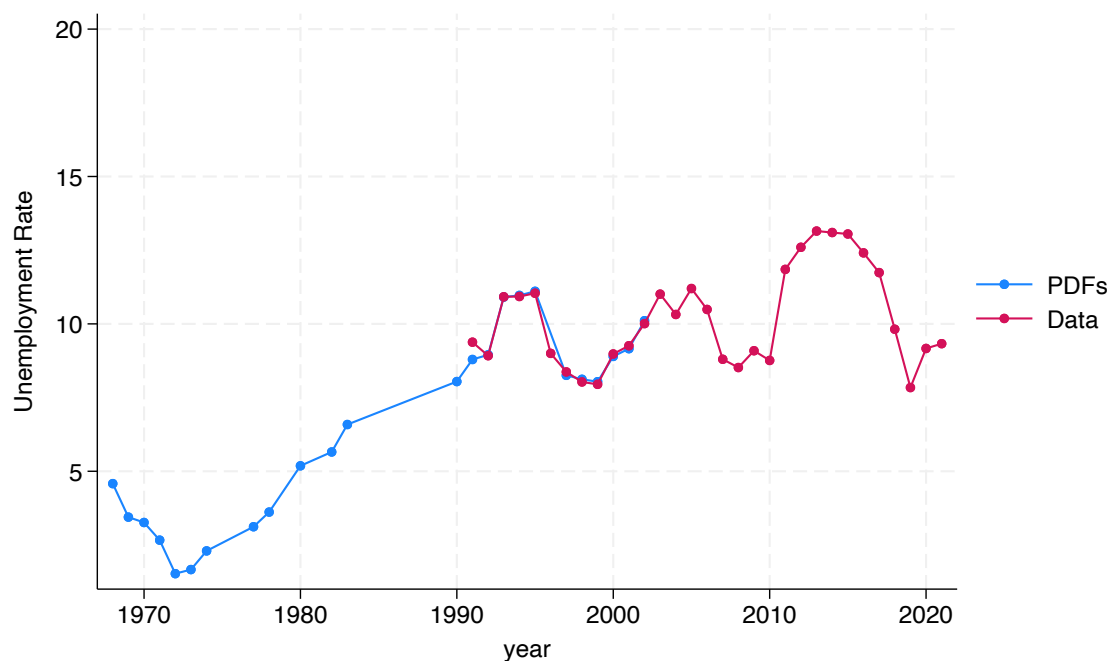
The unemployment rate when the bride was 12 is also associated with increased spending on home furnishings, and a relatively small but significant reduction in the average age gap between husbands and wives. This effect on the age gap could be caused by women waiting slightly longer than men as they wait for their marriageable cohort to accumulate the funds for marriage. Table 2.10 also reports that an increase in the unemployment rate at age 12 increases women's decisionmaking power in their eventual marriages. This could be due to the increased age at first marriage or increases in educational attainment shown above.

2.6 Conclusion

Despite the severity of youth unemployment in many developing countries, there has been little work to date on the effect of entering the labor market during a recession on long run outcomes in low and middle income countries. There is also virtually no work on how recession entrance affects family formation in cultural settings practicing marriage payments. I fill this gap by digitizing historical Egyptian unemployment rates and estimating scarring effect on both labor and marriage market outcomes. I find scarring effects on the male employment rate which persist for approximately 5 years before converging to zero. Male recession entrants who are employed are less likely to work in the public sector and significantly more likely to work in small firms. Women are more likely to work in temporary jobs and less likely to work in the formal sector. Both men and women who come of age during a recession delay marriage significantly.

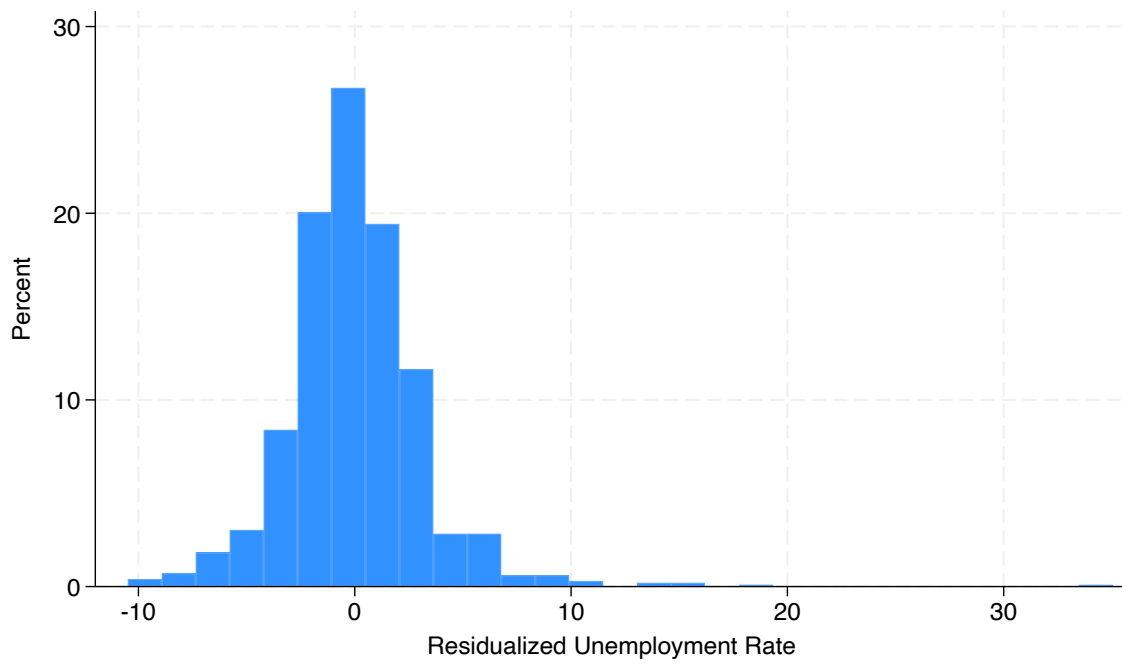
2.7 Figures

Figure 2.1: National Unemployment Rates



Note: This figure plots annual unemployment rates as reported in official statistics from the ILO online ("Data") and according to the digitized bulletins ("PDFs"). Official unemployment rates are available in a machine readable format from 1991, and the bulletins have data for 1968-1974, 1977-1984, 1990-1995, 1997-2002, and 2004. The close overlap between these two sources of data supports the reliability of the bulletins.

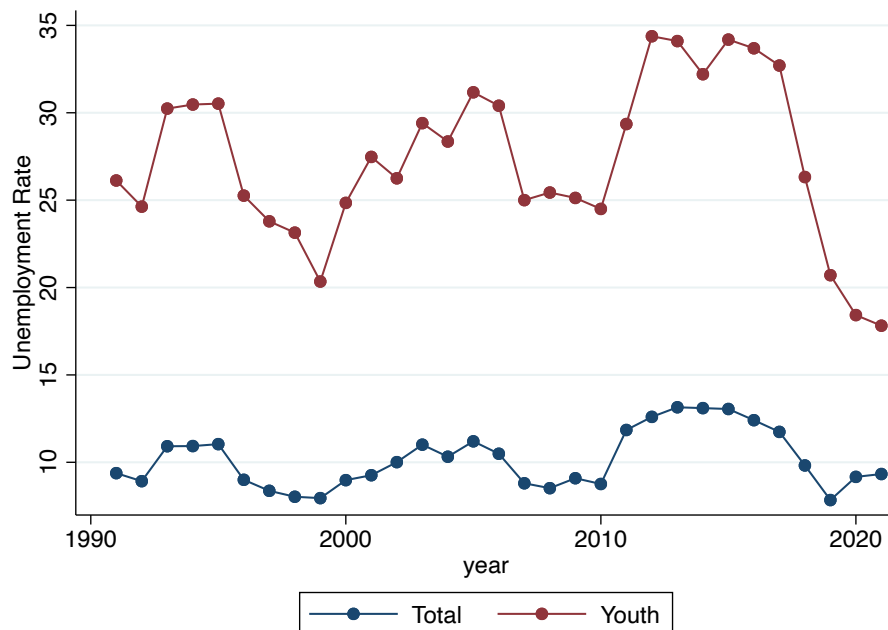
Figure 2.2: Residualized Governorate Unemployment Rates



Note: This figure reports residualized governorate unemployment rates after including year and governorate fixed effects. The dispersion pictured here ensures adequate identifying variation in the regressions of interest.

Figure 2.3: Two Views of Youth Joblessness

Panel A: National unemployment rate, youth vs total



Panel B: Years between graduation and labor market entry, ELMPS 2018



Note: Panel A reports official annual unemployment rates for the working age population (total) and youth (aged 15-24) in Egypt. Panel B reports the difference between the year the respondent exited schooling and the year they began their first job for all individuals who have completed schooling and ever held a job in the 2018 Egyptian Labor Market Panel Survey.

2.8 Tables

Table 2.1: ELMPS Observation by Round

Round	Freq.
98	6,636
06	4,227
12	4,164
18	22,078
98 & 06	4,143
06 & 12	2,796
12 & 18	16,252
98 & 06 & 12	3,073
06 & 12 & 18	12,756
98 & 06 & 12 & 18	10,145

Note: This table reports the number of individuals present in each combination of rounds of the Egyptian Labor Market Panel Survey. For example, there are 6,636 individuals who were only surveyed in the 1998 round and 10,145 individuals surveyed in all four rounds. Individuals lost in one round were not attempted in following rounds.

Table 2.2: Data quality

	mean	p50	p75	p95	min	max	count
Sex consistent =1	0.99	1	1	1	0	1	49165
YOB consistent =1	0.71	1	1	1	0	1	49165
YOB inconsistencies	0.76	0	1	4	0	81	49164
Age at 1st mar consistent =1	0.38	0	1	1	0	1	18348
Age 1st mar inconsistencies	1.86	1	2	8	0	38	18358
Share conflicting employment status	0.21	0.06	0.32	1.00	0	1	31378
Share conflicting employment status (adjacent)	0.16	0.00	0.25	0.78	0	1	19704

Note: Sex consistent reports whether a person's sex is the same across all surveys. YOB consistent reports whether a person's year of birth is the same across all surveys. YOB inconsistencies reports summary statistics on the range of a given person's reported birthdays. Age at first marriage consistent =1 reports whether a person's year of first marriage is the same across all surveys (conditional on being nonmissing). Age at first marriage inconsistencies reports summary statistics on the range of a given person's reported age at first marriage. Share conflicting employment status reports the share of years in which an individual reported two different employment statuses in different survey waves, restricting to years where the individual was 16 or older. To be conservative, a year is counted as conflicted if reports from at least one survey differed, even if the other 2-3 reports are consistent. The following row reports the same statistic, restricted to consecutive surveys. The final row is the most relevant statistic since the retrospective panel is constructed using the closest survey round.

Table 2.3: Descriptive Statistics from each ELMPS Wave

	1998		2006		2012		2018	
	Men	Women	Men	Women	Men	Women	Men	Women
<i>Panel A: Demographics</i>								
Attained Preparatory	0.54	0.34	0.59	0.43	0.63	0.50	0.66	0.54
Attained Secondary	0.47	0.30	0.54	0.39	0.58	0.46	0.61	0.49
Attained Post-Secondary	0.25	0.14	0.25	0.16	0.23	0.18	0.22	0.19
Attained University	0.18	0.09	0.20	0.12	0.19	0.15	0.19	0.16
Employed (past 3 months)	0.88	0.22	0.91	0.28	0.92	0.23	0.90	0.21
Age at First Marriage	26.80	20.08	27.03	20.94	26.80	21.32	26.73	21.39
Observations	12095	11902	18587	18553	24483	24703	30542	30688
<i>Panel B: Marriage Characteristics and Costs (2010 EGP)</i>								
Husband-wife age gap			6.73	6.78	6.31	6.33	6.19	6.09
Husband-wife age gap			6.93	6.96	6.58	6.61	6.54	6.59
Couple lives alone				0.55	0.65	0.63	0.64	0.63
Consanguineous				0.30	0.30	0.30	0.28	0.28
Polygamous wife				0.01		0.01		0.01
Mahr				5,863	1,772	2,044	2,518	2,699
Deferred Mahr				9,829	6,402	6,915	7,174	8,020
Ayma				9,829	37,999	38,012	48,877	59,823
Jewelry				8,608	6,891	7,202	9,215	8,921
Furniture				38,613	23,697	24,310	24,040	25,687
Gihaz				14,682	12,731	13,518	19,443	18,727
Housing				34,347	27,554	28,854	26,750	37,463
Wedding				6,182	6,207	6,322	6,865	19,763
Total bride side				27,543	13,153	13,508	24,954	28,718
Total groom side				78,509	33,814	32,382	57,733	74,972
Observations			0	6,590	4,654	6,367	4,441	4,840

Note: Panel A reports summary statistics from each round of the ELMPS for men and women aged 25-65. Preparatory is equivalent to grade 9, and post-secondary includes any post-secondary or technical degree, typically shorter than a bachelors.

Panel B presents marriage summary statistics for rounds where it is available: a marriage section was added for women in 2006 and added for both genders in 2012. Observations are at the individual level. In 2018, it was administered to all ever-married individuals aged 15-59. The age restriction was 18-39 in 2012 and 16-49 in 2006. Husband-wife age gap is the difference in ages in years between the husband and wife (so older husband is positive). |gap| is the absolute value of the quantity. Consanguineous is an indicator equal to one if the individual is married to a blood relative, including first cousins and more distant relatives. Polygamous wife is an indicator equal to one if the respondent has a co-wife. The subsequent rows report the cost of various wedding expenses or financial sums negotiated over in the marital contract, which are explained in further detail in section 2.4

Table 2.4: Effect of Entry Conditions on Educational Attainment

	Secondary	Post-secondary	University
<i>Men</i>			
Gov. of Birth, age 18 UR	0.00156 (0.00175)	-0.00194 (0.00191)	-0.000244 (0.00164)
Observations	13092	13092	11051
<i>Women</i>			
Gov. of Birth, age 18 UR	0.00808** (0.00301)	0.000833 (0.00143)	-0.00104 (0.00160)
Observations	12688	12688	10599

Note: This table reports results from a regression of an indicator for having attained at least secondary, post-secondary or university education on the unemployment rate at age 18 using specification 2.1. There is one observation per individual, including the highest education they report in any round. Individuals aged 25+ at the time their highest observed education is measured are included in the secondary and post-secondary regressions; individuals aged 28+ are included in the university regressions.

Table 2.5: Long-run Effect of Entry Unemployment Rate on Male Employment

	(1) Employment	(2) WageWork	(3) Self Emp	(4) Self (no employees)	(5) Formal WW
UR*exp=0	-0.00706*** (0.00238)	-0.00562*** (0.00177)	-0.00142 (0.00162)	-0.00247 (0.00149)	-0.000891 (0.00177)
UR*exp=1	-0.00428** (0.00176)	-0.00345** (0.00152)	-0.000818 (0.00129)	-0.00208 (0.00130)	-0.000245 (0.00164)
UR*exp=2	-0.00121 (0.00185)	-0.00184 (0.00159)	0.000623 (0.00124)	-0.000758 (0.00135)	-0.000179 (0.00150)
UR*exp=3	-0.00170 (0.00171)	-0.00274 (0.00163)	0.00101 (0.00114)	-0.000100 (0.00111)	-0.00110 (0.00142)
UR*exp=4	-0.00108 (0.00189)	-0.00250 (0.00182)	0.00140 (0.00102)	0.000655 (0.000984)	-0.00255* (0.00135)
UR*exp=5	0.00134 (0.00159)	-0.000933 (0.00145)	0.00226* (0.00124)	0.00123 (0.00115)	-0.00300** (0.00126)
UR*exp=6	0.00170 (0.00133)	-0.000586 (0.00133)	0.00227** (0.00107)	0.00146 (0.000987)	-0.00300** (0.00110)
UR*exp=7	0.00197 (0.00128)	-0.000679 (0.00136)	0.00264** (0.00111)	0.00152 (0.00100)	-0.00330** (0.00121)
UR*exp=8	0.00226** (0.00107)	-0.000976 (0.00133)	0.00323*** (0.00104)	0.00171 (0.00101)	-0.00348** (0.00135)
UR*exp=9	0.00272** (0.00102)	-0.000647 (0.00125)	0.00337*** (0.00101)	0.00190* (0.00105)	-0.00337** (0.00146)
UR*exp=10+	0.00418*** (0.00144)	0.00118 (0.00130)	0.00302** (0.00130)	0.00290* (0.00163)	-0.00101 (0.00124)
Constant	0.779*** (0.00633)	0.595*** (0.00727)	0.184*** (0.00657)	0.107*** (0.00762)	0.261*** (0.00751)
Current Year	Yes	Yes	Yes	Yes	Yes
Potential Experience	Yes	Yes	Yes	Yes	Yes
Cohort	Yes	Yes	Yes	Yes	Yes
Gov of birth	Yes	Yes	Yes	Yes	Yes
Clustering	Gov Birth	Gov Birth	Gov Birth	Gov Birth	Gov Birth
Entry Yrs	1968-2018 minus 75,76,85-89,96,03				
Obs	305260	304203	304840	304203	304176

Note: This table reports estimates from regression (2.2), which traces out the effect of the unemployment rate at entry over the course of an individual's career. The coefficient associated with UR*exp=n reports the effect of an additional point in the unemployment rate at n years of potential labor market experience. Employment is an indicator equal to one if the respondent has wage work or self employment in a given year. Wage work is an indicator equal to one if the respondent had a wage work job that year. Self employment is an indicator equal to one if the respondent was self employed that year. Self (no employees) restricts to self employment without an employees. Formal wage work refers to wage work with a contract and social security.

Table 2.6: Long-run Effect of Entry Unemployment Rate on Female Employment

	(1) Employment	(2) WageWork	(3) Self Emp	(4) Self (no employees)	(5) Formal WW
UR*exp=0	0.00243 (0.00252)	0.000621 (0.00166)	0.00179 (0.00151)	0.00140 (0.00126)	0.00165 (0.00160)
UR*exp=1	0.00249 (0.00255)	0.000789 (0.00168)	0.00168 (0.00150)	0.00131 (0.00127)	0.00160 (0.00163)
UR*exp=2	0.00212 (0.00260)	0.000553 (0.00168)	0.00157 (0.00154)	0.00121 (0.00131)	0.00147 (0.00154)
UR*exp=3	0.00177 (0.00253)	0.000647 (0.00163)	0.00112 (0.00154)	0.000748 (0.00130)	0.00142 (0.00156)
UR*exp=4	0.00112 (0.00260)	0.000231 (0.00163)	0.000888 (0.00154)	0.000560 (0.00133)	0.000794 (0.00158)
UR*exp=5	0.000652 (0.00262)	0.000125 (0.00162)	0.000515 (0.00152)	0.000219 (0.00131)	0.000434 (0.00162)
UR*exp=6	0.000564 (0.00246)	0.000384 (0.00147)	0.000168 (0.00151)	-0.0000964 (0.00131)	0.000337 (0.00156)
UR*exp=7	0.000528 (0.00246)	0.000163 (0.00150)	0.000327 (0.00157)	-0.0000258 (0.00138)	0.000694 (0.00159)
UR*exp=8	0.0000669 (0.00239)	0.0000542 (0.00154)	-0.0000527 (0.00153)	-0.000351 (0.00135)	0.000902 (0.00152)
UR*exp=9	-0.000340 (0.00224)	-0.000134 (0.00144)	-0.000259 (0.00153)	-0.000493 (0.00136)	0.00110 (0.00145)
UR*exp=10+	0.00174 (0.00190)	0.00237 (0.00160)	-0.000656 (0.00124)	-0.000726 (0.00117)	0.00400** (0.00179)
Constant	0.176*** (0.0146)	0.114*** (0.00950)	0.0625*** (0.00907)	0.0600*** (0.00812)	0.0754*** (0.00982)
Current Year	Yes	Yes	Yes	Yes	Yes
Potential Experience	Yes	Yes	Yes	Yes	Yes
Cohort	Yes	Yes	Yes	Yes	Yes
Gov of birth	Yes	Yes	Yes	Yes	Yes
Clustering	Gov Birth	Gov Birth	Gov Birth	Gov Birth	Gov Birth
Entry Yrs	1968-2018 minus 75,76,85-89,96,03				
Obs	312079	311093	311866	311093	311088

Note: See the table notes for table 2.5.

Table 2.7: Effect of Entry Conditions on Earnings and Wages

	Earnings < p50	Earnings < p70	Earnings < p90	Log Wages
<i>Men, all</i>				
Gov. of Birth, age 18 UR	0.00318** (0.00114)	0.00209 (0.00136)	0.00114 (0.000747)	0.0000338 (0.00243)
Observations	25472	25472	25472	18242
<i>Men, < secondary></i>				
Gov. of Birth, age 18 UR	0.00476 (0.00288)	0.000182 (0.00210)	0.00136 (0.000933)	-0.000185 (0.00364)
Observations	8322	8322	8322	6487
<i>Men, secondary</i>				
Gov. of Birth, age 18 UR	-0.00140 (0.00189)	0.000307 (0.00206)	-0.000944 (0.00120)	0.00281 (0.00374)
Observations	10524	10524	10524	7234
<i>Men, university</i>				
Gov. of Birth, age 22 UR	0.00441** (0.00183)	0.00873*** (0.00205)	0.00645*** (0.00226)	-0.00556 (0.00573)
Observations	6091	6091	6091	4051
<i>Women, all</i>				
Gov. of Birth, age 18 UR	0 (.)	0 (.)	-0.0000462 (0.00116)	-0.00522 (0.00543)
Observations	31310	31310	31310	3914

Note: This table reports results from a regression of the outcome of interest on the unemployment rate at age 18 using specification 2.1. This table uses data only measured contemporaneously in the 4 panel rounds, not in the retrospective job histories. Earnings < p50 is an indicator for having earnings below the median level of male (female) earnings that year, which includes all earnings from wage and self employment. Log wages corresponds to the log of the implied hourly wage calculated from earnings and hours.

Table 2.8: Hazard into Marriage

	Men		Women	
Gov. of Birth, age 18 UR	0.985*** (0.00451)		0.989** (0.00507)	
Gov. of Birth, age 12 UR		0.991 (0.00774)		0.987** (0.00571)
Cohort	Yes	Yes	Yes	Yes
Gov of birth	Yes	Yes	Yes	Yes
Clustering				
Obs	19859	20455	20110	20700

Note: This figure reports hazard ratios from a Cox proportional hazards model on the effect of unemployment rate at entry on the time to marriage. A one point increase in the unemployment rate at entry reduces men's probability of getting married each year by 1.1 percent.

Table 2.9: Effect of Recession Entry on Marriage Outcomes

	Men			Women, UR at 18			Women, UR at 12		
	Effect	Mean	N	Effect	Mean	N	Effect	Mean	N
Husband-wife age gap	0.00 0.03	6.67	6,445	-0.04 0.03	6.37	7,785	-0.06* 0.03	6.37	6,880
Husband-wife age gap	0.00 0.03	6.91	6,445	-0.03 0.03	6.66	7,785	-0.05 0.03	6.66	6,880
Couple lives alone	-0.00 0.01	0.62	4,663	0.00 0.00	0.59	7,677	0.00 0.00	0.59	7,406
Consanguineous	-0.00 0.00	0.30	4,663	-0.00 0.00	0.30	7,677	0.00 0.00	0.29	7,406
Mahr	73.90 106.42	1,973	2,467	100.25 139.29	3,894	5,560	-123.69 273.85	4,760	5,121
Mahr Muakhr	162.60 183.64	7,215	2,519	124.06 151.16	8,683	5,950	-191.17 259.56	9,299	5,594
Ayma	14.71 616.91	40,359	2,649	-335.65 213.01	40,070	4,262	351.30 354.28	39,924	3,518
Jewelry	156.90 168.90	7,919	2,785	-13.84 100.82	8,129	6,320	-232.27 193.80	8,633	5,950
Furniture	24.20 295.37	26,381	2,610	579.30* 305.56	32,958	6,139	1,051.59** 447.83	35,323	5,801
Gihaz	436.66 288.39	15,059	2,510	-109.35 129.76	14,761	6,152	16.23 185.85	15,624	5,814
Housing	-497.14 1,108.89	30,658	2,465	918.55 593.29	31,192	5,909	479.77 1,081.21	32,912	5,607
Wedding	46.06 199.84	6,937	2,634	17.77 72.62	6,126	6,150	-10.43 150.13	6,587	5,814
Total bride side	127.80 231.72	16,243	2,729	243.19 292.32	23,092	6,252	237.71 545.46	24,739	5,896
Total groom side	672.96 1,093.90	40,709	2,811	1,386.57* 794.13	60,508	6,363	1,035.14 1,446.44	66,374	5,995

Note: This table reports estimates from 28 separate regressions of the effect of unemployment rate at entry of various marriage outcomes. The columns marked men (women) use data reported by men (women) pertaining to their marriages and the unemployment rate in the individual's governorate of birth, when the individual was aged 18. To avoid censoring bias, I restrict to individuals observed at an age greater than or equal to the 95th percentile of age at first marriage (33 for men, 29 for women). Husband-wife age gap is the difference between the husband and the wife's age (so older husband is a positive number). The following row is the absolute value of this quantity. Couple lives alone is an indicator equal to one if the couple lives in a home or apartment without other family in the house. Consanguineous is an indicator equal to one if the couple are related by blood (ie, cousins). Mahr is obligatory under Islam and refers to the payment made from the husband to the wife at the time of marriage. Mahr Mukahr is the deferred mahr owed to the wife in the event of divorce (except wife-initiated no-fault divorce) or her husband's death. Ayma refers to a non-religious but culturally customary contract detailing the goods/wealth brought into the marriage by the wife which may be reclaimed by the wife under certain circumstances. Jewelry refers to the value of gold jewelry gifted to the bride by the groom's and his family. Furniture refer to the value of home furnishings purchased for the marital home. Gihaz refers to other preparations for the home (ie, linens, kitchenware, etc). Housing refers to cost of housing incurred for the couple's marital home, and wedding is the cost of the wedding itself. Total bride side sums the amount the bride or her family spent on the furniture, gihaz, housing and wedding. Total groom side sums their groom's side's share of the same expenses, plus the mahr and jewelry. (Mahr muakhr and ayma are not included in this sum since they are only due upon divorce/death.)

Table 2.10: Marital Decisionmaking and Women's Mobility

	Men	Women			
	Decision	Decision	Mobility	Decision	Mobility
Gov. of Birth, age 18 UR	0.00348 (0.00513)	0.00270 (0.00301)	-0.00313 (0.00451)		
Gov. of Birth, age 12 UR				0.0106** (0.00487)	-0.00167 (0.00455)
Cohort	Yes	Yes	Yes	Yes	Yes
Gov of birth	Yes	Yes	Yes	Yes	Yes
Clustering	GB	GB	GB	GB	GB
Obs	5234	12642	12624	11046	11030

Note: This table reports the effect of entry conditions (or entry conditions of women's likely marriage cohort) on marital decisionmaking and women's mobility. The decision columns report impacts on an index of marital decisionmaking, where an increase corresponds to an increase in joint or wife-only decisionmaking.

Chapter 3

Housing Subsidies for Refugees: Experimental Evidence on Life Outcomes and Social Integration in Jordan

This chapter is coauthored with Abdulrazzak Tamim, Emma Smith, Edward Miguel, Samuel Leone, Sandra V. Rozo and Sarah Stillman.

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, Hsiang, and Edward Miguel 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, Hendren, and Katz 2016; Kling, Liebman, and Katz 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 humanitarian aid organizations in Jordan, randomized rental subsidies across 2,870 refugee households residing in Irbid and Mafrqa 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, Hendren, and Katz 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

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

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, Nimoh, and Matthew (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, Miano, and Stantcheva 2023, Dustmann, Vasiljeva, and Damm 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 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

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

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

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 2017; DellaVigna, Otis, and Vivaldi 2020), the research team gathered forecasts from both researchers and policy experts regarding the primary impacts of the 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 re-

port 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, Reisinger, and Shapiro 2015, Baird, De Hoop, and Özler 2013, Egger et al. 2022, M. C. Lehmann and D. T. 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

3.2.1 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 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.

3.2.2 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, Ulrichs, and Holmes 2018; C. Lehmann and D. 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 focused on Syrian refugees in the country’s northern governorates of Irbid and Mafraq, 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

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

(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.

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

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/>

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

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.

programs is the 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.

3.3.1 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

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.

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 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.

3.3.2 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.

3.3.3 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 3.2). In the midline survey, 1,619 participants were surveyed by phone

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.

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

Figure 3.2 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 3.3, 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.

3.3.4 Descriptive statistics

We present characteristics of the sample using baseline vulnerability assessment in Table 3.1.¹⁵ 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.

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

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.

Table 3.1 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 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.

3.3.5 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.

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 C.8).

3.4 Empirical Strategy

3.4.1 First stage compliance with program assignment

Table 3.3, 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.

3.4.2 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 Mafraq 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

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

midline, endline, 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 2012), with a q-value of 0.10 or less.¹⁹

3.4.3 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 3.3, the gaps in attrition are statistically significant, and therefore we estimate bounds following Lee (2009) and report these in appendix C.1. 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.

3.4.4 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

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.

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.

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.

3.4.5 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.

3.5.1 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,

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.

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 ($q\text{-value}=0.07$). 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 $q\text{-value}$ 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

Beyond the primary pre-specified outcomes, all other outcomes with statistically significant estimated program effects pooled across rounds that survive a multiple hy-

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

pothesis 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 C.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.

3.5.2 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.

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 ($q\text{-value} = 0.01$) 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 ($q\text{-value} = 0.02$), 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, $q\text{-value}=0.03$), and this effect is driven by an increase of 0.16 more adolescent boys ($q\text{-value}=0.33$). 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 C.2), and as such should be viewed as suggestive.

Treated households also report experiencing increased COVID-19 exposure dur-

ing 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 C.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 C.2).

3.5.3 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 -value = 0.01 and q -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 C.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 C.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 -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 -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 C.7).

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.

3.5.4 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 -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 respondent 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 3.2). 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 -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.

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.

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 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

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.

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 ($q\text{-value} < 0.08$), 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 C.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 C.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

²⁷This finding is in contrast to the correlational evidence in Ghosn, Braithwaite, and Chu, 2019, who show that a history of displacement did not improve openness towards refugees.

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, Imbens, and Rubin, 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 Table C.9). 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.

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 Figure C.1): 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 Table C.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 Table C.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 C.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 midline 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.

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

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

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 manipulated 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 (Alrababaâ 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, A. M. Ibáñez, and Rozo 2021; Rozo, Quintana, and Urbina 2023; A. 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 Tables and Figures

Figure 3.1: Geographic Location of Treated and Control Communities

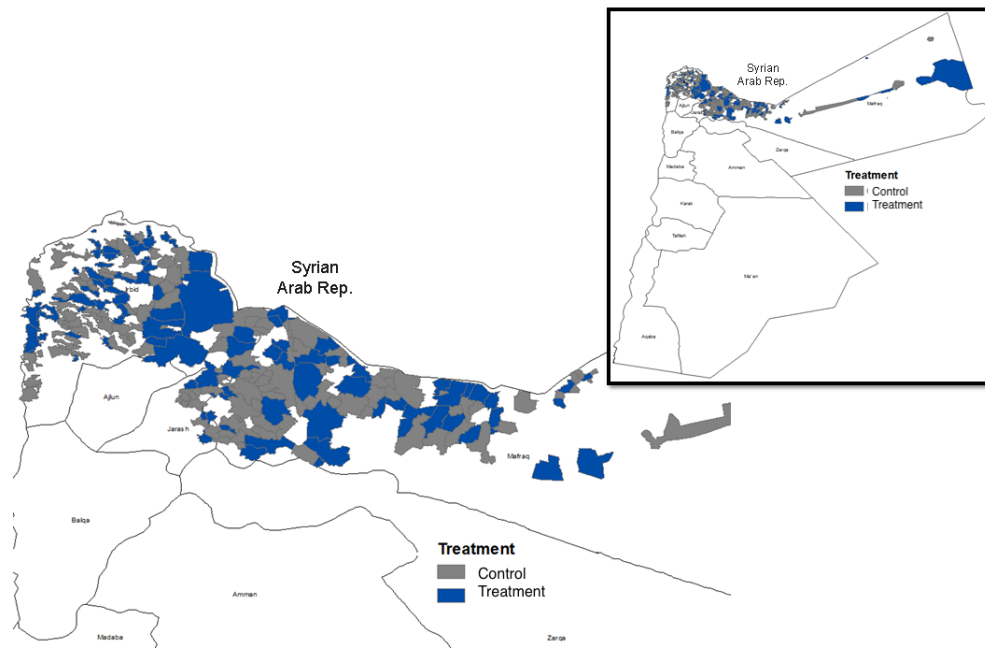


Figure 3.2: RCT, Data Collection, and Forecasts Timeline

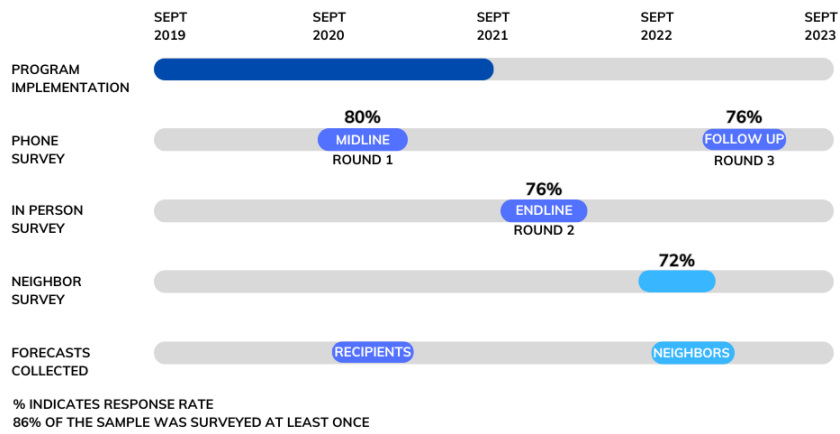


Table 3.1: Average Respondent, Household, and Shelter Characteristics

		Treatment	Control	Diff	(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 3.3). 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	Treatmnt	(se)	FDR q-vals	Contrl 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 (USD PPP)	-84.71***	(33.10)	[0.07]	228.80	(208.80)	1,421	2
Food Consumption (Log USD PPP)	-0.02	(0.07)	[0.63]	4.41	(0.72)	2,753	2, 3
Log Total Consumption (Log USD 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 Child 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-vals	Contrl 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-vals	Contrl mean	(sd)	N
Panel A: Endline						
Focus 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

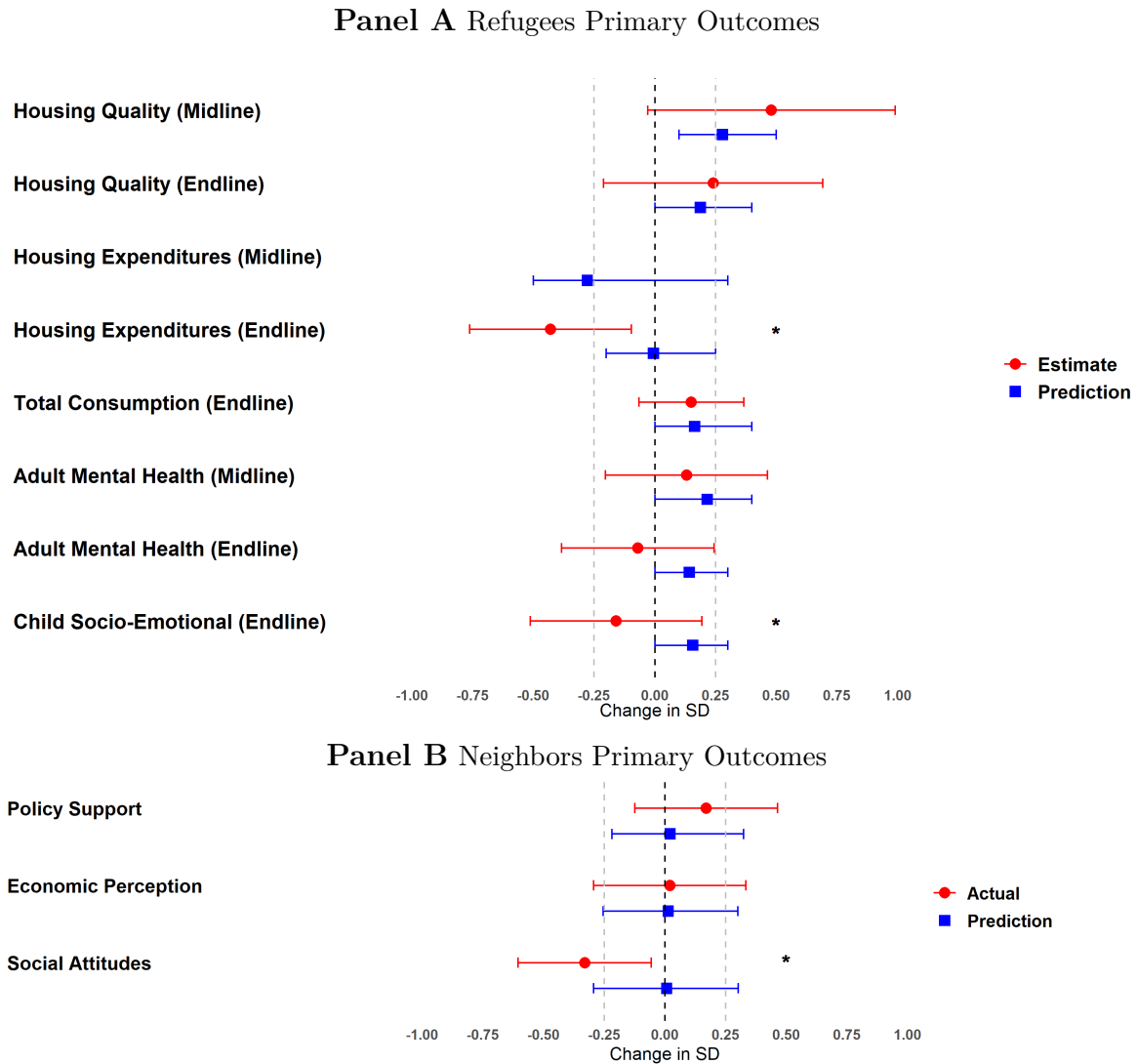
Outcome	TOT	(se)	FDR q-val	Contrl mean	(sd)	N
Panel A: Primary Outcomes						
Social attitudes & perceptions (SD)	-0.33**	(0.14)	[0.08]	-0.00	(1.00)	1,102
Economic attitudes & perceptions (SD)	0.02	(0.16)	[0.56]	0.02	(1.02)	1,102
Policy preferences (SD)	0.17	(0.15)	[0.37]	-0.02	(1.00)	1,102
Altruism 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 is 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 using an outcome family comprised of all the components of the social attitudes and perceptions index.

Figure 3.3: Program Impacts on Primary Outcomes - Estimates vs. 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 $p < 0.05$.

Bibliography

- Abdullah, Mohammed (2022). “هل سيغير غلاء الأسعار في عادات الزواج لدى المصريين؟” *Al Jazeera*. URL: <https://aja.me/9h7f9w> (visited on 01/22/2025).
- Abou Daher, Chaza et al. (2025). “Drivers of Change: Employment Responses to the Lifting of the Saudi Female Driving Ban.” *American Economic Review* 115.9, pp. 3248–3271. DOI: 10.1257/aer.20240119. URL: <https://doi.org/10.1257/aer.20240119>.
- Abramitzky, Ran and Leah Boustan (2024). “Immigrants and their children assimilate into US society and the US economy, both in the past and today.” *PNAS Nexus*.
- Agness, Daniel (2023). “Housing and Human Capital: Condominiums in Ethiopia.” *3ie Registry*.
- Alesina, Alberto, Armando Miano, and Stefanie Stantcheva (2023). “Immigration and Redistribution.” *The Review of Economic Studies*.
- Alrababaâ h, Alaâ et al. (2021). “Attitudes toward migrants in a highly impacted economy: Evidence from the Syrian refugee crisis in Jordan.” *Comparative Political Studies* 54.1, pp. 33–76.
- Altonji, Joseph G., Lisa B. Kahn, and Jamin D. Speer (2016). “Cashier or Consultant? Entry Labor Market Conditions, Field of Study, and Career Success.” *Journal of Labor Economics* 34.S1, S361–S401. DOI: 10.1086/682938. eprint: <https://doi.org/10.1086/682938>. URL: <https://doi.org/10.1086/682938>.
- Anderson, Michael L (2012). “Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects.” *Journal of American Statistical Association*.
- Angrist, Joshua D, Guido W Imbens, and Donald B Rubin (1996). “Identification of causal effects using instrumental variables.” *Journal of the American statistical Association* 91.434, pp. 444–455.
- Arab Barometer (2022). *Arab Barometer Wave VII, 2021–2022*. Accessed: 2025-09-27. URL: <https://www.arabbarometer.org/surveys/arab-barometer-wave-vii/>.
- Arab News (Mar. 2022). “Saudi Arabia Hits Female Labor Force Target Almost 10 Years Early: Uber Report.” *Arab News*. Accessed: 2025-10-10. URL: <https://www.arabnews.com/node/2039271/business-economy>.
- Arellano-Bover, Jaime (2024). “Career Consequences of Firm Heterogeneity for Young Workers: First Job and Firm Size.” *Journal of Labor Economics* 42.2, pp. 549–589. DOI: 10.1086/723500. URL: <https://doi.org/10.1086/723500>.

- Arraf, Jane (2016). "In Egypt, The High Cost Of Romance Is Crippling Hopes Of Marriage." *National Public Radio*. URL: <https://www.npr.org/sections/parallels/2016/10/15/497957693/in-egypt-the-high-cost-of-romance-is-crippling-hopes-of-marriage> (visited on 01/22/2025).
- Ash, Nazanin and Cindy Huang (2018). "Using the Compact Model to Support Host States and Refugee Self-reliance." *World Refugee Council Research Paper Series*.
- Assaad, Ragui, Caroline Krafft, and Irene Selwaness (2022). "The Impact of Marriage on Women's Employment in the Middle East and North Africa." *Feminist Economics* 28.2, pp. 247–279. DOI: 10.1080/13545701.2021.2007415. eprint: <https://doi.org/10.1080/13545701.2021.2007415>. URL: <https://doi.org/10.1080/13545701.2021.2007415>.
- Aygün, Aysun Hızıroğlu et al. (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, p. 103266.
- Bahar, Dany, Ana María Ibáñez, and Sandra 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, p. 102652.
- Baird, Sarah, Jacobus De Hoop, and Berk Özler (2013). "Income shocks and adolescent mental health." *Journal of Human Resources* 48.2, pp. 370–403.
- Baird, Sarah, Craig McIntosh, and Berk Özler (2013). "Cash or condition? Evidence from a cash transfer experiment." *American Economic Journal: Applied Economics* 5.2, pp. 1–26. DOI: 10.1257/app.5.2.1.
- Barron, Kai et al. (2023). "Discrimination, Narratives, and Family History: An Experiment with Jordanian Host and Syrian Refugee Children." *The Review of Economics and Statistics* 105.4, pp. 1008–1016. DOI: 10.1162/rest_a_01090. URL: https://doi.org/10.1162/rest_a_01090.
- Baseler, Travis et al. (2023). "Can Redistribution Change Policy Views?: Aid and Attitudes Toward Refugees in Uganda." *Center for Global Development, Working Paper* 645.
- Beltramo, Theresa, Florence Nimoh, and O'Brien Matthew (2024). "Financial Security, Climate Shocks and Social Cohesion." *Center for Economic Policy Research, Working Paper*.
- Brekke, Kjell Arne and Karine Nyborg (2008). "Attracting responsible employees: Green production as labor market screening." *Resource and Energy Economics* 30.4, pp. 509–526. ISSN: 0928-7655. DOI: <https://doi.org/10.1016/j.reseneeco.2008.05.001>. URL: <https://www.sciencedirect.com/science/article/pii/S0928765508000146>.
- Burke, Marshall, Solomon M Hsiang, and Edward Edward Miguel (2015). "Global non-linear effect of temperature on economic production." *Nature*.
- Bursztyn, Leonardo, Alessandra L. González, and David Yanagizawa-Drott (Oct. 2020). "Misperceived Social Norms: Women Working Outside the Home in Saudi Arabia." *American Economic Review* 110.10, pp. 2997–3029. DOI: 10.1257/aer.20180975. URL: <https://www.aeaweb.org/articles?id=10.1257/aer.20180975>.

- Bustelo, Monserrat et al. (2023). “What Is the Price of Freedom? Estimating Women’s Willingness to Pay for Job Schedule Flexibility.” *Economic Development and Cultural Change* 71.4, pp. 1179–1211. DOI: 10.1086/718645. eprint: <https://doi.org/10.1086/718645>. URL: <https://doi.org/10.1086/718645>.
- Caria, Stefano et al. (Feb. 2025). *The barriers to female employment: Experimental evidence from Egypt*. Working Paper. URL: <https://g2lm-lic.iza.org/wp-content/uploads/2025/03/glmlic-wp092.pdf>.
- Chetty, Raj, Nathaniel Hendren, and Lawrence F. Katz (Apr. 2016). “The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment.” *American Economic Review* 106.4, pp. 855–902. DOI: 10.1257/aer.20150572. URL: <https://www.aeaweb.org/articles?id=10.1257/aer.20150572>.
- Christensen, Peter, Gustavo Nino, and Adam Osman (Sept. 2025). *The Demand for Mobility: Evidence from an Experiment with Uber Riders*. Working Paper. URL: https://www.adam-osman.com/wp-content/uploads/2024/10/Demand_for_Mobility_Christensen_Osman-Sep2025.pdf.
- Clark, Shelley et al. (2019). “The Impact of Childcare on Poor Urban Women’s Economic Empowerment in Africa.” *Demography* 56.4, pp. 1247–1272. DOI: 10.1007/s13524-019-00793-3.
- Corradini, Viola and Giulia Buccione (2023). “Unilateral divorce rights, domestic violence and women’s agency: Evidence from the Egyptian Khul reform.” *Journal of Development Economics* 160.
- Crowne, D. P. and D. Marlowe (1960). “A new scale of social desirability independent of psychopathology.” *Journal of Consulting Psychology* 24.4.
- Cruces, Guillermo, Andrés Ham, and Mariana Viollaz (2012). *Scarring effects of youth unemployment and informality: Evidence from Argentina and Brazil*. Working Paper.
- DellaVigna, Stefano, Nicholas Otis, and Eva Vivalt (2020). “Forecasting the Results of Experiments: Piloting an Elicitation Strategy.” *National Bureau of Economic Research*.
- DellaVigna, Stefano and Devin) Pope (2017). “What Motivates Effort? Evidence and Expert Forecasts.” *The Review of Economic Studies*.
- Dhillon, Navtej and Tarik Yousef (2009). *Generation in Waiting: The Unfulfilled Promise of Young People in the Middle East*. Brookings Institution Press. URL: <http://www.jstor.org/stable/10.7864/j.ctt6wpgg5> (visited on 05/09/2023).
- Dube, Arindrajit, Suresh Naidu, and Adam D Reich (Sept. 2022). *Power and Dignity in the Low-Wage Labor Market: Theory and Evidence from Wal-Mart Workers*. Working Paper 30441. National Bureau of Economic Research. DOI: 10.3386/w30441. URL: <http://www.nber.org/papers/w30441>.
- Dustmann, Christian, Kristine Vasiljeva, and Anna Damm (2019). “Refugee Migration and Electoral Outcomes.” *The Review of Economic Studies*.
- Egger, Dennis et al. (2022). “General equilibrium effects of cash transfers: experimental evidence from Kenya.” *Econometrica* 90.6, pp. 2603–2643.

- Egyptian Social Fund for Development (1999). *The Egyptian Social Fund For Development. A case study of labor-based contractor training in Egypt*. Seventh regional seminar for labour-based practitioners, Lusaka, Zambia.
- Emanuel, Natalia and Emma Harrington (Oct. 2024). "Working Remotely? Selection, Treatment, and the Market for Remote Work." *American Economic Journal: Applied Economics* 16.4, pp. 528–59. DOI: 10.1257/app.20230376. URL: <https://www.aeaweb.org/articles?id=10.1257/app.20230376>.
- Feld, Brian, AbdelRahman Nagy, and Adam Osman (2022). "What do jobseekers want? Comparing methods to estimate reservation wages and the value of job attributes." *Journal of Development Economics* 159, p. 102978. ISSN: 0304-3878. DOI: <https://doi.org/10.1016/j.jdeveco.2022.102978>. URL: <https://www.sciencedirect.com/science/article/pii/S0304387822001201>.
- Garlick, Robert, Erica Field, and Kate Vyborny (Apr. 2025). *Women's Mobility and Labor Supply: Experimental Evidence from Pakistan*. Working Paper 17883. URL: <https://www.jstor.org/stable/pdf/resrep70710.pdf?acceptTC=true&coverage=false&addFooter=false>.
- Ghosn, Faten, Alex Braithwaite, and Tiffany S Chu (2019). "Violence, displacement, contact, and attitudes toward hosting refugees." *Journal of Peace Research* 56.1, pp. 118–133.
- Groh, Matthew, Nandini Krishnan, et al. (2016). "Do Wage Subsidies Provide a Stepping-Stone to Employment For Recent College Graduates? Evidence from a Randomized Experiment In Jordan." *The Review of Economics and Statistics* 98.3, pp. 488–502. ISSN: 00346535, 15309142. URL: <http://www.jstor.org/stable/24917030> (visited on 10/01/2025).
- Groh, Matthew, David McKenzie, et al. (2015). "Testing the importance of search frictions and matching through a randomized experiment in Jordan." *IZA Journal of Labor Economics* 4.7.
- Hagen-Zanker, Jessica, Martina Ulrichs, and Rebecca 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, pp. 57–77. DOI: <https://doi.org/10.1111/issr.12166>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/issr.12166>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/issr.12166>.
- Haushofer, Johannes, James Reisinger, and Jeremy Shapiro (2015). "Your gain is my pain: Negative psychological externalities of cash transfers." *Working Paper*, retrieved on May 13, p. 2016.
- Haushofer, Johannes and Jeremy Shapiro (2016). "The short-term impact of unconditional cash transfers to the poor: Experimental evidence from Kenya." *Quarterly Journal of Economics* 131.4, pp. 1973–2042. DOI: 10.1093/qje/qjw025.
- (2018). "The long-term impact of unconditional cash transfers: experimental evidence from Kenya." *Busara Center for Behavioral Economics Working Paper, Nairobi, Kenya*.

- Hidrobo, Melissa et al. (2014). “Cash, food, or vouchers? Evidence from a randomized experiment in northern Ecuador.” *Journal of Development Economics* 107, pp. 144–156.
- Huckfeldt, Christopher (2022). “Understanding the Scarring Effect of Recessions.” *American Economic Review* 112.4, pp. 1273–1310.
- Hudson, Valerie M. and Hilary Matfess (2017). “In Plain Sight: The Neglected Linkage between Brideprice and Violent Conflict.” *International Security* 42.1, pp. 7–40.
- Hussein, Mohamed H. and Ahmed M. Abdel-Khalek (2021). “Developing a Revised Version of the Arabic Big Five Personality Inventory Using Item Response Theory.” *Mankind Quarterly* 62.2, pp. 327–343. URL: <https://doi.org/10.46469/mq.2021.62.2.7>.
- Ibáñez, Ana et al. (2024). “Life out of the Shadows: The Impacts of Regularization Program on the Lives of Forced Migrants.” *Journal of European Economic Association*.
- International Labour Organization (2025). *Labor force participation rate, female (% of female population ages 15+) (modeled ILO estimate)*. <https://data.worldbank.org/indicator/SL.TLF.CACT.FE.ZS>. ILO Modelled Estimates and Projections database (ILOEST), Indicator Code: SL.TLF.CACT.FE.ZS. Accessed January 7, 2025.
- Jack, B. Kelsey, Kathryn McDermott, and Anja Sautmann (2022). “Multiple Price Lists for Willingness to Pay Elicitation.” *Journal of Development Economics* 159, p. 102977. DOI: 10.1016/j.jdeveco.2022.102977. URL: <https://doi.org/10.1016/j.jdeveco.2022.102977>.
- Jayachandran, Seema (June 2020). *Social Norms as a Barrier to Women’s Employment in Developing Countries*. Working Paper 27449. National Bureau of Economic Research. DOI: 10.3386/w27449. URL: <http://www.nber.org/papers/w27449>.
- Kahn, Lisa B. (2010). “The long-term labor market consequences of graduating from college in a bad economy.” *Labour Economics* 17.2, pp. 303–316. ISSN: 0927-5371. DOI: <https://doi.org/10.1016/j.labeco.2009.09.002>. URL: <https://www.sciencedirect.com/science/article/pii/S0927537109001018>.
- Kaur, Supreet et al. (Nov. 2024). “Do Financial Concerns Make Workers Less Productive?” *The Quarterly Journal of Economics* 140.1, pp. 635–689. ISSN: 0033-5533. DOI: 10.1093/qje/qjae038. eprint: <https://academic.oup.com/qje/article-pdf/140/1/635/61227107/qjae038.pdf>. URL: <https://doi.org/10.1093/qje/qjae038>.
- Kholaif, Dahlia and Tamer El-Ghobashy (2016). *Egypt’s Money Woes Hit a Touchstone of Marriage*. URL: <https://www.wsj.com/articles/egypts-money-woes-hit-a-touchstone-of-marriage-1473290993> (visited on 01/22/2025).
- El-Kholy, Heba Aziz (2000). “A Tale of Two Contracts: Towards a Situated Understanding of “Women Interests” in Egypt.” *Situating Globalization*. URL: <https://api.semanticscholar.org/CorpusID:156430382>.
- Kline, Patrick (2025). “Chapter 8 - Labor market monopsony: fundamentals and frontiers.” *Handbook of Labor Economics*. Ed. by Christian Dustmann and Thomas

- Lemieux. Vol. 6. Handbook of Labor Economics. Elsevier, pp. 655–728. DOI: <https://doi.org/10.1016/bs.heslab.2025.07.007>. URL: <https://www.sciencedirect.com/science/article/pii/S1573446325000070>.
- Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz (2007). “Experimental Analysis of Neighborhood Effects.” *Econometrica* 75.1, pp. 83–119.
- Krafft, Caroline, Ragui Assad, and Khandker Wahedur Rahman (2019). *Introducing the Egyptian Labor Market Panel Survey 2018*. Economic Research Forum Working Paper 1360.
- Kuchibhotla, Murali, Peter F. Orazem, and Sanjana Ravi (2020). “The scarring effects of youth joblessness in Sri Lanka.” *Review of Development Economics* 24, pp. 269–287.
- Kumar, Tanu (2021). “Home-price subsidies increase local-level political participation in urban India.” *Forthcoming, Journal of Politics*.
- Lee, David S. (2009). “Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects.” *The Review of Economic Studies* 76.3, pp. 1071–1102. ISSN: 0034-6527. DOI: 10.1111/j.1467-937X.2009.00536.x.
- Lehmann, Christian and Daniel Masterson (2014). *Emergency Economies: the Impact of Cash Assistance in Lebanon*. Tech. rep. International Rescue Committee.
- Lehmann, M Christian and Daniel TR Masterson (2020). “Does aid reduce anti-refugee violence? Evidence from Syrian refugees in Lebanon.” *American Political Science Review* 114.4, pp. 1335–1342.
- Liu, Kai, Kjell G Salvanes, and Erik Sorensen (2016). “Good skills in bad times: Cyclical skill mismatch and the long-term effects of graduating in a recession.” *European Economic Review* 84.C, pp. 3–17. DOI: 10.1016/j.euroecorev.2015.08.015. URL: <https://ideas.repec.org/a/eee/eecrev/v84y2016icp3-17.html>.
- Ludwig, Jens et al. (May 2013). “Long-Term Neighborhood Effects on Low-Income Families: Evidence from Moving to Opportunity.” *American Economic Review* 103.3, pp. 226–31. DOI: 10.1257/aer.103.3.226. URL: <https://www.aeaweb.org/articles?id=10.1257/aer.103.3.226>.
- Mahmud, Minhaj et al. (Jan. 2020). “What Aspects of Formality Do Workers Value? Evidence from a Choice Experiment in Bangladesh.” *The World Bank Economic Review* 35.2, pp. 303–327. ISSN: 0258-6770. DOI: 10.1093/wber/lhz046. eprint: <https://academic.oup.com/wber/article-pdf/35/2/303/37750536/lhz046.pdf>. URL: <https://doi.org/10.1093/wber/lhz046>.
- Majid, Nomaan (2012). *How not to count the employed in developing countries*. ILO Employment Working Paper No. 136.
- Martinez, Claudia and Marcela Peticara (2017). “Childcare effects on maternal employment: Evidence from Chile.” *Journal of Development Economics* 126, pp. 127–137. ISSN: 0304-3878. DOI: <https://doi.org/10.1016/j.jdeveco.2017.01.001>. URL: <https://www.sciencedirect.com/science/article/pii/S0304387817300019>.

- Martínez, Daniel et al. (2020). “Connecting to Economic Opportunity: the Role of Public Transport in Promoting Women’s Employment in Lima.” *Journal of Economics, Race, and Policy* 3. URL: <https://doi.org/10.1007/s41996-019-00039-9>.
- Mas, Alexandre (2025). “Chapter 5 - Non-wage amenities.” *Handbook of Labor Economics*. Ed. by Christian Dustmann and Thomas Lemieux. Vol. 6. Handbook of Labor Economics. Elsevier, pp. 373–446. DOI: <https://doi.org/10.1016/bs.heslab.2025.07.002>. URL: <https://www.sciencedirect.com/science/article/pii/S1573446325000021>.
- Mas, Alexandre and Amanda Pallais (2017). “Valuing Alternative Work Arrangements.” *American Economic Review* 107.12, pp. 3722–3759. DOI: 10.1257/aer.20161500.
- McKelway, Madeline (Feb. 2025). *How Does Women’s Employment Affect Their Time Use? Evidence from a Randomized Encouragement Design in India*. Tech. rep. Working paper. URL: https://drive.google.com/file/d/1G_ZI0ug2Zzy47YYXiWhXyJ_WX1ZkzBZx/view.
- Metreau, Eric, Kathryn Elizabeth Young, and Shwetha Grace Eapen (July 2024). *World Bank Country Classifications by Income Level for 2024–2025*. <https://blogs.worldbank.org/en/opendata/world-bank-country-classifications-by-income-level-for-2024-2025>. World Bank Blogs.
- Moussa, Wael et al. (2022). “The impact of cash transfers on Syrian refugee children in Lebanon.” *World Development* 150, p. 105711.
- Naidoo, Darian, Truman Packard, and Ilmiawan Auwalin (2015). *Mobility, Scarring and Job Quality in Indonesia’s Labor Market*. World Bank Policy Research Working Paper.
- Nekoei, Arash (2024). *Will markets provide humane jobs? A hypothesis*. Working Paper. URL: <http://dx.doi.org/10.2139/ssrn.3893099>.
- NRC (2015). *In search of a home: Access to adequate housing in Jordan*. Tech. rep. Norwegian Refugee Council.
- OAMDI (2018). *Labor Market Panel Surveys (LMPS)*. Version 1.1 of Licensed Data Files; JLMPS 2016. Egypt: Economic Research Forum (ERF). URL: <http://erf.org.eg/data-portal/>.
- Oreopoulos, Philip, Till von Wachter, and Andrew Heisz (Jan. 2012). “The Short- and Long-Term Career Effects of Graduating in a Recession.” *American Economic Journal: Applied Economics* 4.1, pp. 1–29. DOI: 10.1257/app.4.1.1. URL: <https://www.aeaweb.org/articles?id=10.1257/app.4.1.1>.
- Özler, Berk et al. (2021). “Children on the move: Progressive redistribution of humanitarian cash transfers among refugees.” *Journal of Development Economics* 153, p. 102733.
- Park, Seong-Hi and Hye Yon 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, p. 114037.
- Quattrochi, John et al. (2022). “Cash-like vouchers improve psychological well-being of vulnerable and displaced persons fleeing armed conflict.” *PNAS Nexus* 1.3, pgac101.

- Ramadan, Ahmed (2020). “مبادرات تيسير الزواج.. هل تكون حلا لمواجهة العنوسة في مصر؟” *Al Jazeera*. URL: <https://aja.me/6c1rk5> (visited on 01/22/2025).
- Rothstein, Jesse (2023). *The Lost Generation?* Tech. rep. 5, pp. 1452–1479. DOI: 10.3368/jhr.58.5.0920-11206R1. eprint: <https://jhr.uwpress.org/content/58/5/1452.full.pdf>. URL: <https://jhr.uwpress.org/content/58/5/1452>.
- Rozo, Sandra V, Alejandra Quintana, and María José Urbina (2023). *The Electoral Consequences of Easing the Integration of Forced Migrants*.
- Rozo, Sandra V and Micaela Sviatschi (2021). “Is a refugee crisis a housing crisis? Only if housing supply is unresponsive.” *Journal of Development Economics* 148, p. 102563.
- Salem, Rania (2018). “Matrimonial Expenditures and Egyptian Women’s Power Within Marriage.” *Journal of Family Issues* 39.9, pp. 2615–2638.
- Schwandt, Hannes and Till M von Wachter (Jan. 2020). *Socio-Economic Decline and Death: The Life-Cycle Impacts of Recessions for Labor Market Entrants*. Working Paper 26638. National Bureau of Economic Research.
- Semlali, Amina and Diego F. Angel-Urdinola (2012). *Egypt: Public Employment Services and Publicly Provided Active Labor Market Programs*. World Bank Policy Note 69408.
- Singerman, Diane (2007). *The Economic Imperatives of Marriage: Emerging Practices and Identities among Youth in the Middle East*. The Middle East Youth Initiative Working Paper.
- Stensrud, Mats J. and Miguel A. Hernán (Apr. 2020). “Why Test for Proportional Hazards?” *JAMA* 323.14, pp. 1401–1402. ISSN: 0098-7484. DOI: 10.1001/jama.2020.1267. eprint: https://jamanetwork.com/journals/jama/articlepdf/2763185/jama_stensrud_2020_gm_200003.pdf. URL: <https://doi.org/10.1001/jama.2020.1267>.
- Tauchmann, Harald (2014). “Lee (2009) treatment-effect bounds for nonrandom sample selection.” *The Stata Journal* 14.4, pp. 884–894.
- UNHCR (2023a). *Jordan*. Tech. rep. United Nations High Commissioner for Refugees.
- (2023b). *Syria*. Tech. rep. United Nations High Commissioner for Refugees.
- (2023c). *UNHCR Global Trends*. Tech. rep. United Nations High Commissioner for Refugees.
- Van den Berge, Wiljan (2018). “Bad start, bad match? The early career effects of graduating in a recession for vocational and academic graduates.” *Labour Economics* 53. European Association of Labour Economists 29th annual conference, St.Gallen, Switzerland, 21-23 September 2017, pp. 75–96. ISSN: 0927-5371. DOI: <https://doi.org/10.1016/j.labeco.2018.05.011>. URL: <https://www.sciencedirect.com/science/article/pii/S0927537118300563>.
- Woerner, Wolfgang et al. (2004). “The Strengths and Difficulties Questionnaire overseas: evaluations and applications of the SDQ beyond Europe.” *European child & adolescent psychiatry* 13, pp. ii47–ii54.

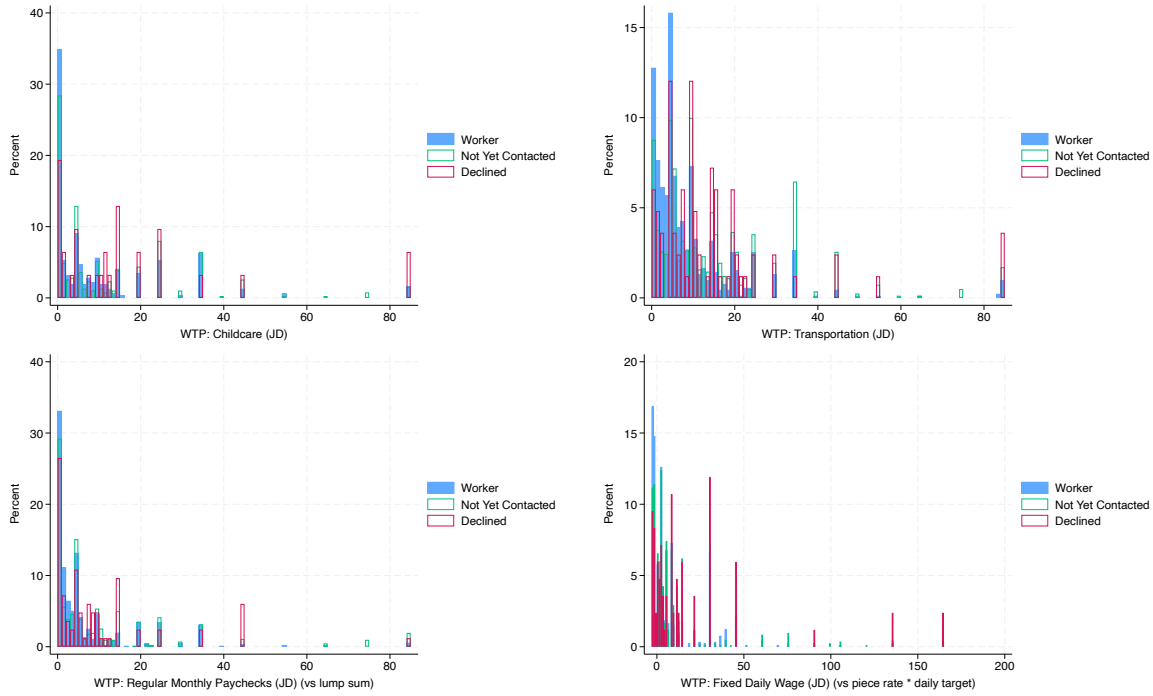
- World Bank (2023). *Who Cares? In Iraq, Jordan And Lebanon*. Tech. rep. URL: <https://documents1.worldbank.org/curated/en/099000502222338765/pdf/P16815701c336d00e095dc093ef29af40a7.pdf>.
- (2024). *Childcare in Jordan: Workplace-Based Nurseries in Private and Public Institutions*. Brief. World Bank. URL: <https://documents1.worldbank.org/curated/en/099062524142032013/pdf/P168157-70e9fe0e-aee7-4e11-b698-69efaf16bd5b.pdf>.
- (2025a). *Ratio of female to male labor force participation rate (%) (modeled ILO estimate) [Indicator Code: SL.TLF.CACT.FM.ZS]*. World Development Indicators database, June 3, 2025. URL: <https://data.worldbank.org/indicator/SL.TLF.CACT.FM.ZS>.
- (2025b). *Unemployment, total (% of total labor force) (national estimate) — Jordan (Indicator SL.UEM.TOTL.NE.ZS)*. <https://data.worldbank.org/indicator/SL.UEM.TOTL.NE.ZS?locations=JO>. Online indicator / dataset. Accessed September 25, 2025.
- Zhou, Yang-Yang (2019). “How refugee resentment shapes national identity and citizen participation in Africa.” *APSA 2018 Annual Meeting Paper*.

Appendix A

Appendix to Job Amenities, Adverse Selection, and Women's Labor Supply in Jordan

A.1 Appendix Figures and Tables

Figure A.1: Distribution of Uncensored WTP



Note: This figure includes histograms of the uncensored WTP that Jordanian women in the sample assign to the four job amenities, reportedly separately by whether the individual was a worker, had been offered a job but declined, or had applied but the firm had not yet contacted them. Individuals with WTP at the top of the price list were asked an open-ended question about the lowest job B wage that would induce them to prefer the higher wage to the amenity. This figure reports WTP including those uncensored estimates. Because WTP above the top of the price list was elicited through an open-ended question, the resulting values should be interpreted as suggestive.

Table A.1: Main Results with Selection-Correction Weights

Productivity Measures (z_{ij})				
Transport	-0.00688*** (0.00227)			
Monthly Pay		-0.00595** (0.00273)		
Fixed Wage			-0.00492* (0.00268)	
Childcare				-0.00267 (0.00459)
Weights	Yes	Yes	Yes	Yes
Obs	5519	5529	5539	1768

Note: This table reports results from a regression of productivity (SDs) on WTP for the amenity in dinars: $z_{ij} = WTP_i + \epsilon_{ij}$. The data is stacked so that the left hand side variable includes all measures of productivity observed for an individual, and the right hand side is WTP. Standard errors are clustered at the individual level. Column 1 reports that an individual with WTP 1 dinar higher is expected to be 0.00688 SDs less productive. Scaled by the standard deviation of WTP for transportation, this implies that an individual with a 1SD higher WTP for transportation will be on average 0.05 SDs less productive.

Table A.2: Heterogeneity by Fixed vs Piece Rate Pay

	Productivity Measures (z_{ij})
WTP	-0.0131*** (0.00456)
Piece Rate	0.0418 (0.0496)
Piece Rate X WTP	0.00864* (0.00496)
Fixed Effects	Location Education Nth Project Industry Experience
Obs	5522

Note: This table provides suggestive evidence that individuals who value transportation and monthly paychecks more are more responsive to external motivation. All workers are more productive when paid piece rates, but workers with high WTP for transport or regular monthly pay are even more responsive to piece rates than the average worker. The table reports estimates from a regression of productivity on WTP for transportation or monthly paychecks interacted with whether the project was paid piece rates or daily wages. WTP for transportation and monthly paychecks are stacked – these amenities are included because these are the amenities which display evidence of inducing adverse selection. Only productivity measures for which information is available on whether the project paid piece rate or fixed wages are included.

A.2 Willingness-to-pay Elicitation Appendix

Jack, McDermott, and Sautmann (2022) describes various pitfalls that may arise during a multiple price list elicitation of WTP that introduce bias. Care was taken in the design of the multiple price list to avoid these sources of bias while remaining practical for elicitation over the phone.

Order Effects. To avoid order effects, both the order of sections eliciting WTP for the 4 amenities and whether each price list was ascending or descending were randomized.

Adaptive Bias. Adaptive surveys may introduce bias when combined with respondents who make mistakes or change their mind halfway through the elicitation. For example, if the respondent accidentally misreports that she prefers job B in a choice between A paying 15JD + transport and B paying 20JD, an adaptive survey will anchor on this amount and narrow in on more precise WTP around this incorrect (too small) value. To avoid this problem while remaining minimally tedious, the survey first asked 4 questions, increasing or decreasing the pay of job B in increments of 5 dinar with no adaptiveness. For example:

If job A pays 15 JD/day and includes childcare, and job B pays **15** JD/day without childcare, which would you prefer?

If job A pays 15 JD/day and includes childcare, and job B pays **20** JD/day without childcare, which would you prefer?

If job A pays 15 JD/day and includes childcare, and job B pays **25** JD/day without childcare, which would you prefer?

If job A pays 15 JD/day and includes childcare, and job B pays **30** JD/day without childcare, which would you prefer?

Assuming there are no logical inconsistencies in her choices, the survey then asks

questions adaptively in 1JD increments until her WTP is determined within a 1 JD range. For example:

If job A pays 15 JD/day and includes childcare, and job B pays **15** JD/day without childcare, which would you prefer? **A**

If job A pays 15 JD/day and includes childcare, and job B pays **20** JD/day without childcare, which would you prefer? **A**

If job A pays 15 JD/day and includes childcare, and job B pays **25** JD/day without childcare, which would you prefer? **B**

If job A pays 15 JD/day and includes childcare, and job B pays **30** JD/day without childcare, which would you prefer? **B** Begin adaptive section:

If job A pays 15 JD/day and includes childcare, and job B pays **23** JD/day without childcare, which would you prefer? **B** If job A pays 15 JD/day and includes childcare, and job B pays **22** JD/day without childcare, which would you prefer? **A**

At the end of the elicitation, the enumerator confirms with the respondent her choices to make sure it aligns with her preferences and that no mistakes were made. “So, when job B pays between 15-22 JD you preferred job A (15 JD with childcare), but when job B pays 23 or more JD, you prefer job B. Is that correct?”

Multiple Switching. In the event of a logical inconsistency or “multiple switching” from A to B then back to A, the survey prompts the enumerator to explain to the respondent that her choice is logically inconsistent. For example, suppose a respondent made the following choices:

15 JD + transportation $\succ$ 15 JD

15 JD + transportation $\prec$ 20 JD

15 JD + transportation $\succ$ 25 JD

15 JD + transportation $\prec$ 30 JD

The survey would read: “You preferred job B over job A when job B paid 20JD, but then you said you preferred job A over job B when job B paid 25JD. Are you

sure about this?” Enumerators were trained to explain to the respondent why this was illogical if needed.

Never Switching. Open ended questions were asked if respondents never switched their choice. If they always preferred job B (implying that they value the amenity at zero), the enumerator asked them to explain why they always preferred job B, and these responses were scrutinized for logical coherency.¹ Respondents who always preferred job A (ie, WTP is larger than the range of the price list) were asked to report their true switch point with the following language: “What would job B’s wage need to be to persuade you to switch from A to B? Your answer will not affect the random draws in any way.”²

A.3 Mathematical Derivations

Equation (1.2) can be derived as follows:

$$\begin{aligned}
 F(w, a) &= Prob(u_i > 0) \\
 &= Prob(w + v_i a - b_i > 0) \\
 &= Prob(w + v_i a > b_i) \\
 &= Prob\left(\frac{w + v_i a - \underline{b}}{\bar{b} - \underline{b}} > \frac{b_i - \underline{b}}{\bar{b} - \underline{b}}\right) \\
 &= Prob\left(\frac{w + v_i a - \underline{b}}{\bar{b} - \underline{b}} > \beta\right)
 \end{aligned}$$

¹Reasons for not valuing childcare include having family members available for caregiving; reasons for not valuing transportation include having her own car, or enjoying a short walk to work, etc. Reasons for not valuing monthly paychecks often mentioned reasons related to saving/self-control.

²While this language does not directly state “what is the *lowest* wage that would prompt you to switch”, it was found to be implied in piloting, and that “lowest” introduced confusion.

$$= \left[\frac{w + v_i a - \underline{b}}{\bar{b} - \underline{b}} \right]^\phi \quad (\text{A.1})$$

The derivation of equation (1.3) is shown below. In an abuse of notation, allow $f(x)$ to be the probability density function of random variable x .

$$\begin{aligned}
U &= \int_{-\infty}^{\infty} \int_{\underline{b}}^{w+va} (w + va - b) f(v, b) dv db && \text{total excess utility} \\
&= \int_{-\infty}^{\infty} \int_0^{\frac{w+va-\underline{b}}{\bar{b}-\underline{b}}} (w + va - (\beta(\bar{b} - \underline{b}) + \underline{b})) f(v, \beta) dv d\beta && \text{from } \beta_i := \frac{b_i - \underline{b}}{\bar{b} - \underline{b}} \\
&= \int_{-\infty}^{\infty} f(v) \int_0^{\frac{w+va-\underline{b}}{\bar{b}-\underline{b}}} (w + va - \underline{b} - (\beta(\bar{b} - \underline{b}))) f(\beta) dv d\beta && \text{by } v \perp\!\!\!\perp \beta \\
&= \int_{-\infty}^{\infty} f(v) \left[\int_0^{\frac{w+va-\underline{b}}{\bar{b}-\underline{b}}} [w + va - \underline{b}] f(\beta) d\beta - \int_{\beta=0}^{\frac{w+va-\underline{b}}{\bar{b}-\underline{b}}} [\bar{b} - \underline{b}] \beta f(\beta) d\beta \right] dv \\
&= \int_{-\infty}^{\infty} f(v) \left[\int_0^{\frac{w+va-\underline{b}}{\bar{b}-\underline{b}}} [w + va - \underline{b}] f(\beta) d\beta - \int_{\beta=0}^{\frac{w+va-\underline{b}}{\bar{b}-\underline{b}}} [\bar{b} - \underline{b}] \phi \beta^\phi d\beta \right] dv && \text{by } f(\beta) = \phi \beta^{\phi-1} \\
&= \int_{-\infty}^{\infty} f(v) \left[(w + va - \underline{b}) \cdot \left[\frac{w + va - \underline{b}}{\bar{b} - \underline{b}} \right]^\phi - (\bar{b} - \underline{b}) \frac{\phi}{\phi + 1} \left[\frac{w + va - \underline{b}}{\bar{b} - \underline{b}} \right]^{\phi+1} \right] dv && (\text{A.2})
\end{aligned}$$

A.4 Full Text of Survey Instrument

The English text of the WTP elicitation section can be found [here](#) and the rest of the survey can be found [here](#). The Arabic versions are available upon request.

Appendix B

Appendix to Economic and Social Impacts of Entering the Labor Market During a Recession: Evidence from Egypt

B.1 Appendix Tables and Figures

Table B.1: Long-run Effect of Entry Unemployment Rate on Male Employment: Less than Secondary

	(1)	(2)	(3)	(4)	(5)
	Employment	Self Emp	WageWork	Formal WW	Public Entprse Formal WW
UR*exp=0	-0.00801 (0.00522)	-0.00407 (0.00273)	-0.00394 (0.00395)	0.00128 (0.00213)	0.00117 (0.000921)
UR*exp=1	-0.00394 (0.00396)	-0.00271 (0.00245)	-0.00122 (0.00345)	0.00221 (0.00208)	0.00162 (0.000994)
UR*exp=2	0.000267 (0.00377)	-0.00136 (0.00241)	0.00163 (0.00330)	0.00374* (0.00217)	0.00225* (0.00114)
UR*exp=3	-0.000210 (0.00399)	-0.00109 (0.00234)	0.000877 (0.00362)	0.00436* (0.00217)	0.00259** (0.00116)
UR*exp=4	0.00102 (0.00390)	-0.000847 (0.00227)	0.00185 (0.00348)	0.00477** (0.00209)	0.00270** (0.00118)
UR*exp=5	0.00262 (0.00359)	-0.000392 (0.00237)	0.00302 (0.00311)	0.00474** (0.00190)	0.00258* (0.00127)
UR*exp=6	0.00291 (0.00325)	-0.000415 (0.00224)	0.00335 (0.00293)	0.00443** (0.00181)	0.00233* (0.00130)
UR*exp=7	0.00340 (0.00288)	-0.000989 (0.00250)	0.00440 (0.00293)	0.00500** (0.00188)	0.00257* (0.00142)
UR*exp=8	0.00334 (0.00273)	-0.000252 (0.00234)	0.00362 (0.00274)	0.00516** (0.00193)	0.00266* (0.00147)
UR*exp=9	0.00374 (0.00286)	0.000342 (0.00257)	0.00342 (0.00286)	0.00551*** (0.00189)	0.00269* (0.00152)
UR*exp=10+	0.00232 (0.00234)	0.000967 (0.00281)	0.00142 (0.00261)	0.00591** (0.00225)	0.00404* (0.00198)
Constant	0.849*** (0.0160)	0.284*** (0.0129)	0.567*** (0.0141)	0.128*** (0.0107)	0.0141 (0.00862)
Current Year	Yes	Yes	Yes	Yes	Yes
Potential Experience	Yes	Yes	Yes	Yes	Yes
Cohort	Yes	Yes	Yes	Yes	Yes
Gov of birth	Yes	Yes	Yes	Yes	Yes
Clustering	Gov Birth	Gov Birth	Gov Birth	Gov Birth	Gov Birth
Entry Yrs	1968-2018	minus 75,76,85-89,96,03			
Obs	129157	128986	128526	128518	128509

Note: This table reports estimates from the same estimating equations as reported in table 2.5, restricted to the sample of men who have attained less than secondary school.

Table B.2: Long-run Effect of Entry Unemployment Rate on Male Employment: Secondary Grads

	(1) Employment	(2) Self Emp	(3) WageWork	(4) Formal WW
UR*exp=0	-0.00801 (0.00522)	0.0000709 (0.00195)	-0.000552 (0.00200)	-0.000301 (0.00220)
UR*exp=1	-0.00394 (0.00396)	-0.000692 (0.00226)	0.00107 (0.00233)	0.00106 (0.00216)
UR*exp=2	0.000267 (0.00377)	0.000168 (0.00208)	0.000712 (0.00249)	0.000330 (0.00220)
UR*exp=3	-0.000210 (0.00399)	0.000630 (0.00192)	-0.00168 (0.00246)	-0.00162 (0.00215)
UR*exp=4	0.00102 (0.00390)	0.000383 (0.00201)	-0.00269 (0.00240)	-0.00375 (0.00227)
UR*exp=5	0.00262 (0.00359)	0.00139 (0.00209)	-0.00220 (0.00226)	-0.00498** (0.00225)
UR*exp=6	0.00291 (0.00325)	0.000976 (0.00219)	-0.00265 (0.00213)	-0.00547** (0.00229)
UR*exp=7	0.00340 (0.00288)	0.00169 (0.00230)	-0.00298 (0.00243)	-0.00651** (0.00250)
UR*exp=8	0.00334 (0.00273)	0.00168 (0.00265)	-0.00289 (0.00266)	-0.00747** (0.00276)
UR*exp=9	0.00374 (0.00286)	0.00168 (0.00278)	-0.00184 (0.00283)	-0.00736** (0.00297)
UR*exp=10+	0.00232 (0.00234)	0.00163 (0.00374)	0.00237 (0.00291)	-0.00597* (0.00344)
Constant	0.849*** (0.0160)	0.154*** (0.0214)	0.621*** (0.0173)	0.293*** (0.0199)
Current Year	Yes	Yes	Yes	Yes
Potential Experience	Yes	Yes	Yes	Yes
Cohort	Yes	Yes	Yes	Yes
Gov of birth	Yes	Yes	Yes	Yes
Clustering	Gov Birth	Gov Birth	Gov Birth	Gov Birth
Entry Yrs	1968-2018 minus 75,76,85-89,96,03			
Obs	129157	102556	102492	102474

Note: This table reports estimates from the same estimating equations as reported in table 2.5, restricted to the sample of men who have just attained secondary school (i.e., not more or less than secondary).

Table B.3: Long-run Effect of Entry Unemployment Rate on Male Employment: University Grads

	(1) Employment	(2) Self Emp	(3) Employer	(4) Self (No emp)	(5) WageWork	(6) Formal WW
UR*exp22=0	-0.00107 (0.00247)	-0.00245 (0.00205)	-0.000669 (0.00103)	-0.00178 (0.00152)	0.00126 (0.00290)	0.00499* (0.00280)
UR*exp22=1	0.000373 (0.00270)	-0.00116 (0.00260)	-0.000570 (0.00100)	-0.000598 (0.00221)	0.00132 (0.00292)	0.00342 (0.00269)
UR*exp22=2	0.00225 (0.00264)	0.000181 (0.00250)	-0.000855 (0.000945)	0.00103 (0.00222)	0.00194 (0.00316)	0.00116 (0.00289)
UR*exp22=3	-0.000679 (0.00295)	0.00119 (0.00252)	-0.000198 (0.00105)	0.00139 (0.00202)	-0.00195 (0.00300)	-0.00366 (0.00286)
UR*exp22=4	-0.00336 (0.00274)	0.00175 (0.00229)	-0.000270 (0.00108)	0.00202 (0.00183)	-0.00512 (0.00301)	-0.00517 (0.00314)
UR*exp22=5	-0.00209 (0.00228)	0.00239 (0.00235)	0.000444 (0.00124)	0.00195 (0.00179)	-0.00446 (0.00291)	-0.00622* (0.00329)
UR*exp22=6	-0.00310 (0.00230)	0.00174 (0.00216)	0.0000925 (0.00139)	0.00166 (0.00165)	-0.00479* (0.00232)	-0.00675** (0.00309)
UR*exp22=7	-0.00303 (0.00193)	0.00166 (0.00212)	0.000602 (0.00147)	0.00107 (0.00143)	-0.00465* (0.00236)	-0.00743** (0.00334)
UR*exp22=8	-0.00232 (0.00151)	0.00215 (0.00216)	0.00150 (0.00172)	0.000658 (0.00127)	-0.00446** (0.00211)	-0.00576 (0.00338)
UR*exp22=9	-0.00313** (0.00150)	0.00278 (0.00213)	0.00175 (0.00164)	0.00103 (0.00141)	-0.00589*** (0.00193)	-0.00735** (0.00311)
UR*exp22=10+	-0.000940 (0.00150)	0.00511* (0.00257)	0.00281 (0.00254)	0.00231 (0.00182)	-0.00591* (0.00286)	-0.00853** (0.00325)
Constant	0.816*** (0.0114)	0.0907*** (0.0142)	0.0469*** (0.0109)	0.0442*** (0.0122)	0.730*** (0.0140)	0.561*** (0.0183)
Current Year	Yes	Yes	Yes	Yes	Yes	Yes
Potential Experience	Yes	Yes	Yes	Yes	Yes	Yes
Cohort	Yes	Yes	Yes	Yes	Yes	Yes
Gov of birth	Yes	Yes	Yes	Yes	Yes	Yes
Clustering	Gov Birth	Gov Birth	Gov Birth	Gov Birth	Gov Birth	Gov Birth
Entry Yrs	1968-2018 minus 75,76,85-89,96,03					
Obs	54369	54336	53947	53947	53947	53947

Note: This table reports estimates from the same estimating equations as reported in table 2.5, restricted to the sample of men who have attained at least a university degree.

Table B.4: Long-run Effect of Entry Unemployment Rate on Male Employment: Post 1989

	(1) Employment	(2) WageWork	(3) Self Emp	(4) Self (no employees)	(5) Formal WW
UR*exp=0	-0.00418** (0.00184)	-0.00470*** (0.00163)	0.000538 (0.00147)	-0.00119 (0.00139)	-0.00370** (0.00172)
UR*exp=1	-0.00331* (0.00161)	-0.00361* (0.00176)	0.000314 (0.00133)	-0.00148 (0.00134)	-0.00282 (0.00180)
UR*exp=2	-0.00222 (0.00182)	-0.00319 (0.00192)	0.000952 (0.00127)	-0.000900 (0.00122)	-0.00254 (0.00166)
UR*exp=3	-0.00398** (0.00190)	-0.00554** (0.00214)	0.00151 (0.00120)	0.0000573 (0.000921)	-0.00358** (0.00159)
UR*exp=4	-0.00334 (0.00234)	-0.00526** (0.00248)	0.00188 (0.00114)	0.000853 (0.000837)	-0.00457*** (0.00159)
UR*exp=5	-0.000432 (0.00212)	-0.00339 (0.00215)	0.00291** (0.00138)	0.00173 (0.00111)	-0.00410** (0.00154)
UR*exp=6	-0.000349 (0.00198)	-0.00331 (0.00215)	0.00292** (0.00120)	0.00181* (0.000969)	-0.00365* (0.00178)
UR*exp=7	0.000714 (0.00152)	-0.00228 (0.00201)	0.00296** (0.00137)	0.00132 (0.000982)	-0.00327* (0.00178)
UR*exp=8	0.000882 (0.00130)	-0.00251 (0.00200)	0.00335*** (0.00117)	0.00132 (0.000961)	-0.00338* (0.00183)
UR*exp=9	0.00179 (0.00132)	-0.00168 (0.00197)	0.00343*** (0.00101)	0.00167* (0.000907)	-0.00329 (0.00193)
UR*exp=10+	0.00250** (0.00117)	-0.000664 (0.00178)	0.00314** (0.00136)	0.00147* (0.000824)	-0.00150 (0.00154)
Constant	0.748*** (0.0109)	0.605*** (0.0163)	0.143*** (0.00887)	0.108*** (0.00615)	0.186*** (0.0134)
Current Year	Yes	Yes	Yes	Yes	Yes
Potential Experience	Yes	Yes	Yes	Yes	Yes
Cohort	Yes	Yes	Yes	Yes	Yes
Gov of birth	Yes	Yes	Yes	Yes	Yes
Clustering	Gov Birth	Gov Birth	Gov Birth	Gov Birth	Gov Birth
Entry Yrs	1990-2018 minus 96,03				
Obs	167149	166820	166820	166820	166819

Note: This table reports estimates from the same estimating equations as reported in table 2.5, restricted to the sample of men who turned 18 after 1989 and before 2018.

Appendix C

Appendix to Housing Subsidies for Refugees: Experimental Evidence on Life Outcomes and Social Integration in Jordan

C.1 Lee Bounds

Table C.1: Pooled Treatment Effects on Primary Outcomes (with Lee Bounds)

Outcome	TOT	(se)	FDR q-val	Lower Bound	Upper Bound	Contrl mean	(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 C.2: Statistically Significant Impacts of the Program at Midline (with Lee Bounds)

Outcome	TOT	(se)	FDR q-vals	Lower Bound	Upper Bound	Contrl 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 consumption (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 in household	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 first 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 number 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). 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 self-explanatory. 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 lockdown: (January 15 - March 15, 2020); during lockdown: (March 15 - May 15, 2020); and after lockdown: (May 15 - interview date). 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 and is coded as 1 if “0”, 2 if “1-5”, 3 if “6-10”, 4 if “11-20”, and 5 if “More than 20”. The other outcomes are as explained in the labels. 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. The lower and upper bounds refer to the ones proposed in Lee, 2009 to correct for attrition. The lower bounds marked by NA for number of children is 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.

Table C.3: Statistically Significant Impacts of the Program 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								
Focus 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 USD 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 (USD 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 C.4: Impacts of the Program on Neighbor’s Attitudes Toward Refugees (with Lee Bounds)

Outcome	TOT	(se)	FDR q-val	Lower Bound	Upper Bound	Contrl 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 & perceptions (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								
Days of media consumption (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.

C.2 Heterogeneous Effects Analysis

Table C.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 C.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.

C.3 Additional Results

Table C.7: Impacts of the Program on Earnings, Labor, & Occupational Choice

Outcomes	TOT	(se)	FDR q-vals	Contrl 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 C.8: Robustness of Baseline Balance

	Outcomes	TOT	(se)	FDR q-vals	Contrl 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: Outcome = Controls + FE + error. From Equation 1, we obtain fitted values. Then, we use them as the outcome variable in Equation 2: Outcome = β · Treatment + FE + 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.

C.4 Addressing Low Take-up Rates

Table C.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)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Pt. Est.	SE	Pt. Est.	SE	Mean	SD	Pt. Est.	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: Badiah N.	High	-0.02	(0.05)	-0.01	(0.04)	0.14	(0.35)	0.01	(0.03)
District: Badiah 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)

Continued on next page

Table C.9 – Baseline averages, continued from previous page

Outcome	Assignment	Noncompliers (A=1, T=0)		Compliers (A=1, T=1)		Control group (A=0, T=0)		(3)-(1)	
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Pt. Est.	SE	Pt. Est.	SE	Mean	SD	Pt. Est.	SE
Team: ICLA Mafrq	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 Mafrq	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)

Continued on next page

Table C.9 – Baseline averages, continued from previous page

Outcome	Assignment	Noncompliers (A=1, T=0)		Compliers (A=1, T=1)		Control group (A=0, T=0)		(3)-(1)	
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Pt. Est.	SE	Pt. Est.	SE	Mean	SD	Pt. Est.	SE
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 recieved. 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 C.10: Variables selected in the lasso step

Covariate	Treatment group		
	All	High	Low
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 * IKSscore2		x	
team 2 * changeshelter w		x	
team 3 * genvul sc		x	
team 5 * PeopleLivingScore		x	
Continued on next page			

Table C.10 – continued from previous page

Covariate	Treatment group		
	All	High	Low
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	
Continued on next page			

Table C.10 – continued from previous page

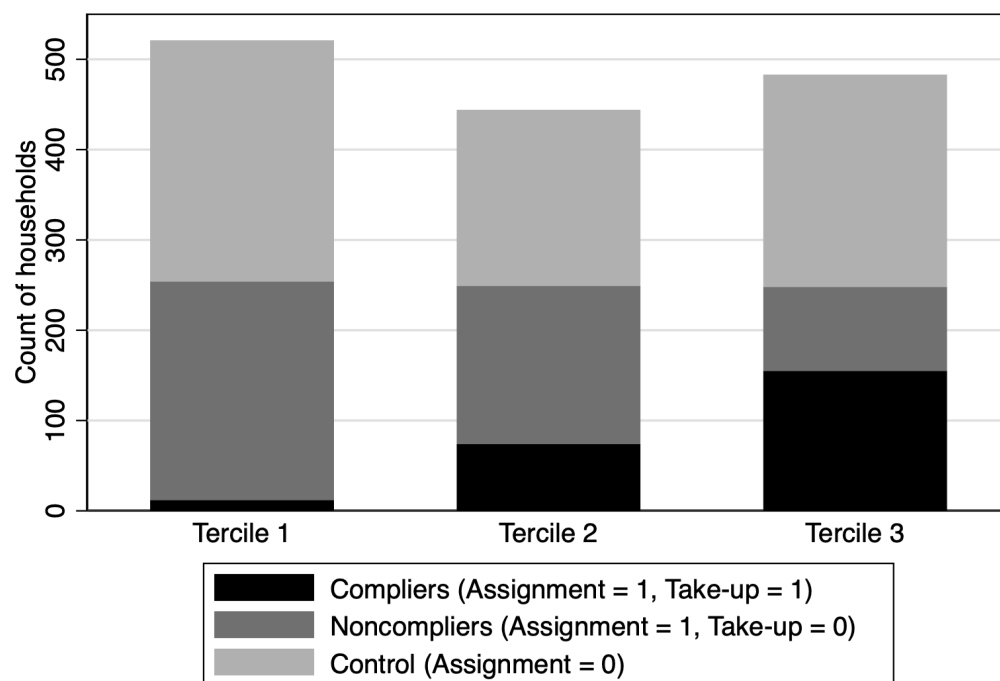
Covariate	Treatment group		
	All	High	Low
PeopleLivingScore * Rentpermonth w		x	
FamilyScore * changeshelter w		x	
SingleHoHScore * flex sc		x	
SingleHoHScore * REScore2		x	
SingleHoHScore * WASHScore2		x	
SingleHoHScore * IKSore2		x	
SingleHoHScore * changeshelter w		x	
DebtScore * REScore2		x	
flex 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	
Continued on next page			

Table C.10 – continued from previous page

Covariate	Treatment group		
	All	High	Low
cons	x		

Note: This table shows the variables selected in the lasso model.

Figure C.1: Actual compliance by predicted tercile (unique households assigned to treatment across all rounds)



N = 1448.

Figure C.2: Predicted tercile using the LASSO on a logistic regression (all rounds)

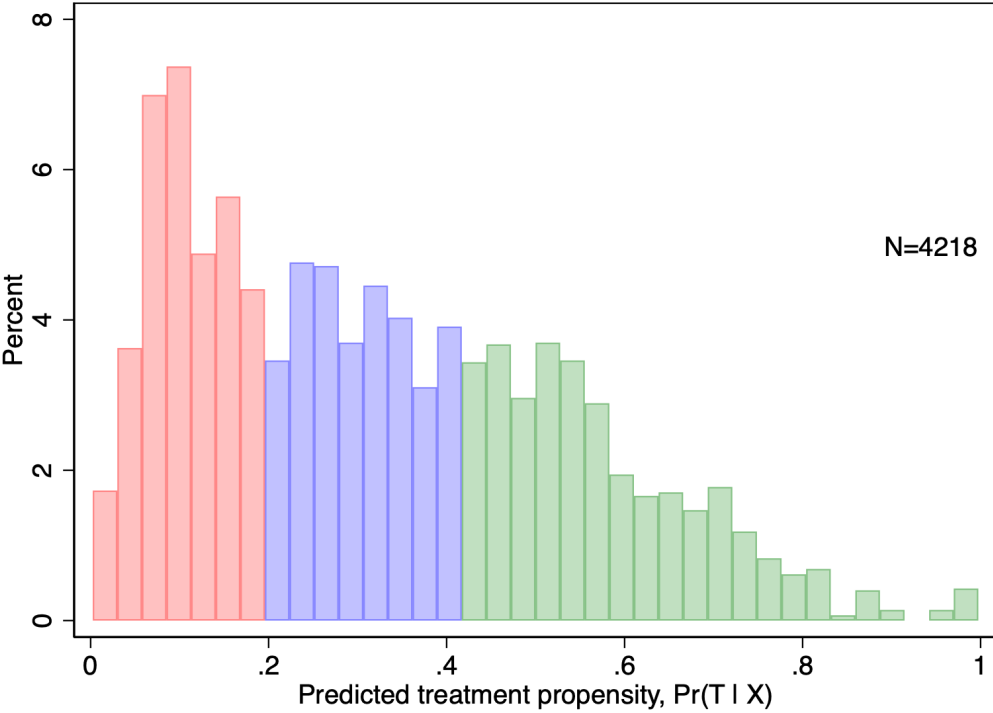


Table C.11: T4, Panel A: Primary Outcomes

Outcome	Assg.	Tercile 2	Tercile 3	Assg. $\times$ Terc. 2	Assg. $\times$ Terc. 3	Contrl Mean	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
Total Monthly Housing Expenditures (USD 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 USD 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 Child 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: Note: 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. Standard errors and standard deviations are shown in parentheses, and FDR q-values are shown in brackets.

Table C.12: T7, Panels A & B: Primary Outcomes and Selected Social Attitudes Index Components

Outcome	Assg.	Tercile 2	Tercile 3	Assg. $\times$ Terc. 2	Assg. $\times$ Terc. 3	Contrl Mean	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 & perceptions (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 Syrian 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.

C.4.1 Intent-to-Treat Results

Table C.13: Intent-to-Treat - Pooled Treatment Effects on Primary Outcomes

Outcomes	ITT (se)	TOT (se)	FDR q-vals	Contrl mean	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
Total Monthly Housing Expenditures (USD PPP)	-28.99*** (11.50)	-84.71*** (33.10)	[0.07]	228.80 (208.80)	1,421	2
Food Consumption (Log USD PPP)	-0.01 (0.03)	-0.02 (0.07)	[0.63]	4.41 (0.72)	2,753	2, 3
Log Total Consumption (Log USD 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 Child 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 from Table 4, showing Intent-to-Treat effects for comparison.

Table C.14: Intent to Treat - Statistically Significant Impacts of the Program at Midline

Outcomes	ITT	(se)	Contrl 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 from Table 5, showing Intent-to-Treat effects instead.

Table C.15: Intent to Treat - Statistically 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 from Table 6, showing Intent-to-Treat effects instead.

Table C.16: Intent to Treat Impacts of the Program on Neighbor's Attitudes Toward Refugees

Outcomes	ITT (se)	TOT (se)	FDR q-val	Contrl 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 from Table 7, showing Intent-to-Treat effects for comparison.