Coordination and Prosperity: Poverty Alleviation and Agricultural Markets in Sub-Saharan Africa

Moritz Poll

B.Sc. Economics, Maastricht University, 2015

M.Phil. Economics, University of Oxford, 2017

M.A. Economics, Brown University, 2021

Thesis

Submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in the Department of Economics at Brown University

PROVIDENCE, RHODE ISLAND

May 2026

This dissertation by Moritz Poll is accepted in its present form

by the Department of Economics as satisfying the

dissertation requirement for the degree of Doctor of Philosophy.

Date: ____________ ____________________________________

Andrew Foster, Advisor

Recommended to the Graduate Council

Date: ____________ ____________________________________

Stelios Michalopoulos, Reader

Date: ____________ ____________________________________

Peter Hull, Reader

Approved by the Graduate Council

Date: ____________ ____________________________________

David Lindstrom, Dean of the Graduate School

Moritz Poll

www.moritz-poll.com
moritz.poll@brown.edu

Academic Appointments

Rice University, Houston, TX
Assistant Professor of Economics 2027 –

University of Chicago, Chicago, IL
Postdoctoral Fellow, Development Innovation Lab 2026 – 2027

Education

Brown University, Providence, RI
PhD Economics 2020 – 2026
MA Economics 2020 – 2021

University of Oxford, Oxford, UK
MPhil Economics 2015 – 2017

Maastricht University, Maastricht, NL
BSc Economics, *Cum Laude* 2012 – 2015

Publications

- **Stress, Ethnicity, and Prosocial Behavior** *with Sara Lowes, Johannes Haushofer, Abednego Musau, David Ndetei, Nathan Nunn, & Nancy Qian*, **Journal of Political Economy Microeconomics.** 2023

Working Papers

- **Poverty alleviation at high frequency (JMP)**
- **Micro-enterprise saturation: Critical mass or overcrowding?**
- **Market Day Scheduling, Market Size, and Rural Economic Development**
  Best Paper Award from the Spatial Structures in the Social Sciences (S4) Institute

Research Grants

- **Weiss Fund** Research Grant	$ 49,600	2026 – 2027
- **Gates Foundation & Innovations for Poverty Action** Grant	$ 961,379	2025 – 2027
- **Livelihood Impact Fund** Grant	$ 250,000	2025 – 2027
- **J-PAL** Learning for All Initiative Grant *(with Hyun Soo Suh)*	$ 74,800	2025 – 2027
- **J-PAL** Gender and Economic Agency Grant *(with Hyun Soo Suh)*	$ 49,696	2025 – 2026
- **CEPR** STEG PhD Research Grant	£ 10,045	2024 – 2025
- **National Science Foundation** DDRI Grant *(Advisor: A. Foster)*	$ 24,985	2023 – 2024
- **Weiss Fund** Research Grant	$ 47,921	2023 – 2024
- **Brown University** internal funding (various grants)	$ 69,681	2022 – 2024
- **CEPR** STEG PhD Research Grant	£ 14,934	2022 – 2023

Fellowships

- **NIH Population Studies and Training Center** Fellowship	\$ 49,012	2024 – 2025
- **Brown University** Research Mobility Fellowship	~ \$ 25,000	2023
- **Watson Institute** GPD Fellowship	~ \$ 50,000	2022 – 2023
- **Brown University** Shoukang Lin President's Graduate Fellowship	\$ 60,944	2021 – 2022
- **German Economic Recovery Program** Fellowship	\$ 51,200	2020 – 2021
- **University of Oxford** Vice Chancellor Fellowship	£ 1,500	2016
- Fellowship of the **German Academic Exchange Service**	€ 15,700	2016 – 2017
- Fellowship of the **German Academic Scholarship Foundation**	€ 26,200	2013 – 2017

Refereeing and Service

Journals: Quarterly Journal of Economics (3x), Journal of Development Economics (2x), Economic Development and Cultural Change (1x), European Journal of Development Research (1x)

Grant Review: SurveyCTO

Invited Talks and Presentations

2026: Allied Social Science Associations (ASSA) Meeting (2 posters); Yale Research Initiative on Innovation and Scale (Y-RISE); Oxford Center for the Study of African Economies (CSAE); CEPR PEDL; IPA Social Protection Researcher Meeting; Midwest International Economic Development Conference (MWIEDC); Population Association of America (PAA)

2025: CSAE; Urban Economics Association (UEA); Brown University PhD Student Conference; North-East Universities Gender Day (NEUGD, lightning round); Advances with Field Experiments (AFE); World Bank-KDI; North-East Universities Development Consortium (NEUDC); Econometric Society

2024: CSAE; Ninth Conference on Network Science in Economics (NSE); German Development Economics Conference; NEUDC; Y-RISE

2020: Sixth NSE; CSAE Conference; Busara Center for Behavioral Economics

2018: CSAE; Busara Center; German Economic Development Institute

2017: CSAE Conference & Seminar; St. Antony's College Graduate Research Conference

Teaching

Teaching Assistant at Brown University:

- Taught and co-designed ECON1000 "Using Big Data to Solve Economic and Social Problems"
- Taught ECON1630 "Mathematical Econometrics"
- Taught ECON1350 "Environmental Economics"

Teaching Assistant at Maastricht University:

- Taught Quantitative Methods

I would like to thank my parents and my brother for letting me roam free and fostering my curiosity, my partner for her love, calm, and unwavering support, my advisors Andrew Foster, Stelios Michalopoulos, and Peter Hull as well as the wider Brown faculty for their guidance and for treating me like a colleague, Johannes for making me fearless and training my eye for interesting questions, Angelica, Corey, and the whole administrative team at Brown for having my back, Melissa, Gertrude, Maxwell, Kennedy, and Kingsley for their trust, and Sleepy Head for her friendship. I dedicate this work to my grandparents in whose footsteps I wander and wonder. And to Marlon who has brought so much light and joy into my life.

Contents

List of Figures

List of Tables

Abstract of "Coordination and Prosperity: Poverty Alleviation and Agricultural Markets in Sub-Saharan Africa" by Moritz Poll, Ph.D., Brown University, May 2026

Human interactions play a vital role in the economy and society at large. In three dissertation chapters, I investigate how the degree of coordination between economic actors can hinder or foster economic prosperity. In chapter 1, I explore the mechanisms behind an anti-poverty program that I helped fund and implement in Malawi. I demonstrate that potential gains from the program are not realized due to a lack of coordination between poverty alleviation and public health efforts. Despite the lack of health impacts, the program is able to lift households out of poverty and could be made 40% cheaper by omitting the public health components – a vital gain for reaching more in-need families. In chapter 2, I document how our ability to successfully scale up this anti-poverty program hinges on whether we are able to get budding entrepreneurs to coordinate which sector of the economy they enter. Failure to do so would lead to intense sectoral concentration in which many of these households would remain trapped in extreme poverty since the demand for products in their local villages simply cannot sustain too many similar firms. In chapter 3, I offer a historical perspective on the role of coordination in creating prosperity. I study how villages in colonial Kenya coordinated the schedules of their weekly markets in order to increase cross-attendance. The resulting flows of traders help explain which villages later became the major trading hubs. I showcase that the study area would have likely seen fewer, larger markets if road infrastructure and marketplace investment had been better coordinated.

Introduction

Worldwide, 840 million people live in extreme poverty. Ultra-poor graduation programs (UPG) have emerged in recent years as one of the most promising poverty alleviation tools at our disposal. UPGs are bundled livelihood interventions with a dual objective: (a) Building a new livelihood by providing monthly cash transfers, entrepreneurship training, and a large cash grant or productive asset, as well as (b) avoiding backsliding into poverty through preventive health coaching and savings groups. These programs have shown tremendous promise across many contexts in lifting households out of poverty. Their popularity among implementers, funders, and policy-makers as the only ultra-poverty alleviation tool with proven long-term effectiveness is soaring, but crucial knowledge gaps persist: If the first Sustainable Development Goal of ending extreme poverty by 2030 is to be achieved, the development community needs to better understand the mechanisms behind their sustained success as well as how they scale. In chapters 1 and 2 of my dissertation I address these evidence gaps.

The first chapter of my dissertation, titled "Poverty alleviation at high frequency" has three objectives: (1) Verify whether the UPG bundle components achieve the goals that they are intended to and identify which (if any) could be cut. (2) Investigate what can account for the long-lived impacts of the UPG compared to simpler interventions. (3) Provide broader lessons on how and why to integrate high-frequency data into the program evaluation toolbox.

Analyzing mechanisms in a traditional randomized control trial (RCT) with one round of follow-up data collection is challenging when evaluating bundled programs: First, many outcomes of interest like earnings fluctuate strongly over time, making a single measurement less informative even if measured without error. Second, long recall periods can introduce measurement error, especially in populations with low literacy and numeracy. Third, by the time evaluations are conducted, all program components are typically implemented and we can only speculate about how they may have interacted. Fourth, cross-randomizing more than two program components to compare every possible combination is very costly, financially and in terms of statistical power, while facing formidable concerns of multiple hypothesis testing and inference on winners. Fifth, implementers often resist unbundling a mature, successful program. In this chapter, I demonstrate how high-frequency data can alleviate these concerns and shed light on the mechanisms behind bundled interventions at a fraction of the cost of a fully cross-randomized factorial design that faces off every possible program component combination.

The UPG components are commonly hypothesized to be complementary because the full bundle outperforms the sum of its parts, especially in the long run. The literature on UPGs has produced a rich set of hypotheses on how these components act and interact: Ongoing unconditional cash transfers (UCTs, or "consumption support") are meant to alleviate the most urgent material needs while soft- or hard-skills group training is being delivered to raise recipients' productivity. The consumption support is also supposed to "protect" a subsequent one-off cash grant or productive asset from being consumed. This big push is intended to overcome entry costs into a productive livelihood that can replace the consumption support. At the same time, considerable resources are spent on one-on-one coaching to prevent adverse health and livelihood shocks or help people cope with them better. This personal outreach also has a social empowerment goal. The formation of savings groups is encouraged to solidify informal social safety nets and allows pooling of resources for further big-push investments. Program recipients are intended to act as agents of change and share what they learned and earned with others.

To test these hypothesized mechanisms, I conducted a large RCT in Malawi, enriched with high-frequency data, in which I compare the full UPG bundle to a control group. I collected panel data from 1,500 households (both treatment and control group) over the duration of two years, amounting to 40,000 home visits. The surveys cover detailed financial diaries and high-quality data on adverse health and livelihood shocks, coping behavior, and preventive actions that would not be possible over the long recall periods in classic RCT designs. I combined them with information on program component sequencing, contextual knowledge, and a series of sub-experiments to shed light on the mechanisms behind the program's success. Attributing patterns in the data causally to one program component and not another requires assumptions. I benefit here from an extensive literature that has studied UPGs and produced hypotheses about the mechanisms at play. I derive testable predictions for them, most of which turn out to be only testable at higher frequency during the intervention or at least more convincingly than in a cross-section afterwards.

First, I assume that my data contain the exhaustive list of mediating variables through which the bundle may impact a final outcome of interest. For example, when I study meal skipping, I assume that additional meals are purchased through a budget constraint that needs to add up and my data capture all of its components. Second, I need to assume that some mediating variables are affected only by some program components. For example, I assume that the only way that savings groups can impact the budget constraint is through deposits to and loans from savings groups; own incomes are affected by trainings and the cash grant, not by savings groups or preventive health coaching; a cash grant in the future can impact current outcomes by borrowing against it, but future training or coaching information cannot have retroactive effects. Attribution then becomes a matter of confirming some and ruling out other channels. In this vein, tightly estimated null results are very helpful for my causal argument. This is where my approach contrasts starkly with a factorial experimental design: Holding the budget of the program evaluation fixed, I concentrate the full statistical power in two treatment arms (full bundle and control group), increasing my chance of either detecting a mechanism or ruling it out with meaningfully tight confidence intervals, but at the expense of making the above assumptions. A factorial design dispenses with such assumptions and has a more straightforward causal interpretation, but each program component doubles the number of treatment arms, thus halving the number of observations per arm.

My high-frequency data allow me to document a precise and very low marginal propensity to consume (MPC) the seed capital grant. I compare my findings to the only other study to date that has had the data to estimate precise MPCs in a poverty alleviation program at the time of grant disbursement. Egger et al. (2022) study a pure cash grant of a similar size in Kenya and find that 93% of it is on average consumed, while they detect no impact on productive investments. In my study, upon impact, 60.5% is invested and 15.7% of the cash grant is saved. I show that the ongoing consumption support indeed has a higher MPC than the seed capital grant. Propensity to save increases after cash transfers cease and households rely on their own incomes. It is only after the last bundle component is implemented that treatment group incomes start diverging from the control group. The fact that this bundled intervention induces a very large investment response provides a vital puzzle piece for a major debate in the poverty alleviation literature: Why do pure cash transfer programs not have long-term impacts among extremely poor populations while bundled interventions do?

Immediately after program launch, the prevalence of severe coping behaviors such as meal skipping, falls sharply in the treatment group. Four of the five bundle components are phased in at the start of the program: UCTs, entrepreneurship training, preventive health coaching, and formation of savings groups. The large one-off grant follows after trainings are completed. All of these might in principle enable higher consumption early. I causally attribute lower meal skipping to the UCTs as follows: I observe detailed and complete income and spending data. I demonstrate that at this stage of the program, treated households' own earnings are not yet exceeding those of the control group and they are not borrowing against the future large capital grant from their savings group or the market, leaving only the UCTs as a plausible explanation for reduced meal skipping. This matters if we are otherwise tempted to simplify the program by lumping all cash payments together, especially if participants might benefit from trainings more when they are not worrying about where their next meal comes from. By contrast, the one-on-one coaching component, intended to avoid backsliding into poverty by teaching preventive health behaviors and described by many implementers as the heart of the UPG approach, may not be working as well as previously thought. The program is successful in encouraging households to adopt a range of shock prevention behaviors, but this does not appear to be a consequence of the one-on-one coaching visits themselves. Almost all behavioral change follows the group training and I find little behavior change following the costly one-on-one coaching visits. I also detect no complementarities between coaching and cash transfers: Since the order of coaching topics is randomized, I can directly test whether households are more likely to follow recommended behaviors while receiving contemporaneous cash transfers (for the first six of the twelve coaching months).

More importantly, despite the adoption of shielding behaviors, treated households are no less likely to subsequently experience adverse shocks in any life domain. Treatment and control group experience adverse shocks at the same rate, including ones that are directly targeted by coaching, such as malaria, diarrhea, or respiratory diseases. This is true even for households who were randomly assigned to be coached on these behaviors while still receiving cash transfers, when they might have been more likely to follow the NGO's advice. Here, the conclusion that the coaching does not have the intended effect of lowering the arrival rate of shocks follows straightforwardly from the observation that the full bundle does not lower it.1 When households experience a shock, the propensity to resort to severe coping behaviors like meal skipping increases by similar amounts in the treatment and control group. When experiencing a severe adverse health event, households appear to largely absorb the health care costs and lower incomes in reduced consumption, while managing to keep their productive assets and savings intact.

Lastly, I uncover a source of tremendous bias for cost-effectiveness analysis when annual data snapshots are being used in contexts with outcome seasonality. I show that the earnings gap between treatment and control group depends strongly on the agricultural season in which they are collected. Using my high-frequency data, I demonstrate that the program pays for itself in 28 months, but using data from the lean season (when newly started businesses have fewer customers) increases this estimate to 71.8 months.

In the second chapter of my dissertation, titled "Micro-enterprise saturation: Critical mass or overcrowding?", I turn to the scaling properties of the graduation program. Poverty alleviation in the absence of stable employment opportunities from the private or public sector relies on individual entrepreneurship by poor people themselves. Policies that support entrepreneurship in developing countries are growing in popularity, but economic theory does not offer a clear-cut prediction of their impact at scale. On the one hand, the reason we are seeing few firms in contexts of extreme poverty might be that firms are not individually profitable as they depend on the weak demand of other similarly poor people. It is possible that a large-scale support program would create a critical mass of entrepreneurs whose income generation and associated spending multipliers make each other viable. The logic of a big-push individual-level intervention would then carry over to the aggregate (local) economy. On the other hand, encouraging the start-up of many small firms in close proximity may be a recipe for overcrowding and artificially fierce competition in a small village economy. Failure rates for micro-enterprises are high, especially soon after launch and anecdotally too many UPG-beneficiary households start similar businesses in close proximity. This would suggest tight upper bounds to how far these programs can be scaled – and more broadly to the approach of alleviating poverty bottom-up from the household level.

These two scenarios (critical mass versus overcrowding) motivate the second chapter of my dissertation. They imply contrasting policy prescriptions for whether entrepreneurship should be fostered by saturating a few locations or scaling over a larger geographic area at low saturation. In the above-described large randomized control trial in Malawi, I exogenously vary the program saturation of an entrepreneurship support training to assess whether business performance increases or decreases in the share of treated neighbors. Saturation at the household cluster level is randomly assigned to be either 50% or 100% of eligible households, corresponding to 18.5% and 37% of the population, respectively. I also test a simple coordination intervention, added to the group business training, encouraging households to coordinate with each other when choosing their business venture, rather than duplicating existing businesses.

The coordination intervention did not succeed in shifting entrepreneurs-to-be out of the retail sector. While there are weak indications of sectoral shifts in both the business plans that program participants submit as well as the livelihoods households report very shortly after the cash grant, these differences dissipate within just a few months. My study demonstrates that the barrier is not simply an information friction, but I provide evidence that the lack of hard skills confines these entrepreneurs in the petty trading sector.

Eighteen months after the start of the intervention, the saturation results are nuanced: In terms of earnings, the businesses in high and low saturation clusters initially track each other's performance closely over time, but show signs of divergence, with low saturation cluster entrepreneurs earning at times up to twice as much. The cumulative earnings are not significantly different from one another, but the point estimate for the high-saturation group is about 21% lower than that of the low-saturation group. It is important to note that an implementer may prefer to implement at high saturation even if earnings were significantly lower as long as earnings losses do not offset savings from economies of scale. In the program I analyze, the cost savings at high saturation are about 30%. Firms in the high saturation clusters are significantly more likely to exit, with households reverting to casual labor. An even starker indication of overcrowding is that the higher saturation induces out-migration.

My results in chapters 1 and 2 have important implications for poverty alleviation policy: First, the UPG bundle induces much larger productive investments and has longer-lasting impacts than pure cash transfer interventions. We may therefore want to explore whether cheaper, simpler bundles can emulate this investment response. Second, a key area for cost savings appears to be the one-on-one coaching, which has admirable behavioral impacts, but falls short from lowering the arrival rate of shocks that we ultimately care about. Such interventions may better be carried out at a district or regional scale, rather than by small organizations to the most vulnerable among many vulnerable people. Third, outcome seasonality may severely bias the results of existing program evaluations, especially cost effectiveness. High-frequency data can abate these biases and their falling cost of implementation should make them a standard part of the evaluation toolbox. Fourth, the majority of subsistence farmers lacks the hard skills to enter any non-agricultural sector other than retail and the resulting sectoral concentration cannot be overcome simply by encouraging entrepreneurs-to-be to diversify their business. But providing diversified hard skills trainings poses a formidable implementation challenge for anti-poverty programs. Fifth, local demand is insufficient to provide demand for more than a handful of retailers, which constitutes an important bottleneck for scaling up poverty alleviation programming that operates through firm creation.

In chapter 3, titled "Market Day Scheduling, Market Size, and Rural Economic Development", I shift to a historical lens and explore the long-term economic development consequences of weekly market day scheduling in the distant past.

In a world with agglomeration gains on the one hand and transport costs on the other, there is a tradeoff between the access to markets and their size. Many small markets, scattered across a geographic area, have the advantage that buyers and sellers can reach markets at low transport cost. The probability of finding a trading partner and transacting at a favorable price, however, is greatly diminished in a small market. Reliably finding customers is vital to any business, but to smallholder farmers and entrepreneurs, it is existential.

I study how this tradeoff was brokered in the context of rural market days in Western Kenya and how that has affected local economic development. Villages in this context are too small to sustain daily trade. Large shares of the consumption basket are perishable, but inadequate availability of refrigeration limits bulk purchases or storage. On any given day, a share of consumers need to purchase, while sellers sit on post-harvest perishable stockpiles, or daily fish catch, and sellers of more long-lived products still rely on regular sales for their subsistence. Rotating, periodic market days have emerged here and across the world as a solution to this challenge.

The periodic "pulsation of economic activity" that market days spark has been described as “one of the basic life rhythms in all traditional agrarian societies”. At their economic core, they solve a coordination problem: Bringing buyers and sellers together in the same place at same time. Rural periodic markets are also the final link in the food supply chain for millions of people. Market days are the site of myriad matching problems and the livelihoods of vulnerable market participants on both sides of the transaction depend on reliably finding someone to trade with at a predictable price. This process is challenging in thin markets with few buyers and sellers and can be eased through market integration that increases market attendance. In this chapter, I study an age-old, straightforward to implement, and essentially costless way of increasing the attendance to weekly markets: Scheduling market days in such a way that they do not clash between neighboring villages, thus allowing buyers and sellers to cross-attend. I investigate the impact of this strategy on present-day trade patterns: How many more sellers and buyers can a market attract if other markets on the same day are further away, thus increasing the catchment area? Conversely, if two nearby markets are splitting their surrounding area between each other on market day, how much does this hurt their attendance? To answer these questions causally, I identify a natural experiment in how market schedules were set in Western Kenya over the past century, which allows me to show that weekday scheduling in the distant past meaningfully impacted present-day market attendance by shaping market catchment areas.

I measure a market’s catchment area as the area within which it is the nearest market to reach on its market day. A catchment boundary is the line along which two markets that share a market day are equidistant (Voronoï, 1908a,b). Catchments are thus fully determined by only two factors: The locations and schedules of all markets. Ignoring other determinants of market attendance, such as population density within the catchment or transport infrastructure is not a bug but a feature of my analysis. Over the course of a century, these are endogenous to market locations and schedules. By focusing on the spatio-temporal primitives of market catchment areas, I am restricting my analysis to a proxy of a potentially richer measure of market potential, but one for which I can pin down sources of potential bias, which I can remedy.

A simple regression analysis that relates outcomes of interest to market catchment area would not identify its causal effect. Geographic fundamentals or initial conditions (e.g. in population density) may both endow some markets with a larger catchment area while simultaneously affecting the outcomes we are interested in. My goal in this study is to resolve this bias and arrive at the causal effect of market catchment area.

An in-depth literature review in anthropology, sociology, geography, history, and economics, as well as over 1,900 key informant interviews with village elders and market administrators in the study area allow me to reconstruct the market scheduling process for all 280 weekly markets in my market census. Market day scheduling was a function of the distance to existing markets at the time of market inception, region-wide weekday preferences, as well as idiosyncratic preference shocks, such as pre-existing village meetings or the weekday on which the butcher slaughtered. Using the method proposed in Borusyak and Hull (2023), I leverage this contextual knowledge to separate the observed variation in market areas into a component that was to be expected given the known geography and scheduling process, and a component of quasi-random deviations from this expectation which are the culmination of a cascade of idiosyncratic weekday preference shocks. In the analysis that follows, I use this latter component to identify the causal effect of market area and frictions on various outcomes of interest. In practice, rather than compare markets with small and large catchment areas, I compare those that we would have expected to have the same size of market area, was it not for idiosyncratic weekday preferences on their part as well as those of all of their neighbors. The high present density of markets means that eventually a market can be reached anywhere in Western Kenya on any given day of the week and all markets face competition on the same day. The identification assumption here is not that markets chose their market day at random but rather that after accounting for the scheduling process, exactly who clashed with whom and how big a market catchment area that implied is as good as random.

I find that, indeed, distance to existing markets predicts market day scheduling decisions and that simulations from the estimated scheduling model can be used to purge out imbalances in predetermined market characteristics. The quasi-random component of the scheduling decision model is then used in the remainder of the analysis. I show that the variation in market area that can be attributed to idiosyncratic weekday preferences persistently and causally impacts market attendance by both buyers and sellers and that this is driven by outsiders cross-attending from other towns. I also show that market towns that are endowed with a larger market area than the scheduling model would have us expect have a larger population today and are brighter in nighttime satellite imagery, a proxy for economic activity.

Finally, I discuss the tradeoff between travel cost and agglomeration gain that these markets are brokering. Consider drawing market catchment areas with the objective to maximize agglomeration gains subject to travel cost. One would make catchment areas so large that those furthest from the market are just made indifferent about attending the market. Those even further away would ideally be part of another market's catchment. My results suggest that a sizable number of people live in boundary areas where more than one market is competing for their attendance in a context where we are worried about markets being too thin. What's more, market schedules have at this point been mostly set and highly persistent for over 50 years. Markets in this context largely predate transport infrastructure and I document that potential gains from schedule harmonization are driven by markets that have recently been connected to a tarmac road and now face lower transport costs. If market catchment areas were optimal before the construction of most road infrastructure, my results suggest that trade activity today is scattered across suboptimally many markets and potential agglomeration gains are being left on the table.

Chapter I

Poverty alleviation at high frequency

Abstract I use high-frequency survey data in a large randomized control trial in Malawi to investigate the mechanisms behind a successful, multifaceted livelihood program. The program bundles small, monthly, unconditional cash transfers (UCTs), with entrepreneurship classes, one-on-one life skills coaching, formation of savings groups, and a big one-off seed capital grant. I track 1,500 extremely poor, rural households at fortnightly frequency across about 40,000 home visits. I find strong evidence that the livelihood components drive program success: The UCTs reduce economic hardship while participants attend trainings. Subsequently, participants invest a much higher share of the seed capital grant into their new businesses than is common in pure cash transfer interventions. Once UCTs cease, households immediately replace them with newly generated income and reinvest that in turn. The coaching component, which aims to reduce the arrival rate of health and livelihood shocks, appears to be ineffective: While treated households adopt preventive measures, their shock arrival rates remain identical to the control group and I find no evidence of complementarities with contemporaneous cash transfers. The propensity to resort to severe coping behaviors in response to a shock increases by similar amounts in treatment and control. The program pays for itself in a matter of 25 months, which would have been dramatically over- or underestimated in traditional baseline-endline RCT designs that cannot measure earnings seasonality but which are the norm in program evaluation. Throughout, the high-frequency data enable conclusions about mechanisms that would have otherwise stayed opaque.

1 Introduction

In 2022, 840 million people lived under $3.00 per day (2021 PPP, Foster et al., 2025). Ultra-Poor Graduation programs (UPGs) have been shown to alleviate extreme poverty lastingly and across many countries (Banerjee et al., 2015; Bandiera et al., 2017; World Bank, 2024). UPGs are bundled interventions with a dual objective: (a) Building a new livelihood by providing monthly cash transfers, training, and a large cash grant or productive asset, as well as (b) avoiding backsliding into poverty through preventive health coaching and savings groups. If the first Sustainable Development Goal of ending extreme poverty by 2030 is to be achieved, the development community needs to better understand the mechanisms behind this promising tool. Complexity and costs are key obstacles to scale and under globally shrinking aid budgets, cutting non-performing program components would allow more people in need to benefit from them.

Analyzing mechanisms in a traditional randomized control trial (RCT) with one round of follow-up data collection is challenging when evaluating bundled programs: First, many outcomes of interest like earnings fluctuate strongly over time (McKenzie, 2012), making a single measurement less informative even if measured without error. Second, long recall periods can introduce measurement error, especially in populations with low literacy and numeracy. Third, by the time evaluations are conducted, all program components are typically implemented and we can only speculate about how they may have interacted. Fourth, cross-randomizing more than two program components to compare every possible combination is very costly, financially and in terms of statistical power, while facing formidable concerns of multiple hypothesis testing (List et al., 2019) and inference on winners (Andrews et al., 2024). Fifth, implementers often resist unbundling a mature, successful program. In this paper, I demonstrate how high-frequency data can alleviate these concerns and shed light on the mechanisms behind bundled interventions at a fraction of the cost of a fully cross-randomized factorial design.

This paper has three objectives: (1) Verify whether the UPG bundle components achieve the goals that they are intended to and identify which (if any) could be cut. (2) Investigate what can account for the long-lived impacts of the UPG compared to simpler interventions. (3) Provide broader lessons on how and why to integrate high-frequency data into the program evaluation toolbox.

The UPG components are commonly hypothesized to be complementary (Heil et al., 2024) because the full bundle outperforms the sum of its parts, especially in the long run (Banerjee et al., 2022). The literature on UPGs has produced a rich set of hypotheses on how these components act and interact: Ongoing *unconditional cash transfers* (UCTs, or "consumption support") are meant to alleviate the most urgent material needs while soft- or hard-skills *group training* is being delivered to raise recipients' productivity. The consumption support is also supposed to "protect" a subsequent *one-off cash grant* or *productive asset* from being consumed. This big push is intended to overcome entry costs into a productive livelihood (Balboni et al., 2022) that can replace the consumption support. At the same time, considerable resources are spent on *one-on-one coaching* to prevent adverse health and livelihood shocks or help people cope with them better. This personal outreach also has a social empowerment goal. The *formation of savings groups* is encouraged to solidify informal social safety nets and allows pooling of resources for further big-push investments. Program recipients are intended to act as agents of change and share what they learned and earned with others.

To test these hypothesized mechanisms, I conducted a large RCT in Malawi, enriched with high-frequency data, in which I compare the full UPG bundle to a control group. I collect panel data from 1,500 households (both treatment and control group) over the duration of two years, amounting to 40,000 home visits (31,877 visits completed at the time of writing). The surveys cover detailed financial diaries and high-quality data on adverse health and livelihood shocks, coping behavior, and preventive actions that would not be possible over the long recall periods in classic RCT designs. I combine them with information on program component sequencing, contextual knowledge, and a series of sub-experiments to shed light on the mechanisms behind the program's success.

Attributing patterns in the data causally to one program component and not another requires assumptions. I benefit here from an extensive literature that has studied UPGs and produced hypotheses about the mechanisms at play. I derive testable predictions for them, most of which turn out to be only testable at higher frequency during the intervention or at least more convincingly than in a cross-section afterwards. First, I assume that my data contain the exhaustive list of mediating variables through which the bundle may impact a final outcome of interest. For example, when I study meal skipping, I assume that additional meals are purchased through a budget constraint that needs to add up and my data capture all of its components. Second, I need to assume that some mediating variables are affected only by some program components. For example, I assume that the only way that savings groups can impact the budget constraint is through deposits to and loans from savings groups; own incomes are affected by trainings and the cash grant, not by savings groups or preventive health coaching; a cash grant in the future can impact current outcomes by borrowing against it, but future training or coaching information cannot have retroactive effects. Attribution then becomes a matter of confirming some and ruling out other channels. In this vein, tightly estimated null results are very helpful for my causal argument. This is where my approach contrasts starkly with a factorial experimental design: Holding the budget of the program evaluation fixed, I concentrate the full statistical power in two treatment arms (full bundle and control group), increasing my chance of either detecting a mechanism or ruling it out with meaningfully tight confidence intervals, but at the expense of making the above assumptions. A factorial design dispenses with such assumptions and has a more straightforward causal interpretation, but each program component doubles the number of treatment arms, thus halving the number of observations per arm.

I offer three lessons about UPGs and the RCT methodology more broadly: First, high-frequency data allow me to draw a very precise picture of how recipients use their funds and I document a much larger investment response to this bundled program than has been found for pure cash transfers. Second, the program's success is driven by the components that create a new livelihood, while the (costly) attempt by the program to prevent setbacks and adverse shocks fails. Third, I demonstrate that the results of traditional baseline-endline RCT designs in contexts with outcome seasonality can be severely biased depending on when exactly evaluations are conducted. Ongoing monitoring data remedy this bias and should become the standard in program evaluation. I now discuss these findings in more detail and illustrate under which assumptions my findings are causal.

My high-frequency data allow me to document a precise and very low marginal propensity to consume (MPC) the seed capital grant. To the best of my knowledge, only one other study to date has had the data to estimate precise MPCs in a poverty alleviation program at the time of grant disbursement. Egger et al. (2022) study a pure cash grant of a similar size in Kenya and find that 93% of it is on average consumed, while they detect no impact on productive investments. In my study, upon impact, 60.4% is invested and 17.2% of the cash grant is saved. I show that the ongoing consumption support indeed has a higher MPC than the seed capital grant. Propensity to save increases after cash transfers cease and households rely on their own incomes (Table 1). It is only after the last bundle component is implemented that treatment group incomes start diverging from the control group. The propensity to invest right at the moment of the seed grant can be thought of as an insightful surrogate indicator for impact longevity (Athey et al., 2019). The fact that this bundled intervention induces a very large investment response provides a vital puzzle piece for a major debate in the poverty alleviation literature, which I discuss in more detail below: Why do pure cash transfer programs not have long-term impacts among extremely poor populations (Haushofer and Shapiro, 2018; Aggarwal et al., 2022) while bundled interventions do (Bandiera et al., 2017; Banerjee et al., 2021; Balboni et al., 2022)?

Immediately after program launch, the prevalence of severe coping behaviors such as meal skipping, falls sharply in the treatment group. Four of the five bundle components are phased in at the start of the program: UCTs, entrepreneurship training, preventive health coaching, and formation of savings groups. The large one-off grant follows after trainings are completed. All of these might in principle enable higher consumption early. I causally attribute lower meal skipping to the UCTs as follows: I observe detailed and complete income and spending data. I demonstrate that at this stage of the program, treated households' own earnings are not yet exceeding those of the control group and they are not borrowing against the future large capital grant from their savings group or the market, leaving only the UCTs as a plausible explanation for reduced meal skipping. This matters if we are otherwise tempted to simplify the program by lumping all cash payments together, especially if participants might benefit from trainings more when they are not worrying about where their next meal comes from (Mullainathan and Shafir, 2013).

By contrast, the one-on-one coaching component, intended to avoid backsliding into poverty by teaching preventive health behaviors and described by many implementers as the heart of the UPG approach, may not be working as well as previously thought. The program is successful in encouraging households to adopt a range of shock prevention *behaviors*, but this does not appear to be a consequence of the one-on-one coaching visits themselves. Almost all behavioral change follows the group training and I find little behavior change following the costly one-on-one coaching visits. I also detect no complementarities between coaching and cash transfers: Since the order of coaching topics is randomized, I can directly test whether households are more likely to follow recommended behaviors while receiving contemporaneous cash transfers (for the first six of the twelve coaching months).

More importantly, despite the adoption of shielding behaviors, treated households are no less likely to subsequently experience adverse shocks in any life domain, including ones that are directly targeted by coaching, such as malaria, diarrhea, or respiratory diseases. This is true even for households who were randomly assigned to be coached on these behaviors while still receiving cash transfers, when they might have been more likely to follow the NGO's advice. Here, the conclusion that the coaching does not have the intended effect of lowering the arrival rate of shocks follows straightforwardly from the observation that the full bundle does not lower it.[1] When households experience a shock, the propensity to resort to severe coping behaviors like meal skipping increases by similar amounts in the treatment and control group. When experiencing a severe adverse health event, households appear to largely absorb the health care costs and lower incomes in reduced consumption, while managing to keep their productive assets and savings intact.

Lastly, I uncover a source of tremendous bias for cost-effectiveness analysis when annual data snapshots are being used in contexts with outcome seasonality. I show that the earnings gap between treatment and control group depends strongly on the agricultural season in which they are collected. Using my high-frequency data, I demonstrate that the program pays for itself in 25 months, but using data from the lean season (when newly started businesses have fewer customers) increases this estimate to 71.8 months.

My results have important implications for poverty alleviation policy: First, the UPG bundle induces much larger productive investments and has longer-lasting impacts than pure cash transfer interventions. We may therefore want to explore whether cheaper, simpler bundles can emulate this investment response. Second, a key area for cost savings appears to be the one-on-one coaching, which has admirable behavioral impacts, but falls short from lowering the arrival rate of shocks that we ultimately care about. Such interventions may better be carried out at a district or regional scale, rather than by small organizations to the most vulnerable among many vulnerable people. Third, outcome seasonality may severely bias the results of existing program evaluations, especially cost effectiveness. High-frequency data can abate these biases and their falling cost of implementation should make them a standard part of the evaluation toolbox.

Bundled livelihood interventions like the UPG are one of the most well-evidenced poverty alleviation tools in development economics and I contribute to our understanding of how and why they work as well as they do. First developed by the NGO BRAC in Bangladesh, UPGs

[1] This would not be true if one thought that the remaining bundle components in turn increase the likelihood of getting sick through heightened exposure. In this setting that seems implausible as the studied diseases all have different channels of transmission.

have been implemented in over 400 programs worldwide (World Bank, 2024). Their high cost of \$400-\$1,700 per household is a key barrier to scale (World Bank, 2021). While simpler in-kind, conditional and unconditional cash transfers have been shown to provide effective poverty relief and invigorate micro-enterprises in the short run (Sulaiman et al., 2016; Baird et al., 2011; Blattman et al., 2014; Haushofer and Shapiro, 2016; Cahyadi et al., 2020; Brooks et al., 2022; Egger et al., 2022; Crosta et al., 2025), in-kind transfers or strong-handed conditionalities may induce distortionary effects (Bryan et al., 2021), and the evidence on long-term effects is mixed at best and favors settings in which households are above the international extreme poverty line to begin with (Gertler et al., 2012; De Mel et al., 2012; Aizer et al., 2016; Ikegami et al., 2017; Haushofer and Shapiro, 2018; Araujo et al., 2019; Millán et al., 2020; Aggarwal et al., 2022; Fiala et al., 2022; Altındağ and O'Connell, 2023). Skills training, coaching, or access to loans or savings facilities by themselves show few if any impacts (Blattman and Ralston, 2015; Berge et al., 2015; Valdivia, 2015; Bauchet et al., 2015; Chowdhury et al., 2017; Brooks et al., 2018; Sedlmayr et al., 2020; Giné and Mansuri, 2021; Leight et al., 2021, 2022). By contrast, UPGs seem to be standing the test of time while yielding results that are larger than the sum of their parts in response to the complex challenge of multidimensional poverty (Sulaiman et al., 2016; Banerjee et al., 2015; Bandiera et al., 2017; Banerjee et al., 2021; Balboni et al., 2022; Chowdhury et al., 2017; Sedlmayr et al., 2020; Bedoya et al., 2019).

I make a concrete suggestion which of the many possible component combinations may be worth trying: While coaching has been a component of many successful bundled livelihood interventions (Blattman et al., 2016; Burchi and Strupat, 2018; Sedlmayr et al., 2020; Banerjee et al., 2022; Bossuroy et al., 2022), my findings, backed up by Beam et al. (2025), suggest that individual household visits may not be as central as previously thought. This would make room for sizable budget reductions and allow the intervention to achieve much greater scale by bringing governments on board as implementers, as in Botea et al. (2021).

My high-frequency panel data collection is the major contribution of this paper on the data side. It is most closely related to the financial diaries of Collins et al. (2009), the ICRISAT data used in Merfeld and Morduch (2022), and the Townsend (2016) Thai data used in Kinnan et al. (2024), among others. The financial diaries are very detailed but collected on a much smaller sample. The ICRISAT data have a comparable sample size and span a longer time frame, but do not cover the detailed array of shocks that I track. The Townsend data cover a set of key shocks, network linkages (which my data lack), and track households for a much longer time period, though each of my treatment arms is about twice their sample size. All of those data sets are observational without a randomized intervention overlayed. Aggarwal et al. (2022) are among the first to track an RCT at high frequency in this context, though their data are collected by phone surveys and from only two households per village, while I track the entire treatment group and pure control group with high-frequency household visits. High-frequency data are rapidly becoming more affordable to collect. They increased the combined intervention and evaluation budget of this study by around 20%. By contrast, conducting a multi-armed experiment testing every possible combination of the five UPG program components in $2 \times 2 \times 2 \times 2 \times 2 = 32$ treatment arms but holding statistical power for comparisons between them constant would have required a budget more than 15 *times* larger.

This paper proceeds as follows. Section 2 describes the setting, research design, and intervention. Section 3 discusses the analysis, and the results can be found in Section 4. Section 5 concludes.

2 Experimental design

2.1 Setting and census

This study takes place in Mangochi District in southern Malawi, one of the poorest districts in the country. To determine program eligibility and convey an accurate picture of the local economy, I conducted a household census of all 29,000 households in the area, covering a population of 138,000 individuals (Appendix Table A1). The majority of the population are subsistence or pre-commercial farmers (Appendix Figure A1). Outside of the harvest season, few have any income sources other than ad hoc daily labor (*ganyu*). About 40% of aggregate production in the study area is consumed without being traded in a market. Three quarters of households report that at least one member skipped a meal in the past week (Appendix Table A2). This is in line with the fact that agricultural yields are low. The 2018 harvest lasted fewer than 20% of farmers longer than six months and four months after the 2019 harvest only half of registered households had any food reserves at all. Six months after harvest, investments in the next planting cycle (seeds, fertilizer, and labor) need to be made, but that is when a majority of the rural poor are running dangerously low on food and money, which they compensate for with casual labor, drawing them away from their own fields. Existing businesses are overwhelmingly one-person enterprises and a sizable share depend on demand within their own village.

2.2 Sample selection and treatment assignment

The sampling process is described in Appendix Figure A2. Using the Malawi-validated Simple Poverty Score Card (Schreiner, 2019) and self-reported income, I identified about 40% of households as ultra-poor and eligible for the program.[2] Using k-means geographic clustering, these roughly 11,000 households were assigned to 271 clusters. Because households tend to live nearby one another, clusters often roughly correspond to villages when these are small and remote. In larger towns, villages either contain several clusters or a cluster combines households from more than one village. To determine the sample, I deployed a custom algorithm to select 72 clusters into the study in a way that maximizes the distance between clusters and minimizes the potential for spillovers, with buffer clusters in

[2]The implementing partner excluded households from participation that have no mother of a child under the age of five as of August 1, 2023; did not appear to be ultra-poor based on either their stated income or the Simple Poverty Scorecard; in which the female was unable to engage in physical labor; whose data appeared implausible to the research team or enumerator; or who were not willing to be registered or could not be found during several (mop-up) visits. The latter two criteria were very rare.

between that are not part of the study.[3] The minimum distance between study clusters corresponds to the distance beyond which Egger et al. (2022) no longer find spillover effects of cash transfers in western Kenya.[4] In my analysis, I will therefore allow for spillovers within but not across clusters. Though this assumption might not exactly hold in practice (for example, information is likely to travel farther than cash), the assumption is that spillovers are substantially greater between neighbors within a cluster than across clusters.

Clusters were grouped into sets of three matched tuples by proximity that were assigned to a program implementer as well as being randomly assigned to one of three cohorts to be launched a few months apart from each other (see Appendix Figures A3 and A4). The cohorts are timed such that one receives the large cash grant during the planting season, one during the harvest season, and one during the agricultural lean ("hungry") season. In this paper, this allows me to separate time relative to seasonal events like the harvest from time relative to program components. I investigate the optimal timing of this program in Poll (2026b). Within each implementer's area, the three clusters were assigned at random (Bruhn and McKenzie, 2009; Bai, 2022) to either high treatment saturation (everybody treated), low treatment saturation (half treated, randomly selected at the individual level) or pure control group (nobody treated). For the purposes of this paper, this allows me to investigate spillovers from treated onto control group households. The impact of different levels of saturation on competition between new businesses is analyzed in a companion paper (Poll, 2026a). For the paper at hand, I pool data across all three cohorts as well as pooling treated households from high and low saturation clusters into a single treatment group and pure and spillover control groups into a single control group (unless otherwise noted).

Both the treatment and control groups in my study are visited regularly for monitoring visits and asked about various recommended behaviors. I classified these into six topics

[3] After cluster selection and baseline survey but before the start of the intervention, the local government designated one of our clusters as a flood-prone area and ordered it to be vacated. As our study may have encouraged people to stay in an unsafe area, we shifted to a neighboring cluster in this one instance.

[4] If anything, the study area's road network is substantially less developed, the terrain is sliced up by seasonal rivers that make effective distances larger than straight-line distances, and personal ownership of bicycles or even motorbikes is lower in southern Malawi than in western Kenya.

concerning shock prevention (sleeping under bed nets, hygiene practices such as building and using a handwashing station or cooking in a dedicated kitchen to reduce indoor air pollution, use of safe water sources and chlorination, vaccines and regular medical checkups, leak-proofing the roof, and contraception), as well as six topics that enable a resilient response when a shock occurs or are miscellaneous (having small livestock and food reserves, joining a savings group, having steady and diversified nutrition to form bodily calorie reserves, diversifying crops in case one crop has a bad season, sending children to school, giving children toys and sufficient clothing). During each monitoring visit, one of these topics is brought up by the facilitator. If a control group household does not practice the behavior, the facilitator notes this on their tablet without comment or intervention. In the treatment group, this triggers an individualized coaching session for the household. I randomize the order of these topics at the individual level, starting with either the six topics on shielding or the combination of resilience and miscellaneous topics. Each set takes about six months to complete, so the topics are taught either in parallel with the six months of consumption support or afterward. I hypothesize that recommendations made concurrently with cash payments are more likely to be implemented.

This research design element serves two purposes. First, it allows me to track coaching-specific behavioral change independently from group training-induced behavioral change. Second, if those who are coached on shielding behavior first do adopt these actions and they serve their purpose, this will enable me to generate exogenous variation in the likelihood of different shocks occurring. The ideal experiment for investigating the impact of shocks would directly subject households at random to different shocks. This would overcome the challenge that shock exposure is an endogenous decision (for example, entering animal husbandry exposes one to livestock-related shocks). Since this is neither ethical nor feasible, my timing randomization is a reasonable second-best under the implementer-imosed constraint that all households had to be coached on every topic eventually.

To balance the sample on a number of key outcome and mechanism variables (see

Table A4), I first run the treatment assignment algorithm 100,000 times (seeding each with the counter of its iteration). I then estimate baseline balance under each theoretical treatment assignment and discard assignments that lead to a statistically significant imbalance in any of the key mechanism variables at the 10% significance threshold (Morgan and Rubin, 2012). In other words, I enforce a visually clean balance table. I then randomly pick one of the permissible treatment assignments. As pointed out in both Bruhn and McKenzie (2009) and Morgan and Rubin (2012), classical inference that does not take this rerandomization into account would on average (though not in every case) be too conservative, leading to confidence intervals on the average treatment effect with overly wide coverage and too few rejections of null hypotheses. A Fisher (1935) exact test has the correct coverage, though it tests the sharp "randomization hypothesis" that the treatment had no effect on any one individual rather than no average treatment effect (Ritzwoller et al., 2025). Traditionally, researchers using rerandomization methods have done so somewhat ad hoc, rerandomizing until the balance table is satisfactory. My approach is more transparent. Given R simulated assignments and k baseline balance table variables, the likelihood of discarding an assignment at the 10% significance level is $q = 1 - (1 - 0.1)^k$ and I can therefore use the remaining $R \times (1 - q)$ simulated assignments for the randomization inference (Young, 2019).

2.3 Intervention

The intervention consists of five main components shown in Figure 1 and modeled after the original UPG developed by the NGO BRAC in Bangladesh (Banerjee et al., 2015; Bandiera et al., 2017) with two key modifications noted below. At the onset of the program, participants are encouraged to join savings groups. For the first six months, they receive cash transfers of $75 PPP per month for six months. These are unconditional and labeled for consumption. During this time, households also receive a four month entrepreneurship soft skills training covering topics such as goal setting, market research, accounting, or microfinance. In parallel (not depicted here and different from other existing UPGs) program participants (all of whom are mothers of children under five) receive an additional four-month training course on early childhood development that also covers health topics. Otherwise, the program is comparable to BRAC's and other UPGs in the literature: Treated households receive six monthly cash transfers of $20 (labeled for investments in child health and well-being as well as business experimentation prior to the large grant). At the end of the training, instead of BRAC's livestock and related training, households receive a one-off labeled[5] cash grant of $300 intended for starting a business, corresponding to about $1,100 PPP.[6] Flexibility is given to participants to leverage their own informational advantage regarding their abilities and preferences, and they receive four months of weekly general-purpose business training covering financial literacy, market analysis, budgeting, recordkeeping, funding sources, and the development of business ideas. The seed capital and consumption support combined amount to $420 (over $1,500 PPP), three times larger than the poverty trap threshold estimated in Balboni et al. (2022), equivalent to the large grants in Haushofer and Shapiro (2016) and marginally lower than those in Egger et al. (2022). The structure of cash transfers is closely aligned with evidence on how poor households themselves prefer to receive such funds, combining a regular buffer with a large lump sum that arrives after sufficient time for deliberation (Kansikas et al., 2023). For the duration of the entire first year, households in the treatment group also receive monthly one-on-one home visits with the goal of coaching the household on preventive health behaviors and other ways of avoiding or coping with adverse shocks. Treated households can also use these visits to ask questions about training contents or their businesses.

A team of external enumerators conducts three rounds of household surveys, baseline, midline, and endline, for the purpose of program evaluation. In addition to these main-wave

[5]The grant was largely unconditional, but the NGO encourages recipients to use it for starting a business or supporting an existing one. To receive the grant, one must attend the training and be able to adequately verbalize a concrete business plan, but participants can miss sessions without penalty and are coached until their plan is satisfactory and the actual implementation of the plan is not afterward enforced. This follows the recommendation in Benhassine et al. (2015).

[6]The $300 were converted at the official exchange rate to MWK 530,000 and the World Bank estimates the 2024 PPP conversion factor to be 475 MWK/USD-PPP.

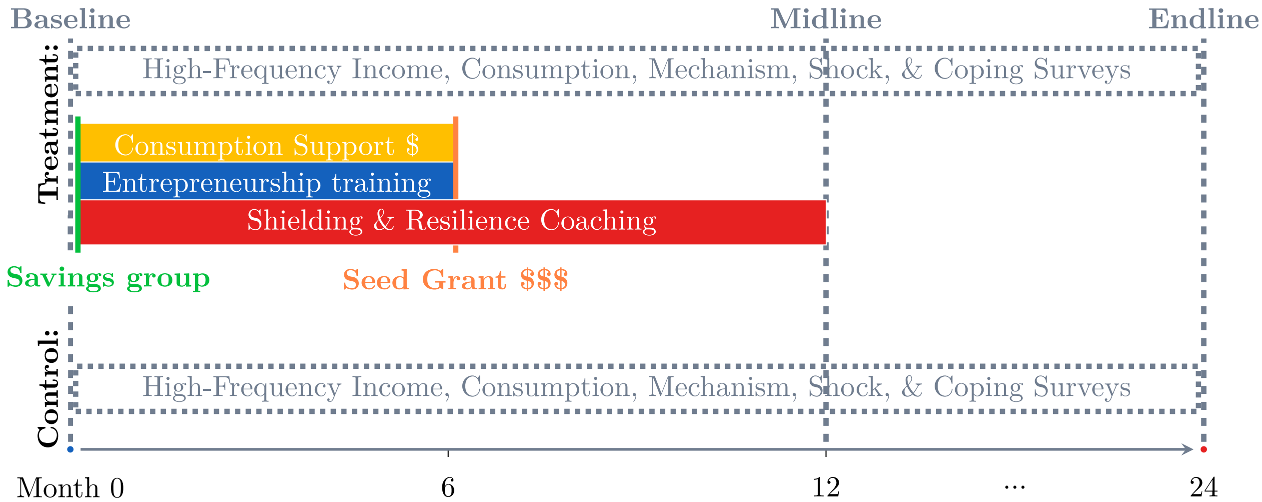


Figure 1: Sequencing of program components and data collection

Note: This figure describes the sequence of program components and their relative timing to data collection.

surveys, all households in the study, both treatment and pure control group, receive fortnightly monitoring visits from program implementers.[7] For the treatment group, these visits tend to coincide with coaching, though all surveys are conducted first, before coaching topics are being discussed. During these monitoring visits, implementing staff conduct short surveys about consumption and investment choices, livelihood activities, as well as shocks, coping, and preventive behavior. Both the treatment and control group receive a basic phone for scheduling surveys and receiving cash transfers or survey incentives through mobile money. Local childcare centers are supported financially and with the same early childhood development (ECD) training in both treatment and control groups.

2.4 Size of the intervention

In the census that I conduct for the purpose of sample selection (Appendix Table A1), I estimate the regional GDP of the study area. I use self-reported annual income from all roughly 29,000 households, and an estimate of the value of non-marketed subsistence agriculture. While income estimates in a survey are likely to be underreported if respondents

[7] The spillover control group is not visited at high frequency, only for the main wave surveys.

understand that the data will be used for targeting an assistance program, Appendix Table A2 shows that the poverty rate estimated from these self-reports aligns closely with that from a proxy means test validated for Malawi by Schreiner (2019). Appendix Table A3 documents the GDP estimates for the study area. Column (2) shows local production at the exchange rate at which the cash transfers in this study were converted from USD to MWK. The combined cash transfers per treatment household correspond to approximately one year's GDP per capita for the average five-person household in the study area and (because of negative selection), about 20 months of expenditure of a control group household. My study injects around 3.3% of one year's local marketed GDP into the study area in cash transfers alone, though this estimate falls to 1.9% when including an estimate of subsistence production in my measure of GDP. This estimate does not include survey incentives, local procurement, or spending by study staff.

3 Empirical strategy

3.1 Conceptual framework

The goal of this study is to shed light on the mechanisms behind this bundled livelihood intervention without unbundling it. Doing so in a causal and not merely descriptive manner requires not only high-frequency data, but also a number of assumptions. For each of the mechanisms I test, I need to assume that I know and measure all variables that matter for the mechanism. For example, when I study which program component is responsible for the reduction in meal skipping early on in the program, I assume first that additional food purchases happen through a budget constraint that I fully observe:

$$\begin{aligned} \text{Food}_t + \text{Other}_t + \text{Net savings}_t + \text{Investment}_t + \text{Loan repayment}_t = \\ \text{Own income}_t + \text{Cash transfers from NGO}_t + \text{Loans}_t \end{aligned} \tag{1}$$

Second, I need to make assumptions about how program components map to these observed variables. For example, I assume that the trainings may increase productivity and thus own income while the savings groups might enable borrowing. If in a given period t, I observe increased spending among the treatment group while they receive cash transfers from the NGO, but own incomes and loans do not differ from those of the control group, I can then conclude that the increased consumption was due only to the cash transfers.

3.2 Hypotheses

The graduation model has been extensively evaluated and the literature has produced an extensive list of hypothesized mechanisms that I alluded to in the introduction. Here, I restate the hypotheses about program complementarities from Banerjee et al. (2022) and Heil et al. (2024) as well as from extensive discussions with my implementing partner and other NGOs implementing similar programs. For each hypothesis, I derive a testable implication and describe how I intend to test it.

(H1) *The monthly UCTs abate economic hardship promptly.* Using baseline and high-frequency data on coping behaviors, I compare levels of coping for treatment and control groups. I measure economic hardship as severe coping behaviors like meal skipping. I assume that they can be abated with costly consumption that passes through a budget constraint and that I accurately measure all components of this budget constraint. Observing whether the additional consumption is financed by increased earnings, savings, borrowing, or the UCTs allows me to attribute differences to components.

(H2) *The UCTs protects the seed capital from being consumed.* I do not directly test this hypothesis (e.g., by omitting the consumption support from some participants' bundles) but instead compare the marginal propensity to consume the UCTs to that of the cash grant.

(H3) *The new livelihood that the seed capital and training jointly enable, fully and* *immediately replaces the UCTs.* I test this by looking for slumps in month-by-month earnings after the withdrawal of UCTs for the treatment group relative to the control group.

(H4) *The one-on-one life-skills coaching encourages households to adopt shock-shielding behaviors which reduce the arrival rate of shocks.* Using experimental variation in timing of the coaching, I test whether and when shielding behaviors are adopted. Comparing two treatment arms in which the shielding aspect is stressed either concurrently with or after the payment of consumption support, I test whether it affects shielding behavior or shock arrival rates differentially. Finally, I compare the entire treatment group to the control group in terms of both their compliance with recommended behaviors and their shock arrival rates.

(H5) *The program empowers recipients to believe that they can change their lives for the better.* To test this, I analyze the sentiment in answers to short, open-ended questions about what is on study participants' minds. I again look for patterns in the data that suggest that upticks in some components of overall sentiment coincide with the introduction or withdrawal of particular program components.

(H6) *Savings groups facilitate further aggregation of small saved sums into large productive lump-sums.* I document this using digitized savings group ledger books and benchmark the amounts against the one-off cash grant.

(H7) *The program produces positive spillovers on neighboring ultra-poor households.* I test this by comparing the spillover control group in the low-saturation clusters to the pure control group. This group is not monitored at high frequency, so I will only use main-wave survey data here.

Throughout, while some of these mechanisms could be tested using only a cross-section of the data in a traditional RCT, most are more convincingly tested while the component in question is being rolled out and comparing this to outcomes before or afterwards.

3.3 Reduced form analysis

As pre-specified,[8] I evaluate the impact of the program using the following OLS specification:

$$y_i = \beta_0 + \beta_T \text{Treatment}_i + \beta_{\text{BL}} y_i^{\text{BL}} + \beta_{\text{M}} \mathbb{1}[y_i^{\text{BL}} \text{ missing}] + \beta_{\text{FE}} S_i + \varepsilon_i \tag{2}$$

Standard errors are clustered at the 26-household cluster level at which the treatment was assigned. For the program evaluation, I estimate equation (2), in which y_i refers to post-treatment outcomes that are evaluated at midline and endline, while y_i^{BL} controls for the level of the outcome variable at baseline with the goal of absorbing residual variation in the outcome (if available). This variable is omitted for outcomes that were not collected at baseline. If an outcome was generally collected at baseline but is missing for certain observations, the observations will not be dropped from the estimation, but their baseline levels are dummied out: Their $y_i^{\text{BL}} = 0$ and $\mathbb{1}[y_i^{\text{BL}} \text{ missing}] = 1$. In accordance with the stratified randomization, I control for strata fixed effects S_i. I do not controlling for other baseline characteristics apart from the outcome itself and thanks to the treatment assignment stratification, I do not anticipate meaningful imbalances in relevant variables at baseline. To assess baseline sample balance, the baseline outcomes move to the left-hand side of the equation without control variables.

To evaluate the cross-randomized ordering intervention, I estimate

$$\begin{aligned} y_i = &\beta_0 + \beta_{\text{T}} \text{Treatment}_i + \beta_{\text{S}} \text{Shielding coaching first}_i + \\ &\beta_{\text{T}\times\text{S}} \text{Treatment}_i \times \text{Shielding coaching first}_i + \beta_{\text{BL}} y_i^{\text{BL}} + \beta_{\text{M}} \mathbb{1}[y_i^{\text{BL}} \text{ missing}] + \beta_{\text{FE}} S_i + \varepsilon_i \end{aligned} \tag{3}$$

Standard errors are again clustered at the 26-household cluster level at which the program was assigned, even though the cross-randomized coaching ordering is assigned at the individual level. Apart from the treatment interaction, this specification is identical to

[8] This study was pre-registered at https://www.socialscienceregistry.org/trials/11789.

equation (2). The coefficient of interest is $\beta_{\mathrm{T\times S}}$.

Finally, to evaluate spillovers onto nearby non-treated households, I compare control group households in the low-saturation clusters to those in pure-control clusters using the below pre-specified specification:

$$y_i = \beta_0 + \beta_T \text{Treatment}_i + \beta_{\mathrm{S}} \text{Spillover}_i + \beta_{\mathrm{BL}} y_i^{\mathrm{BL}} + \beta_{\mathrm{M}} \mathbb{1}[y_i^{\mathrm{BL}} \text{ missing}] + \beta_{\mathrm{FE}} S_i + \varepsilon_i \quad (4)$$

Standard errors are once again clustered at the 26-household cluster level at which treatment was assigned, even though within low-saturation clusters, treatment status is assigned at the individual level. The coefficient of interest is β_{S}.

4 Results

I document the following main results. First, the program alleviates extreme poverty in line with benchmark programs in its class, bostering generalizability of my findings. Second, the cash and training components achieve their stated objectives. Third, in contrast to existing evidence on pure cash transfer interventions without the accompanying bundle, this UPG induces a very high propensity to invest. Fourth, the individual coaching component does not shield participants from shocks nor increase shock resilience above and beyond what the livelihood components achieve. Fifth, the program has spillovers on untreated households, but the majority of them are positive. Sixth, cost-effectiveness analysis without high-frequency data can be severely biased in contexts with seasonality in outcomes.

4.1 Sample balance and data quality

In Table A4, I show sample balance among the 1,872 study households according to equation (2), where the baseline outcomes become the dependent variables. As discussed above, the clean balance table follows mechanically from the randomization approach. Attrition at the one-year follow-up (midline) survey is 2.7% and balanced between treatment and control group.

High-frequency data collection can be taxing for respondents. The implementers managed to consistently sample around 80% of the sample in any given month across both control and treatment group (Appendix Figure A15, top panel). Temporary dips can be attributed to countrywide fuel shortages, staff absences, and holidays, while the gradual but modest downward trend reflects out-migration. In addition to being paid an incentive for every home visit, respondents were scored by the data collector at the end of each survey. As I show in the bottom panel of Appendix Figure A15, ratings are consistently high and do not differ by treatment status.

4.2 Short-run impacts and benchmarking

At the time of the one-year follow-up survey, treated households' economic position has improved dramatically. Appendix Table A5 shows that most now run their own business and that livelihoods have shifted out of agriculture and into the retail sector. Household incomes have increased by about 60% and both business and personal asset holdings have gone up by sizable amounts. Livestock ownership nearly doubled and savings improved markedly.

At baseline, households routinely need to engage in severe coping behaviors such as skipping meals, begging, or consuming seed stock for the next harvest (Figure A9). The top panel of Figure 4 shows marked decreases in these practices among the treatment group, one year into the program, indicating heightened food security.

In Figure 2, I put these and other economic and social outcomes of the program into context by benchmarking them against the pooled results from the Banerjee et al. (2015) six-country UPG study (at the two-year mark).[9] All variables are standardized to the control group mean and standard deviation. Where available, I use the same indicator as in the Banerjee et al. study. When I report the closest equivalent, the red label is in italics. Effects

[9]This study's two-year follow-up for cohort 1 is scheduled for January 2026 at which point this figure will be updated to compare two-year results for both studies.

on assets, savings, and earnings are all aligning closely with Banerjee et al. (2015) but impacts on per capita consumption and food security are much larger. Had Malawi been a part of that study, it would have been by far the poorest country in the sample and my study population is quite uniformly poor, which explains why effects on consumption and food security are so much larger than in Banerjee et al. (2015). The benchmark UPG evaluated gave households livestock rather than a cash grant, which explains why the livestock impacts are larger there while effects on microenterprise income dominate in this study. Impacts on perceived well-being and female empowerment are similarly modest in both studies.

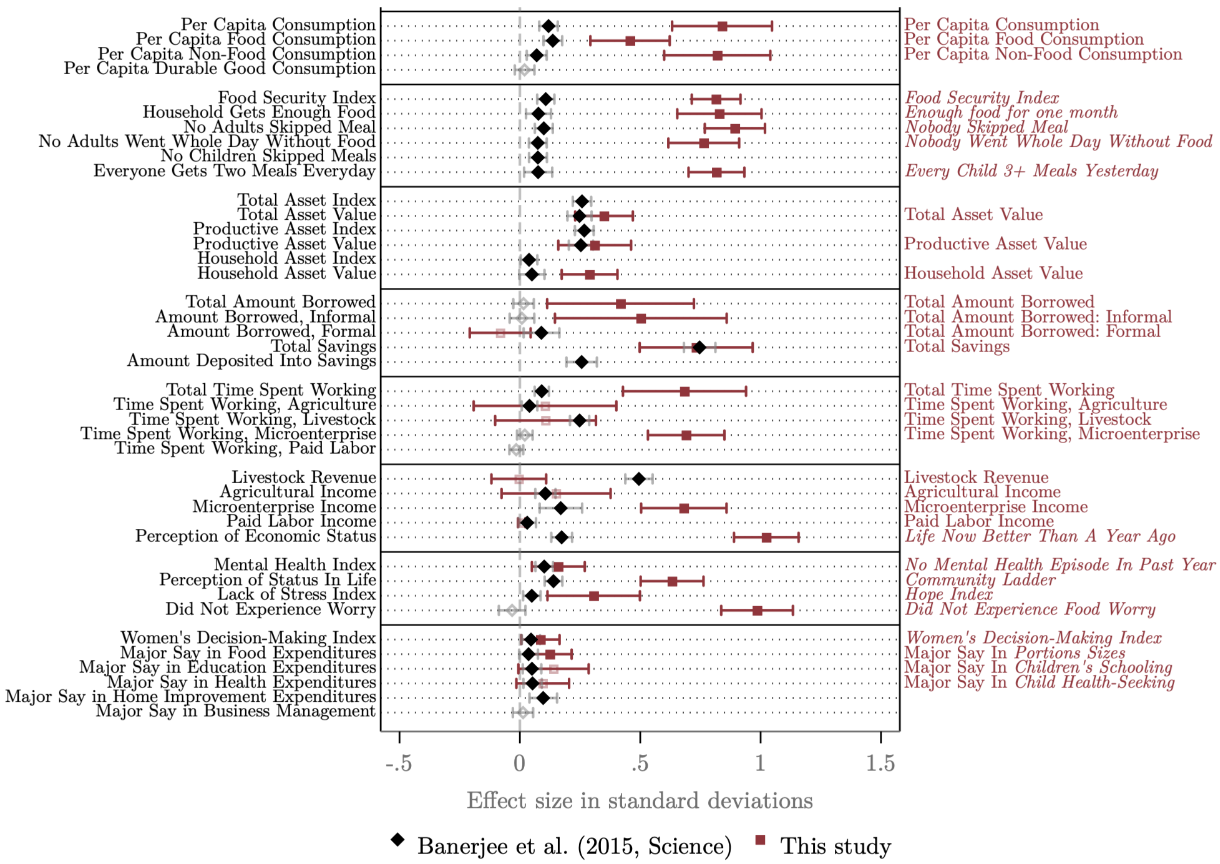


Figure 2: Benchmarking of effect sizes

Note: This figure compares the results from the one-year follow-up survey of this study to the two-year follow-up results from Banerjee et al. (2015). Black and gray point estimates and whiskers correspond to treatment effect sizes (in standard deviations) from that paper's Appendix Figure S2. Dark and light red markings correspond to effect sizes of the closest equivalent measure in this study. Where variable definitions differ, this is indicated by italics on the right-hand side.

As its stands and under the assumption that these results persist for another year, this would suggest that the studied intervention acts very similarly to the well-studied BRAC UPG and there is some reason to believe that its results may last in a similar fashion (Bandiera et al., 2017), though I can of course not show that definitively at this point in time. If the studied program is indeed representative of the graduation model more broadly, insights about the mechanisms that make it effective would be generalizable to other programs around the globe. Rather than unbundling it and potentially losing important complementarities (Banerjee et al., 2022), I make progress toward understanding the mechanisms behind this poverty alleviation program by leveraging my high-frequency panel data.

4.3 Poverty alleviation at high frequency

I next leverage the 31,877 household visits conducted (at the time of writing) during the implementation to shed light on the mechanisms behind the UPG's success. I provide evidence for or against each of the hypotheses in Section 3.

Livelihoods. The UPG raises incomes net of program cash transfers. Figure 3 decomposes earnings over time into what the control group earns (dark blue), how much the implementing NGO pays the treatment group in consumption support and cash grant (light yellow), as well as, crucially, the income that treated households generate themselves above and beyond that of the treatment group (dark yellow). Note that the NGO handed out 75% of the cash grant at the end of trainings and the remainder at the end of the first program year (second light yellow hump).

What is perhaps most striking about Figure 3 is the smooth transition from the first six months of cash transfers, followed by the cash grant, to the subsequent own income stream. Even before the seed capital grant, treated households begin experimenting with new livelihoods, which reap some additional income in the month before the grant. When cash transfers cease, own income more than fully replaces it and there is no gap between transfers and own income, suggesting that the program does not create lasting dependence on the cash transfers. The increased earnings toward the end of the graphed timeline come from the harvest for cohort 1, which more than doubles the income of both intervention groups.[10]

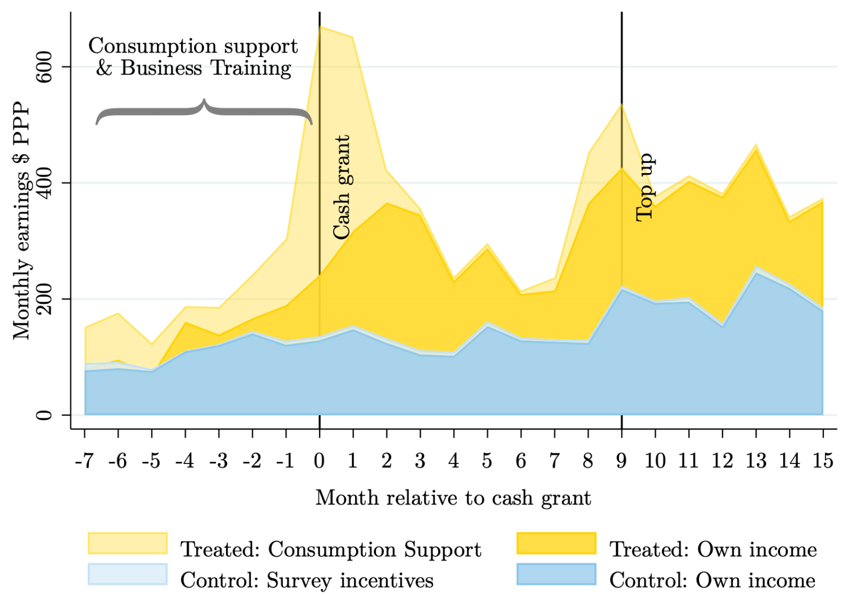


Figure 3: Household income over time

Note: This figure shows how incomes develop over time for the control group (dark blue) and the treatment group (light yellow: consumption support; dark yellow: own income generation). The two light yellow spikes show the capital seed grant disbursement, split into $225 (75%) at the end of training and $75 (25%) after the first program year.

Appendix Figure A12 describes the churn in livelihoods over the course of the study. Households rely on more than one livelihood at once, so the shares of households operating in each sector sum to more than 100%. The left panel shows that the control group relies primarily on primary sector activities such as agriculture, fishing, and firewood/charcoal production, as well as casual labor, consisting of either work on richer farmers' fields or construction jobs, paid by the day. Reliance on casual labor tends to drop at times when agricultural activity (on one's own farm) peaks. From the start, the treatment group reduce their reliance on casual labor gradually in favor of retail, though even non-retail services gain popularity. Unlike with the control group, the timing of the cash grant marks a strong shift in livelihoods for the treatment group. The abrupt trend break is strongly suggestive of a

[10]Cohorts 2 and 3 have not yet reached this point of the program, but I expect these cohorts to smoothen out the seasonality of incomes.

causal link through which the seed capital grant removes startup barriers to entry into the retail sector. Agriculture remains important and shows strong seasonality, while retail businesses are paused during agricultural peak seasons, but later revived, though the seasonality is smoothed out somewhat by the various cohorts. In Poll (2026a), I explore the reasons for and competitive consequences of this mass entry into petty trading.

Marginal propensity to consume and invest. How are the newly earned funds used? According to Table 1, the control group spends around two-thirds of its spending on food while investing less than a tenth, and saving less than a twentieth. Thanks to the random assignment of the UPG, I can safely assume that in the absence of the program, the treatment group would have spent their money in the same proportions as the control group. Comparing the observed spending shares of the treatment group therefore allows me to back out the MPC. I calculate the share of the additional funds that are now available to each treated household relative to an average control group household and then average them over the full sample to obtain the bottom panel of Table 1:

$$\mathrm{MPC}_{\mathrm{Treat}}^{\mathrm{Food}} = \frac{1}{N_{\mathrm{Treat}}} \sum_{i=1}^{N_{\mathrm{Treat}}} \frac{C_{\mathrm{Treat}}^{\mathrm{Food}} - \frac{1}{N_{\mathrm{Ccontrol}}} \sum_{i=1}^{N_{\mathrm{Control}}} C_{\mathrm{Control}}^{\mathrm{Food}}}{Y_1 - \frac{1}{N_{\mathrm{Control}}} \sum_{i=1}^{N_{\mathrm{Control}}} Y_{\mathrm{Control}}}$$

The additional funds that the treatment group has available in the early months of the program skew much more heavily toward savings and investments. Less than half of the ongoing consumption support is consumed immediately. The marginal propensities to spend are all positive but fall notably for food as the additional income rises, underscoring its status as a necessity. At the same time, the MPC for food stays positive over the full domain of additional funds given in this study, so its Engel curve does not appear to bend backward. Strikingly, households invest around 60% of the grant productively while also saving a sizable share. For comparison, in the only other study computing an estimate for a similar concept, Egger et al. (2022) find that a pure cash grant without any kind of accompanying bundle induces no productive investment response at all.[11] One may not be willing to credit the full difference between no investment without the bundle to about 60% investment in this study just to the UPG bundle. After all, these are two different studies, one happening in Kenya and one in Malawi. But the contexts are comparable enough that the UPG bundle, providing guidance on how to productively use the seed capital, must account for at least a sizable portion of this difference: The grant size in the Egger et al. (2022) study is comparable in size to the one studied here, and both take place in regions of extreme poverty near major lakes in the Bantu-speaking part of the Great Rift Valley. That the cash grant is treated as an investment opportunity here lends support to the complementarities with other program components.

An encouraging insight for the longevity of this program's impacts, in line with what the literature has found for other UPGs, is that the propensity to invest remains high even after the cash support from the NGO ceases. Both the investment response to the grant itself and later investment behavior out of own income might serve as early surrogate outcomes to aid in predicting long-run outcomes of these kinds of programs (Athey et al., 2019). Recall that without high-frequency data, columns (1) for the control group and (4) for the treatment group over a year after the start of the program would be all that is observable to a researcher.

Severe coping behavior. Severe coping behaviors, such as meal skipping or begging, are an important indicator for extreme poverty and fall dramatically over the same period, as the middle panel of Figure 4 illustrates. I plot the evolution of a coping index that uses a World Food Program (WFP) classification displayed in Figure A9, which I supplement with study-area-specific coping strategies. The shown time series assigns to each household the rank (0–3) of the most severe coping strategy they engaged in that month, akin to how WFP assesses whether they need to provide food assistance to an area.[12] Treated households

[11] Their estimate in Table A3, panel B is in fact slightly negative, but not significantly different from zero.
[12] Readers who prefer a more easily interpretable, though less comprehensive, measure of coping may refer to Appendix Figure A10, which shows the same pattern for meal skipping alone.

Table 1: Marginal propensity to consume, invest, and save

	(1) Control Group Full study	(2) Treated CS Mo 1–6	(3) Treated Grant Mo 7	(4) Treated Post-grant Mo > 7
SHARE OF SPENDING ALLOCATED TO...				
Food	64.8%	48.0%	18.2%	44.7%
Non-food consumption	22.7%	22.7%	13.6%	19.9%
Productive investment	8.4%	15.5%	52.1%	17.4%
Savings	4.1%	13.7%	16.1%	18.0%
ADDITIONALLY ALLOCATED FUNDS...				
relative to Control		+ 74.3%	+ 600.8%	+ 130.8%
MARGINAL PROPENSITY TO SPEND ON...				
Food		13.1%	9.2%	17.9%
Non-food consumption		18.2%	13.2%	17.0%
Productive investment		41.9%	60.4%	36.6%
Savings		26.9%	17.2%	28.6%

Note: This table describes households' spending allocation and marginal propensity to consume, invest, and save. The top panel describes consumption shares for the control group across the whole study period in column (1), for the treatment group in the months leading up to the large cash grant (but receiving consumption support, CS) in column (2), the treatment group in the month they receive their grant in column (3), and finally for the treatment group in the months after the cash grant in column (4). The bottom panel calculates the marginal propensity to spend on the same categories out of the funds that households had available above and beyond what the control group have.

reduce their coping behavior promptly with the first tranche of consumption support. We know from Figure 3 that at this point in the program they have not yet started generating any additional income of their own, and Appendix Figure A11 shows that treated households are if anything borrowing less than control group households, so they are not borrowing against the large cash grant. Note that this study started around the time of the 2024 growing season when the El Niño phenomenon caused a drought and the country was still recovering from the 2023 tropical cyclone Freddy. The 2025 harvest was much better. In line with this fact, I document a gradual reduction in coping behavior severity even among the control group. The treatment group, however, almost immediately relies on fewer such severe coping behaviors. Under the assumption that there are no major parts of the budget constraint that remain unobserved in my surveys, and having demonstrated that own earnings take several months to manifest while treated households are not more likely to borrow, I conclude that the consumption support must be driving these early improvements in food security. Why does it matter when during the program basic needs are being covered? We might suspect that pressing basic needs prevent households from fully benefiting from other program components – for example because in the absence of the consumption support they would have to spend their time putting food on the table or because they take the training more seriously while getting paid.

4.4 Shock incidence and impacts

Adverse shocks are common in my study setting. The typical household experiences 3.5 health episodes per year,[13] of which 29.6% are self-described as very or extremely severe. In addition, unexpected expenditures, crop and asset damages, livestock deaths, and conflicts are not uncommon. Severe adverse shocks are widely understood to have both immediate and lasting consequences for poor households. It is therefore not surprising that the architects of the graduation model sought to protect participants from setbacks. A

[13] This number is not an integer since households were observed for more than a year and I scale episode counts accordingly.

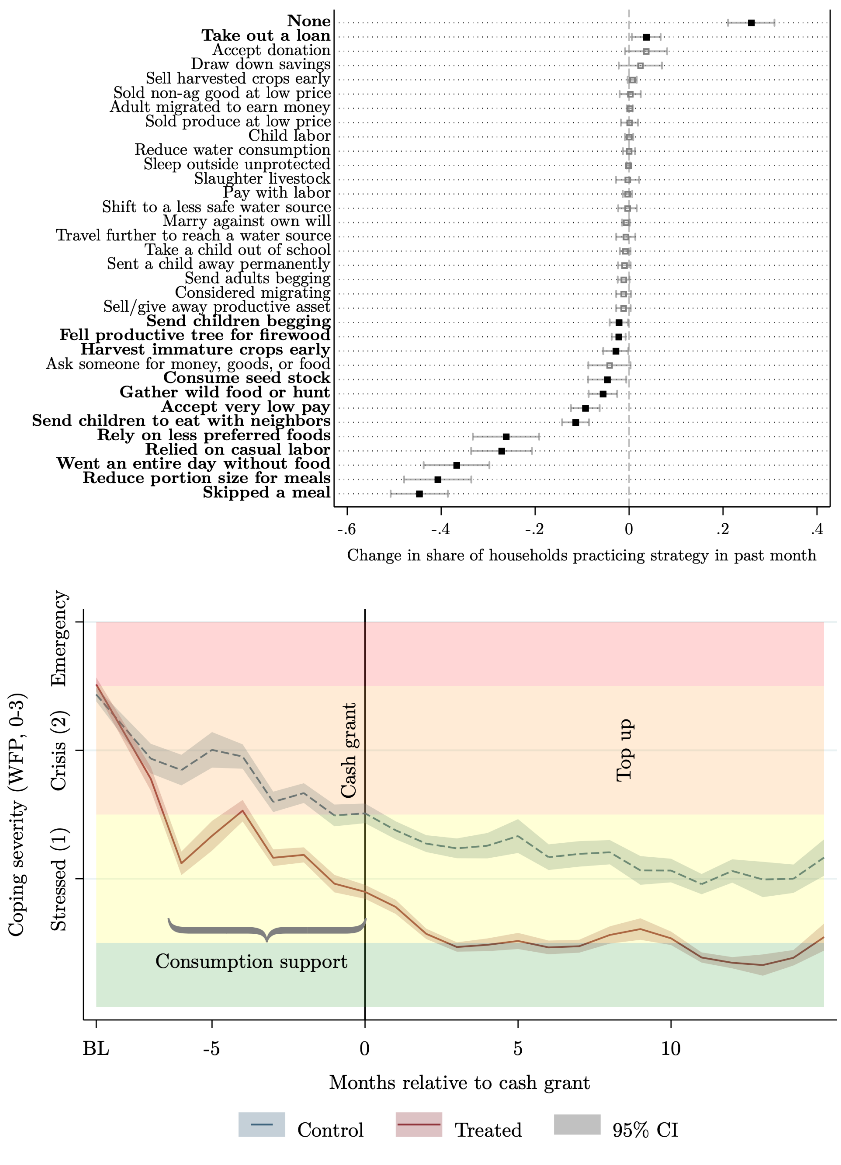


Figure 4: Effects on severity of coping behavior

Note: This figure shows the effects of the program on households' reliance on severe coping mechanisms.

significant portion of the program budget is therefore allocated to one-on-one life-skills coaching meant to encourage households to adopt preventive health behaviors such as sleeping under a bed net or building a handwashing station, as well as resilience-building measures such as maintaining food reserves or rearing small livestock. The coaching component is intended to improve resilience to such shocks and to lower the arrival rate of shocks. I now test these hypotheses in turn.

First, I explore what happens when households experience a severe shock. Severe shocks are defined here as (a) the death of or (b) divorce from a breadwinner; (c) a health episode that the household member rates as “extremely severe or potentially lethal”; (d) medical complications during a child birth; (e) damages to assets; (f) deaths of livestock; (g) crop losses; (h) unexpected large expenses (for the last four the minimum is a damage that the respondent estimates would cost more than MWK 10,000, or about $25 PPP, to replace); (i) a violent conflict, or (j) a conflict with an employer that results in complete non-payment of owed casual labor wages. In months without a shock, about 60% of the control group find themselves forced to skip at least one meal, while for the treatment group, this probability is reduced to around 35% (Appendix Figure A13). In periods with a severe shock, the probability of meal skipping rises by about 10 percentage points for either group.

My data also allow me to explore the impacts of shocks on the study households over time. I focus particularly on health shocks. I follow Kinnan et al. (2024) in designating the health shock that was associated with the highest health spending as the most severe health shock that a household experienced over the study period. Like these authors, if study participants mention symptom onset in a prior month to peak health spending, I designate that month as the timing of the health shock. I then use staggered differences-in-differences (Callaway and Sant’Anna, 2021), comparing households that do experience such a health shock to those that have not yet experienced one, thus allowing for the possibility that people who never report any health expenditures to be different from the rest of the sample in some fundamental way. Figure 5 summarizes the results. As a first sanity check, I plot health care spending as a dependent variable against the incidence of health shocks (top left). Health care costs rise sharply with the onset of a health shock. The treatment group is able to spend about twice as much on health care in response to the health episode. Households in my study appear to mostly absorb these health shocks in disutility. There is an uptick in severe coping behaviors like meal skipping in response to the shock. Their productive investments, on the other hand, do not appear to be affected and impacts on earnings are noisy, though the point estimates at times reach magnitudes similar to those of the consumption support that the NGO provides. Kinnan et al. (2024) show starker income scarring, which is at least in part because they use the most severe health shock reported for each household in the Townsend Thai data over fourteen years, while my households have been tracked for a little under two years.

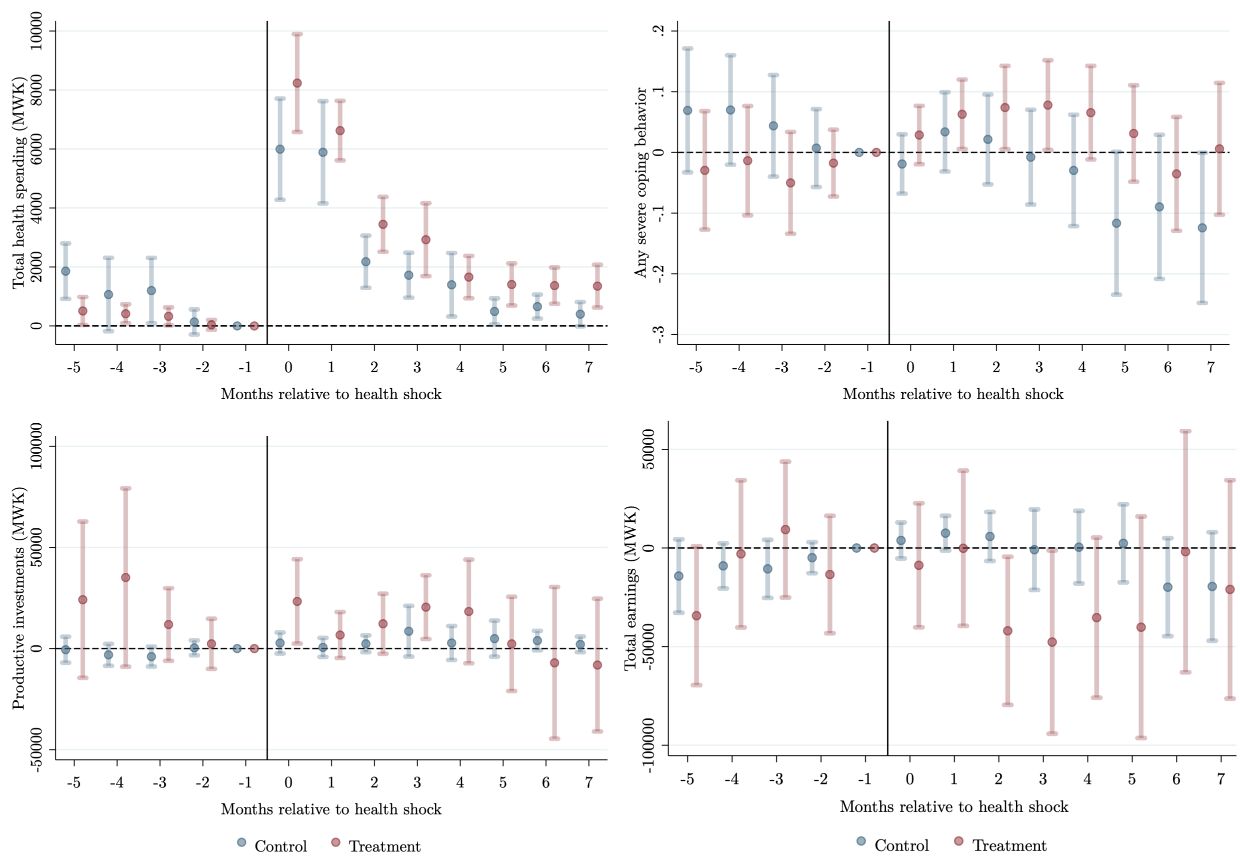


Figure 5: Impact of severe health shocks

Note: This figure shows the impacts of severe health shocks on households' livelihoods and finances.

The impact on shielding behavior, shock incidence, and resilience. Considering these negative impacts of shocks, preventive behavior is critical, and Figure 6 shows that the program successfully encourages households to adopt recommended behaviors. In what follows, I will first focus on three particular groups of preventive behavior that map directly onto some of the most common health episodes afflicting poor households here and across sub-Saharan Africa: (a) Bed nets and the removal of stagnant water are meant to reduce malaria incidence and while bed net use increases significantly, most control group households remove stagnant water absent the intervention. (b) Cooking in a separate room from where household members are sleeping is meant to reduce indoor air pollution and with it respiratory diseases, as would the provision of a blanket to children in the colder months. The program achieves these goals. (c) Handwashing stations with soap near a latrine with a lid/slap in addition to using borehole water and treating it with chlorine or boiling it before use are all intended to lower the incidence of diarrheal diseases. Program participants are significantly more likely to adhere to all of these practices.

Appendix Figure A7 shows that by contrast to the successful behavior change, the arrival rates of malaria, respiratory diseases, and diarrhea, which the coaching intervention tries to specifically address, are unaffected by the program. Apart from a spike in diarrheal diseases in the control group that the treatment group does not experience, none of the diseases are affected. More broadly, the UPG program does not lower the arrival rate of shocks in any particular domain nor of shocks overall (Appendix Figure A6). The figure plots timelines of shock arrival rates for various domains of life, such as health (which the coaching directly targeted), livestock, crops, or damages (which the coaching did not target, though the broader program might have allowed investments in pesticides, animal vaccinations, or dwelling fortifications). The time series of arrival rates of shocks in the treatment and control groups lie precisely on top of each other for the duration of the study. It is therefore not surprising that shock prevention cannot explain any of the poverty alleviation gains found earlier: Decomposing the reduction in severe coping behavior formally using a

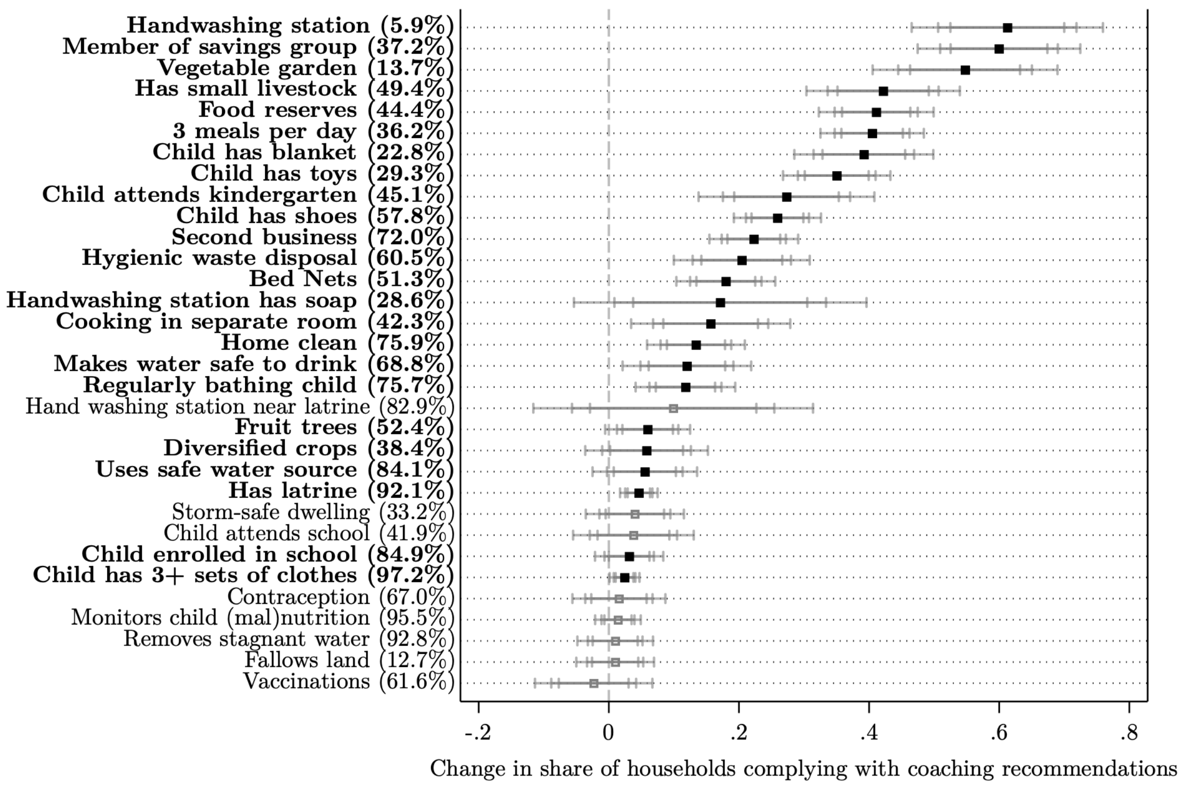


Figure 6: Compliance with recommended coaching behavior

Note: This figure displays the change at midline in he share of households complying with recommended behaviors to prevent adverse shocks or build resilience. The spikes denote 90, 95, and 99% confidence intervals. In parentheses, I note control group compliance levels.

Kitagawa-Oaxaca-Blinder decomposition, confirms that the gap is not driven by a reduction in the arrival rate of shocks (Appendix Table A6).

The reasons for the persistently high shock arrival rate are nuanced. I explore them in more detail for the case of malaria. Malawi has a high incidence of malaria with about 228 cases per 1,000 people per year nationwide,[14] though the study area – by the lakeside and down in the hotter Rift Valley – lies above the national average at around 350 per 1,000 people and year (Weiss et al., 2019). Using my high-frequency monitoring for the RCT sample households, selected for being among the poorest, the self-reported incidence is about 443 per 1,000 people and year.

Bed nets are widely accepted as effective a lowering malaria incidence. The seminal Cochrane meta-study finds reductions in malaria incidence of around 17% from insecticide-treated bed nets compared to no nets (Pryce et al., 2018). The point estimate for my study is smaller, at 9.3%, but within the Cochrane review's confidence interval. Several explanations present themselves for the smaller effects: First, the studies in Pryce et al. (2018) are predominantly from the 1990s when availability of bed nets in the general population was lower and the comparison group likely had no bed nets at all. In the subset of studies in which insecticide-treated nets are compared to non-treated nets, the drop in incidence is a more moderate 9%. This may be a more reasonable benchmark for my setting since 69% of individuals in the control group sleep under a bed net and in 51% of control group households, all household members sleep under one. Second, in 21 of the 23 Cochrane review's studies (accounting for 99.9% of the overall sample size) nets are distributed to all households in a community. My intervention treats only a subset of households, so unenrolled neighbors would likely act as a reservoir for malaria even under perfect bed net compliance by the treatment group. A similar argument can be made for the potential spillovers of diarrheal and respiratory diseases, which are directly communicable between people. The small and uncertain impacts of the graduation program on these diseases

[14] According to the World Health Organization in 2023, as retrieved from the World Bank DataBank.

therefore raise the question whether disease prevention is viable at the individual household level. The argument here is not that disease prevention is generally ineffective or even that it cannot complement poverty alleviation. But when it makes up 40% of the budget of an intervention that is targeted at the poorest subset of the population, it appears to realize none of the gains that have been demonstrated in studies that treat an entire population.

4.5 Further insights about the graduation program

Social empowerment and sentiment analysis. While it seems the ongoing coaching does not achieve its main goal of reducing the shock arrival rate or building resilience, it is possible that this program component empowers program participants. Respondents are asked at every meeting to reflect on the most important life events, current emotions, hopes and worries. I analyze the sentiment of these short reflections in Appendix Section B. I conclude that while the program raises overall sentiment consistently, this is driven by two separate forces: an increased sense of optimism while the NGO hands out consumption support, which fades quickly thereafter while the coaching continues; and a boost in internal locus of control – the idea that one determines one's own fate – that only emerges after respondents start their own businesses, when the coaching has already been ongoing for about ten months. While I cannot definitively conclude that the coaching had no additional effect on this feeling of empowerment, the timing of impacts lines up closely with other program components.

Savings groups and aggregation. The formation of savings groups is meant to sustain some of the graduation components beyond the initial intervention period. The seed capital grant or asset transfer of the graduation model is meant to overcome the widely documented difficulty poor households face in mobilizing large lump sums (Collins et al., 2009). Savings groups have been seen as an important component of graduation programs (Heil et al., 2024), even though their unbundled stand-alone impact appears to be small (Karlan et al., 2017; Banerjee et al., 2022). In this literature, savings groups have been hypothesized as allowing households to either form a buffer against shocks or serve as an opportunity to recreate the UPG's large one-off payment, potentially essential for further business growth. My team digitizes the ledger books of all savings groups started under the investigated program. These savings groups were structured as follows: All treatment group participants in a given cluster were encouraged to form a savings group. Compliance was practically universal at the beginning and even after one year, 96.4% of treated participants (compared to 96.4% in the control group) were members of a savings group. The group members received a training on the functioning of a savings group and were given equipment, such as a lock box with several keys and a ledger book. Members made weekly contributions (with a minimum mandatory contribution and then voluntary purchases of so-called shares). The savings group would remain active for one year (by default, but some groups adopted other durations). Throughout the lifetime of the savings group, members could ask for loans which they had to explain and pay back with an interest rate. At the end of a savings group's cycle, all savings were shared out according to the ownership of shares. Program implementers encouraged starting a second round of the savings group after share-outs and so far all groups did re-form for a second season.

Savings groups do appear to have an aggregation effect: The median contribution is MWK 2,400, while the median loan is an order of magnitude larger at MWK 20,000 and the ultimate share-outs are even larger at a median of MWK 79,300. Business investments are the most common purpose of loans, followed by bridging liquidity impasses for food purchases. At least in the first year of the savings groups for which I present data here, the aggregated amounts still fall short of reproducing the MWK 400,000 that was paid by the NGO as the main seed capital tranche.

Spillovers. Using specification (4), I find evidence of modest spillovers of the UPG onto non-treated peers. Rather than present the selected set of outcomes in which β_{S} is significant, I plot the t-statistic from equation (4) for all outcome variables that were collected during main-wave surveys[15] as a histogram in Appendix Figure A14. Outcomes are flipped as needed to make positive t-statistics "good" outcomes. Under the null hypothesis of no spillovers, we would expect a combined mass of 2.5% of t-statistics on either side of the interval [-1.96,1.96], but I find excess mass on both sides, more so for positive outcomes. I classify a majority of them as knowledge, attitudes, and aspiration variables, as well as economic variables. I find nearly no spillovers on health or early childhood development variables. The two notable outliers show that spillover control households are substantially less likely to own a phone or have access to mobile money, but that finding is mechanical as the program distributed phones only to households tracked at high frequency (pure control and all treated, but not spillover control households).

4.6 Cost-effectiveness

In low-numeracy and low-literacy contexts with predominantly informal labor, collecting high-quality income and consumption data is difficult in traditional RCT designs. When twelve or more months pass between main-wave surveys, researchers have the option to either ask for income and spending estimates retrospectively over the full study period (creating recall uncertainty) or rely on shorter recall periods like a month and extrapolate those earning and spending estimates (risking dramatic bias in contexts in which incomes are highly seasonal or vary over the intervention period). My data allow me to contrast both of these approaches with the dominant strategy of high-frequency short-recall data.

My cost-effectiveness analysis is guided by the following question: For how many months will income gains among the treatment group (relative to the control group) need to be sustained for the program to pay for itself? I divide the full program implementation cost by the number of served households (not including any research-specific costs like surveys or those that relate to the control group). In doing so, I conservatively include all

[15] The survey contains about 500 outcome variables, but in a first step, I remove all those that have a correlation coefficient with another variable of more than 0.5 in absolute value. I end up with roughly 350 variables.

community-level costs, even though their gains accrued to the control group as well.[16]

Respondents overestimate their own earnings over the full-year recall period – systematically more so among the control group. I estimate that the midline survey income gap needs to be sustained for a total of 63.8 months from the launch of the program for households to earn back what the program cost. If instead I use the one-month recall data,[17] the number of months required for households to earn back the program cost varies strongly by the timing of the survey relative to the agricultural season. The midline survey for cohort 1 was in the lean season and if I restricted myself to these data alone, I would estimate 71.8 months of required sustained income gains. Cohort 2 had its midline survey three months after harvest when farmers were still selling their harvested crops; the number of months for which that income gap would need be sustained plummets to 19.1 months. While we may reasonably expect researchers to understand this shortcoming of naively projecting out monthly income gain data over a full year, there is no obvious way that one would scale the estimate when the seasonality is not measured (Merfeld and Morduch, 2022). Turning to my high-frequency survey data, the naive approach would be to average the income gains (dark yellow area in Figure 3) over the full program duration and then project them out. Doing so, I obtain an estimate of 25.8 months of sustained income gains needed for the program to pay for itself. But Figure 3 suggests that this would be too conservative. The first six months see only very minute income gains, so it would be better to project out the higher post-grant income gains, which yield my final estimate of 25 months for households to earn back the program cost.

[16]Alternatively, one could assume either that contributions to the local childcare centers are not income enhacing or that they have no complementarities with the other program components. In that case, one could deduct those from the program costs and arrive at a higher cost-effectiveness estimate. Note also that I am taking the existence of the implementing NGO as a whole as given, meaning I am only including project-specific overhead. This is roughly in line with how my implementing partner is funded, covering its overall functioning from smaller unrestricted charitable giving and its programs through dedicated restricted funding. This cost-effectiveness calculation should be thought of as an estimate for the restricted funding.

[17]The midline survey did not collect one-month recall data. Instead, I use the high-frequency survey most closely adjacent to the midline survey and scale the period that it covered since the last such survey to 30 days because these modules were not always scheduled exactly monthly, but were collected cumulatively.

Optimal program design. The graduation program is a costly intervention with per-household costs ranging from $400 to $1,700, depending on implementer, program components, and location (World Bank, 2021). The studied program falls well within this range.[18] Leveraging my intimate familiarity with this program's implementation, I compute that about 39% of its costs (at the current number of enrolled households) is spent on coaching visits. Without the need for ongoing one-on-one coaching, implementers would only need to be hired for about six months rather than two years, and several key pieces of equipment could be recycled between cohorts rather than reaching the end of their productive life within one cohort. After taking into account how various cost components scale, I conclude that the number of served households could have been doubled under the same budget envelope if, as I suggest, the one-on-one coaching had no meaningful impact on participants. This is all the more an appealing option as the shocks that the coaching seeks to prevent do not appear to have permanently scarring effects on households' livelihoods.

5 Discussion

The graduation model is one of the most promising tools for alleviating extreme poverty available in the development economics toolbox. Understanding how it operates and what makes it successful is just as critical as evaluating which of its many costly components might not yield the desired effects. A number of conclusions emerge from this study.

First, the three main livelihood components of the program appear effective and strongly complementary: Households reduce their reliance on severe coping behaviors like meal skipping dramatically and promptly upon enrollment, using the UCTs before generating any additional income of their own. The cash grant allows the launch of new businesses and the new income more than fully replaces the consumption support immediately after. At the one-year mark, the program appears to behave very similarly to the well-studied BRAC

[18] Exact cost figures are under review by the implementing partner and will be included in a future version of this paper.

UPG, despite the notable design change of a cash grant in lieu of livestock. While only time will tell whether this remains true, this suggests that my results are generalizable. Unlike other leading social assistance programs like pure UCTs, UPGs have a track record of lasting impact.

Second, despite the fungibility of cash, the marginal propensity to invest the seed capital productively was very high in this study. I provide suggestive evidence that the prior six months of consumption support along with labeling the seed capital for business contributed to this. This places my findings in an interesting tension a key finding of frontier UCT research where Egger et al. (2022) find an MPC of nearly 95% over two years. Over the approximately twelve months that I have been able to follow the households in my study thus far, the MPC is below 20%. Egger et al. (2022) argue that part of what underpins their large multiplier estimates of UCTs is a high MPC. A high MPC though would also explain the absence of lasting impacts in most UCT programs to extremely poor populations. Meanwhile, the consumption support in the UPG (in the presence of apparently unavoidable shocks), seems to have protected the seed capital from being consumed. This offers a tantalizing explanation for the long-run impacts that UPGs have produced and that have eluded the literature on pure cash transfers.

Third, it is surprising that households in the studied UPG appear to be unable to shield themselves from health or livelihood shocks in any way, despite receiving one of the most comprehensive and costly poverty alleviation programs in existence. I estimate that around 40% of this program's budget goes to one-on-one coaching. After taking various fixed costs into account, I find that the program could have been extended to twice as many beneficiaries without this program component. Other studies (Banerjee et al., 2022; Beam et al., 2025) are more directly designed to speak to this aspect, but my high frequency data suggest that households adopt the recommended behaviors already after the group training. Further, the timing of the coaching itself does not coincide with increased adoption, and some behaviors are dropped again regardless of coaching. While there is a strong sense in the implementation community that coaching matters, the present study does not strengthen the case for its cost-effectiveness. It may be too early to judge the impacts of coaching, but most of the encouraged behaviors, if effective and adhered to, should show results quickly, especially hygiene and bed nets. Despite demonstrated changes in behaviors from the bundle as a whole, none of these translate into lower shock arrival rates. The shocks that the coaching seeks to avoid have modest impacts on consumption but do not scar livelihoods lastingly.

In conclusion, I shed light on the mechanisms that make UPGs one of the most promising poverty alleviation tools available tested to date. I validate a number of hypotheses about their inner working, while failing to find evidence for others. My results on the propensity to consume and invest funds received in this program or subsequently earned offer a tentative explanation for the stark contrast in impact longevity between the UCT and UPG literature.

Appendix of chapter 1

A Additional figures and tables

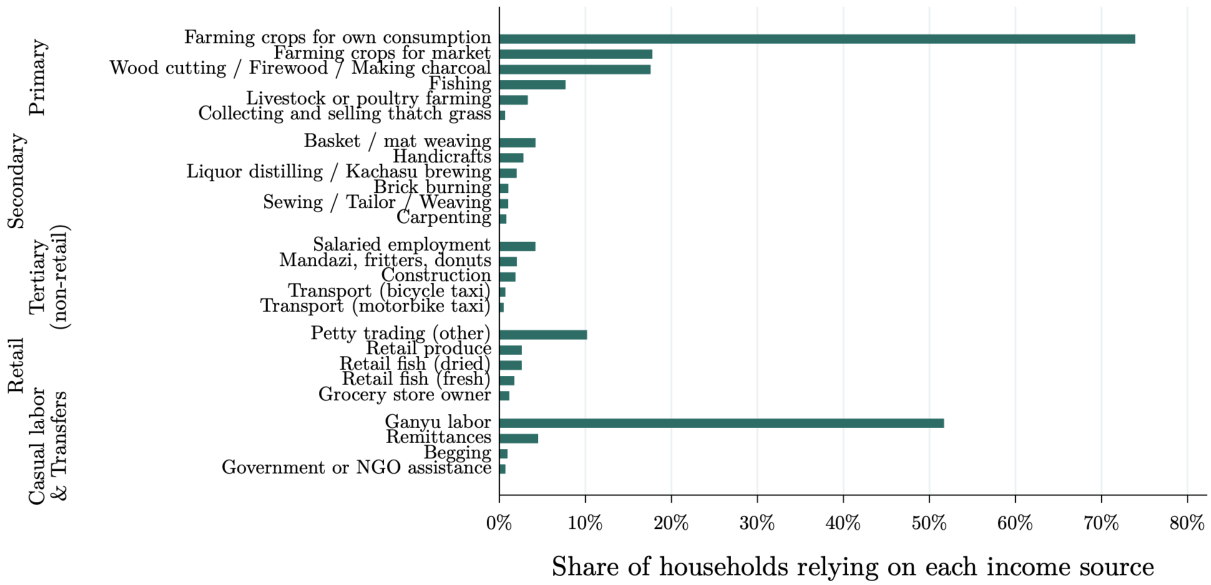


Figure A1: Most common income sources in the study area (census)

Note: This figure describes the income sources of the population as stated during the area-wide census. People are predominantly farmers or casual "'ganyu" laborers. Outside of the primary sector, retail is another common income source.

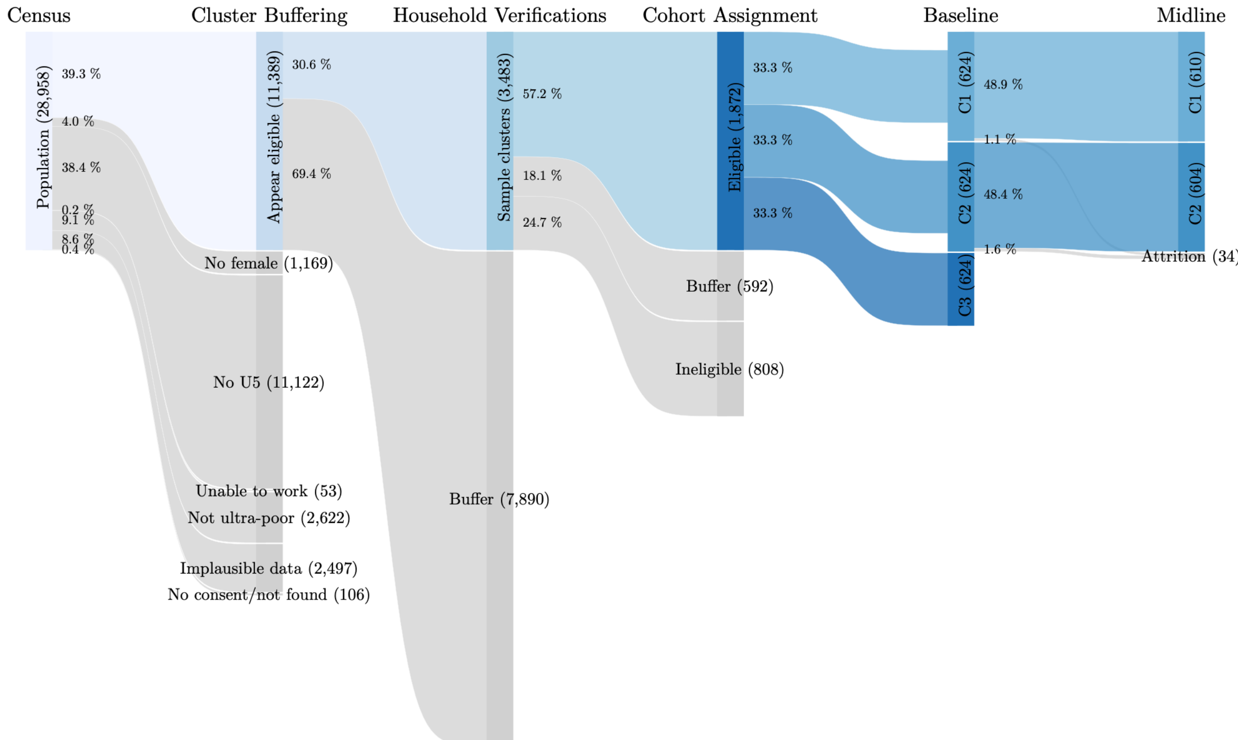


Figure A2: Sampling funnel

Note: This figure describes how the sample is selected from the census population. Starting from the census population of about 29,000 households, about 60% were deemed ineligible, mostly for lack of a child under age five. Among apparently eligible households, 271 clusters of 40-50 households were formed by proximity using *k*-means clustering. A second custom algorithm then selected 72 of these clusters in such a way as to maximize the distance between the nearest two clusters, in order to minimize spillover concerns (cf. Appendix Figure A3). The other clusters serve as buffer zones. Among the selected clusters, household verifications were conducted that resulted in some households being deemed ineligible under the same criteria as in the census, while selecting the 26 eligible households closest to the cluster centroid and excluding the remainder as additional buffer households on the periphery of the clusters. The 72 clusters were grouped into triplets by proximity and assigned an implementing staff. They were also assigned at random to three cohorts to undergo the intervention at different times in the agricultural season. The full sample was surveyed at baseline and the graph shows attrition in the subsequent survey rounds.

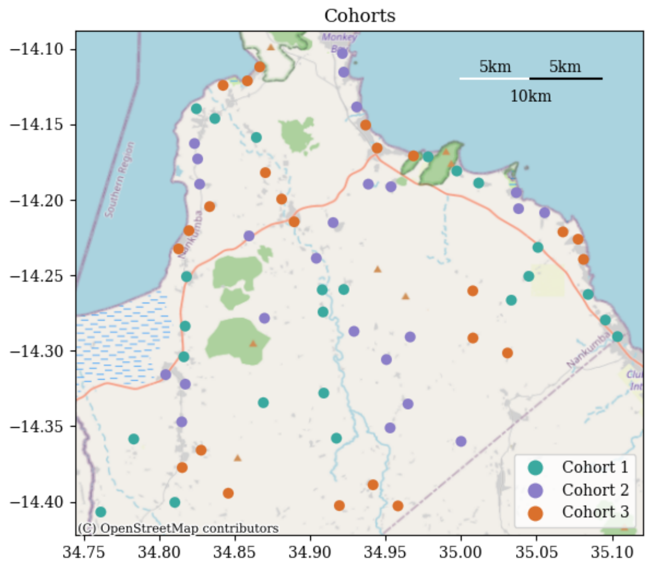


Figure A3: Clusters by cohort

Note: This map shows the 72 study clusters of 26 households each that make up the study sample, color-coded by the cohort they are assigned to.

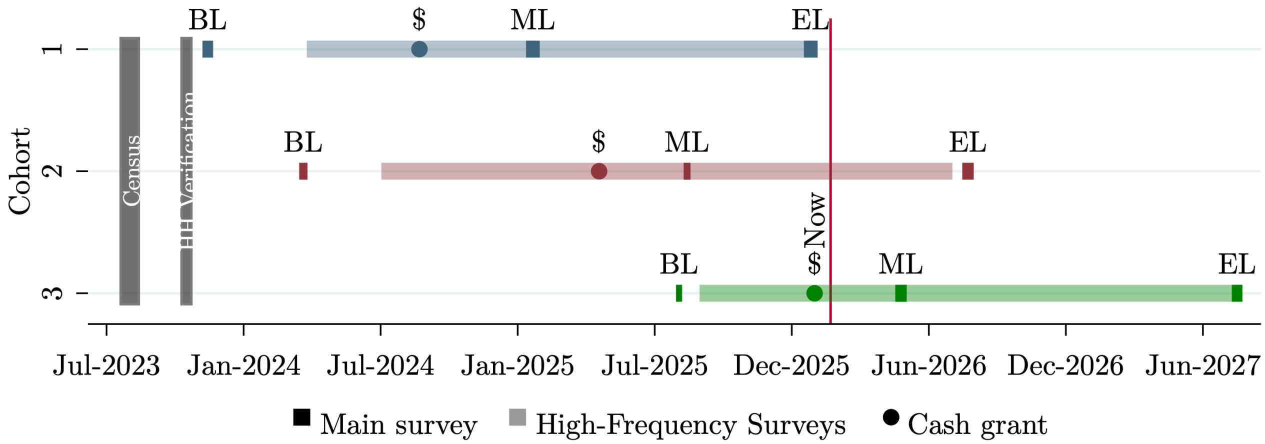


Figure A4: Study timeline

Note: This figure describes the study timeline. A census of the study area was conducted in August 2023. Household eligibility verifications followed in October 2023. The baseline survey for cohort 1 was conducted in December 2023 and high-frequency surveys in this cohort started in April 2024. The large one-off cash grant was paid out in August 2024. The midline survey was collected in January 2025, followed a year later by the endline survey.

Table A1: Study population

	(1) Count
HOUSEHOLDS	
Number of surveys conducted	28,853
Female primary adult present	27,684
U5 present	16,882
No children present	3,254
INDIVIDUALS	
Population	138,548
Adults	65,162
Children	73,386
U5s	20,998
Number of villages	162
Number of CBOs	25
Number of ADCs	5

Census of all households in Mangochi district's subdistrict ("Traditional Authority") Nankumba. The urban center Monkey Bay and the tourist destination Cape Maclear were omitted from this census as well as an area in Southwestern Nankumba where another NGO is implementing a livelihood program. Area Development Committees (ADCs) are administrative subdivisions of a subdistrict, while Community-Based Organizations (CBOs) are associations of several villages organizing development initiatives or providing local public goods together.

Table A2: Poverty

	(1) Percent
Skip meal (past two weeks)	76.8
Female illiterate	37.7
Thatch roof	64.5
No furniture	80.3
No phone	51.9
No bedding	15.7
<2.15 PPP 2017 USD (self report)	92.7
<2.15 PPP 2017 USD (PMT)	85.1
# children / adult (if > 0)	1.5
# income sources	1.5

Poverty measures from the census in Table A1.

Table A3: Local GDP

	(1) MWK	(2) USD	(3) USD BM	(4) USD PPP
GDP (millions)	12,707	12.07	7.06	35.56
Subsistence production (millions)	8,614	8.18	4.79	24.11
GDP/capita	91,717	87.10	50.95	256.69
Subsistence production/capita	62,171	59.04	34.54	174.00
GDP/adult	195,009	185.19	108.34	545.79
Subsistence production/adult	132,189	125.54	73.44	369.97

GDP estimate for the study area from August 2022 to July 2023. The estimate in the first row is based on households' self-reported income over the past 12 months. The second row is an estimate by respondents of the value of the subsistence agricultural yield that was not marketed. Household work is not included. The following rows divide these numbers by the overall and adult population, respectively. Column (1) shows the estimates in local currency (Malawi Kwacha). In column (2), I express these amounts in USD using the official exchange rate of 1,153 MWK/USD that was in effect at the time of the census. This rate is best suited for benchmarking the size of the intervention against local GDP as it is what the NGO used to convert the funds. For column (3), I use the black market exchange rate at the time for conversion to USD to give a sense of what this GDP could buy in the US. For reference, Malawi's GDP per capita at this rate was USD 389 in 2023. In column (4), I convert the GDP using a weighted average of the World Bank's GDP purchasing power parity conversion factor for 2022 and 2023 of 357.3 MWK/USD. For reference, Malawi's GDP per capita at this conversion rate was USD 1,829 in 2023. These figures illustrate the equivalent of what can be bought at local prices with this production value.

Table A4: Balance

	(1) Control Mean (SD)	(2) Treatment (SE)
Household and participant characteristics		
Household size	5.4	-.1175
	(1.8)	(.09726)
Age	32	-.5641
	(8.4)	(.4244)
Years of education	5.7	-.1218
	(2.9)	(.1906)
Literate	.69	.005342
	(.46)	(.02665)
Livelihood and assets		
Hours worked (7 days)	22	-.3643
	(18)	(.7828)
HH income (12 months, MWK)	475,629	11,315
	(480,571)	(47,515)
Total savings, assets, livestock, minus debt (MWK)	256,874	-17,471
	(365,607)	(24,353)
Largest purchase (12 months, MWK)	38,607	-2,816
	(121,528)	(6,344)
Poverty and food security		
Thatch / tarp / no roof	.75	.02137
	(.44)	(.02314)
Food insecurity score (0-8)	7.5	.04915
	(1.3)	(.0641)
Health and non-health shocks		
Health episodes (12 months)	16	-.578
	(13)	(.9197)
Non-health shocks (12 months)	21	-.3771
	(18)	(.8531)
Observations		1,872
F		0.58
$p(F = 0)$		.85

This table describes the baseline balance. The main panel reports the control group means and coefficients of the treatment indicator in a regression predicting each baseline outcome variable. At the bottom of the table, an F statistic is reported for a regression in which all outcomes jointly predict the treatment. Standard errors are clustered at the household cluster level and reported in parentheses. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Midline economic outcomes

	(1) Control mean (SD)	(2) Treatment Effect (SE)	(3) N (BL)
Does business	.6851	.2216***	1214
	(.4649)	(.02781)	✓
		[0]	
Casual labor, charity	.804	-.1775***	1214
	(.3973)	(.03445)	✓
		[0]	
Farming, fishing, forestry	.6131	-.08489**	1214
	(.4875)	(.03684)	✓
		[.02997]	
Manufacturing	.1424	-.00317	1214
	(.3497)	(.01904)	✓
		[.8771]	
Non-retail services	.1457	.1099***	1214
	(.3531)	(.01865)	✓
		[0]	
Retail	.1977	.4375***	1214
	(.3986)	(.03544)	✓
		[0]	
HH income (12 months, MWK, winsorized)	664,083	411,049***	1213
	(647,484)	(58,886)	✓
		[0]	
Total assets (MWK)	391,459	185,231***	1214
	(473,494)	(29,801)	✓
		[0]	
Business assets (MWK)	25,543	75,314***	1214
	(114,443)	(17,590)	✓
		[0]	
Agricultural Assets (MWK)	14,960	5,062***	1214
	(19,353)	(1,364)	✓
		[0]	
Personal assets (MWK)	341,534	108,968***	1214
	(387,610)	(20,457)	✓
		[0]	
Any livestock	.4941	.4216***	1214
	(.5004)	(.04001)	✓
		[0]	
Livestock value (MWK)	41,333	72,140***	1214
	(91,642)	(10,446)	✓
		[0]	
Any savings	.3166	.4943***	1214
	(.4655)	(.03558)	✓
		[0]	
Total savings (MWK)	12,791	59,465***	1214
	(32,717)	(8,941)	✓
		[0]	

Notes: The last column indicates the number of observations. All regressions are implemented as an ANCOVA specification, controlling for strata fixed effects. A ✓ indicates that the outcome variable is available for all observations at baseline. If it was not collected at baseline, a ✕ is shown, and if it is partially available, missing observations are set to zero and dummied out (✓). Where available, regressions control for baseline levels of the outcome. Standard errors are clustered at the strata level. Fisher (1935) exact test p-values from 1000 simulated alternative treatment assignments are shown in brackets. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

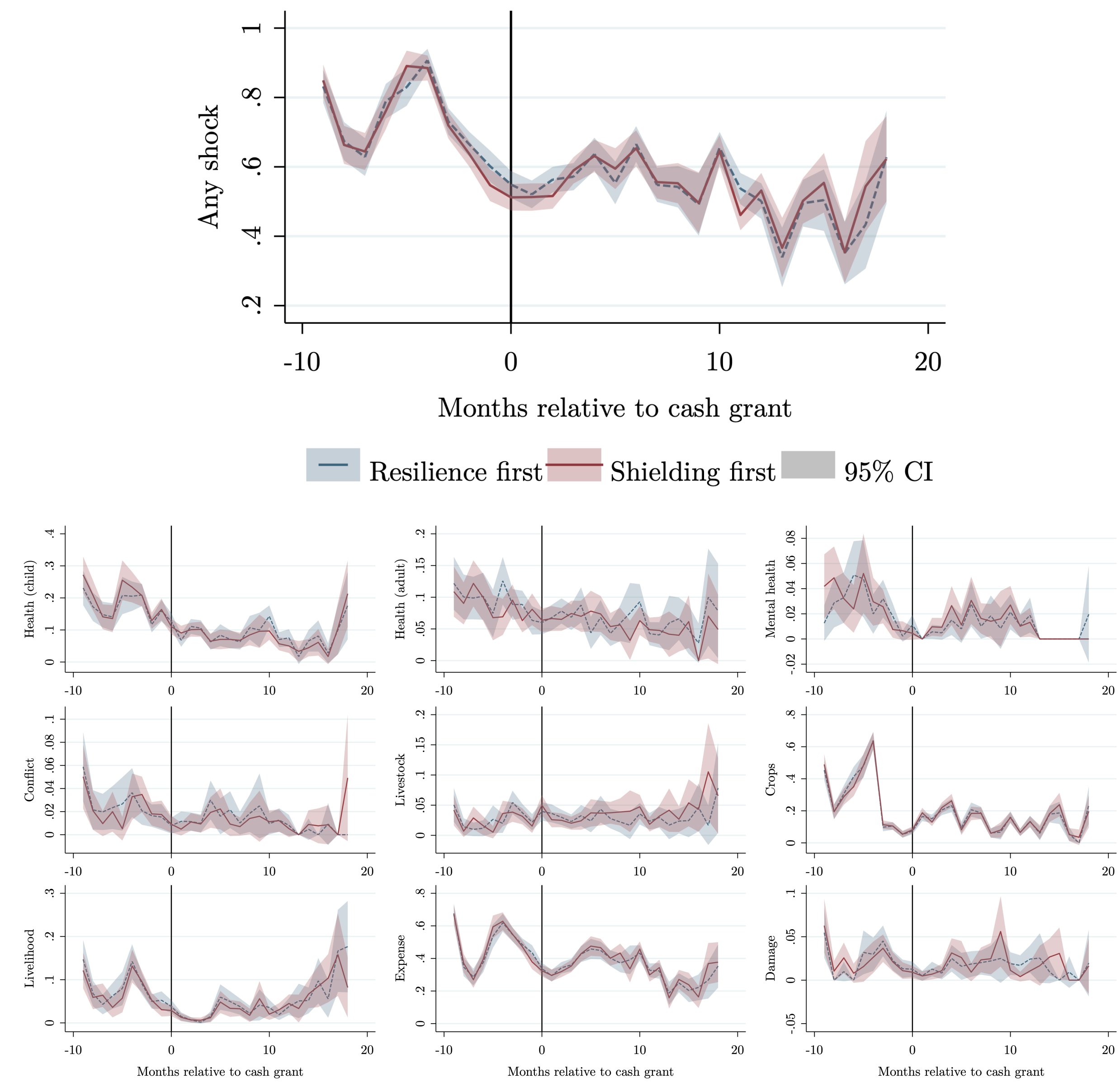


Figure A5: Arrival rate of shocks by ordering of the coaching intervention

Note: This figure plots shock arrival rates for the treatment group separately by whether they were first coached on shielding behaviors that prevent shocks or resilience that improves ability of coping with a shock conditional on it occurring, relative to the timing of the cash grant. The ordering of coaching did not impact the arrival rate of shocks.

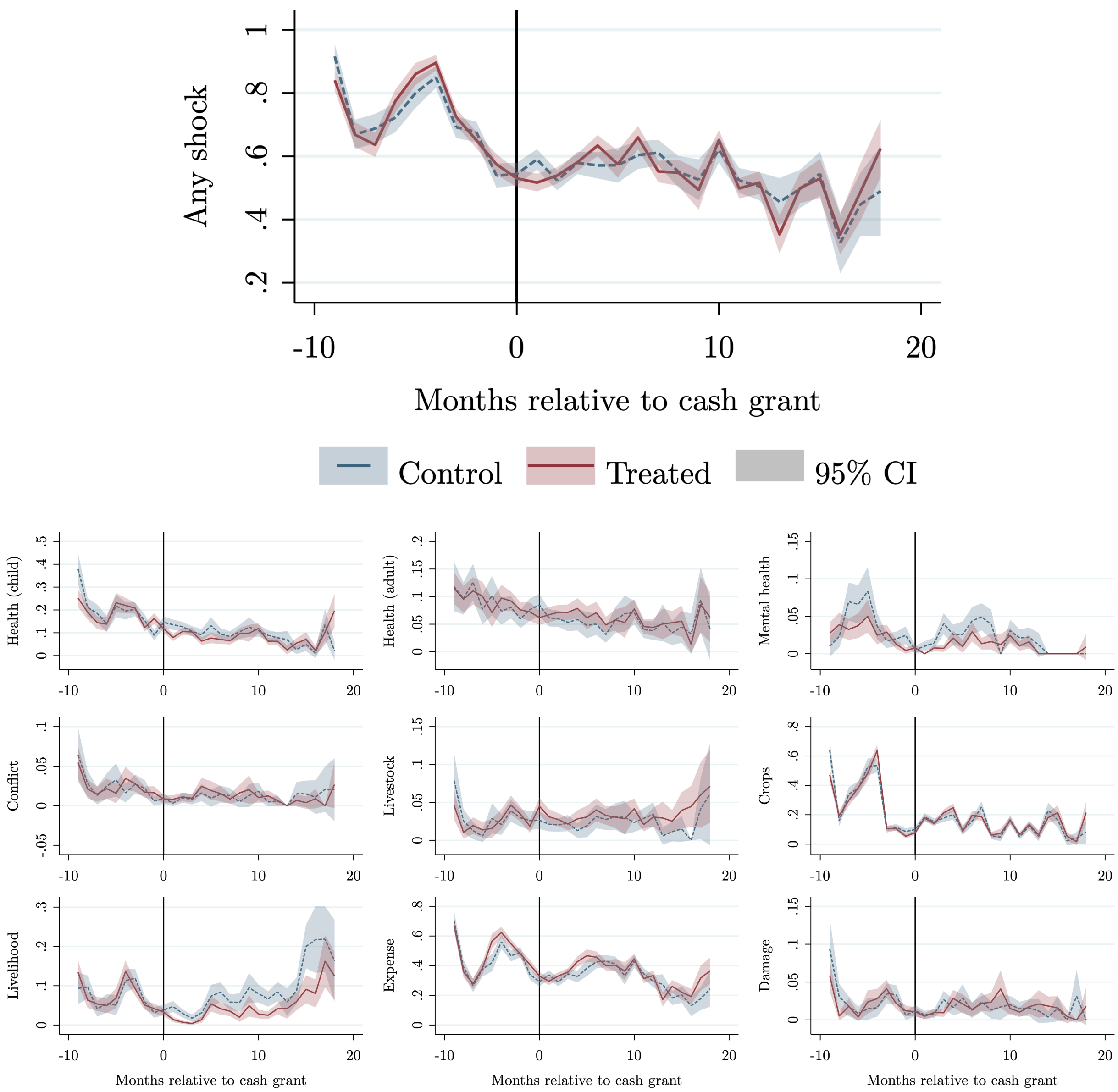


Figure A6: Arrival rate of shocks by treatment arm

Note: This figure plots shock arrival rates for treatment and control group relative to the timing of the cash grant. The program does not lower shock arrival rates.

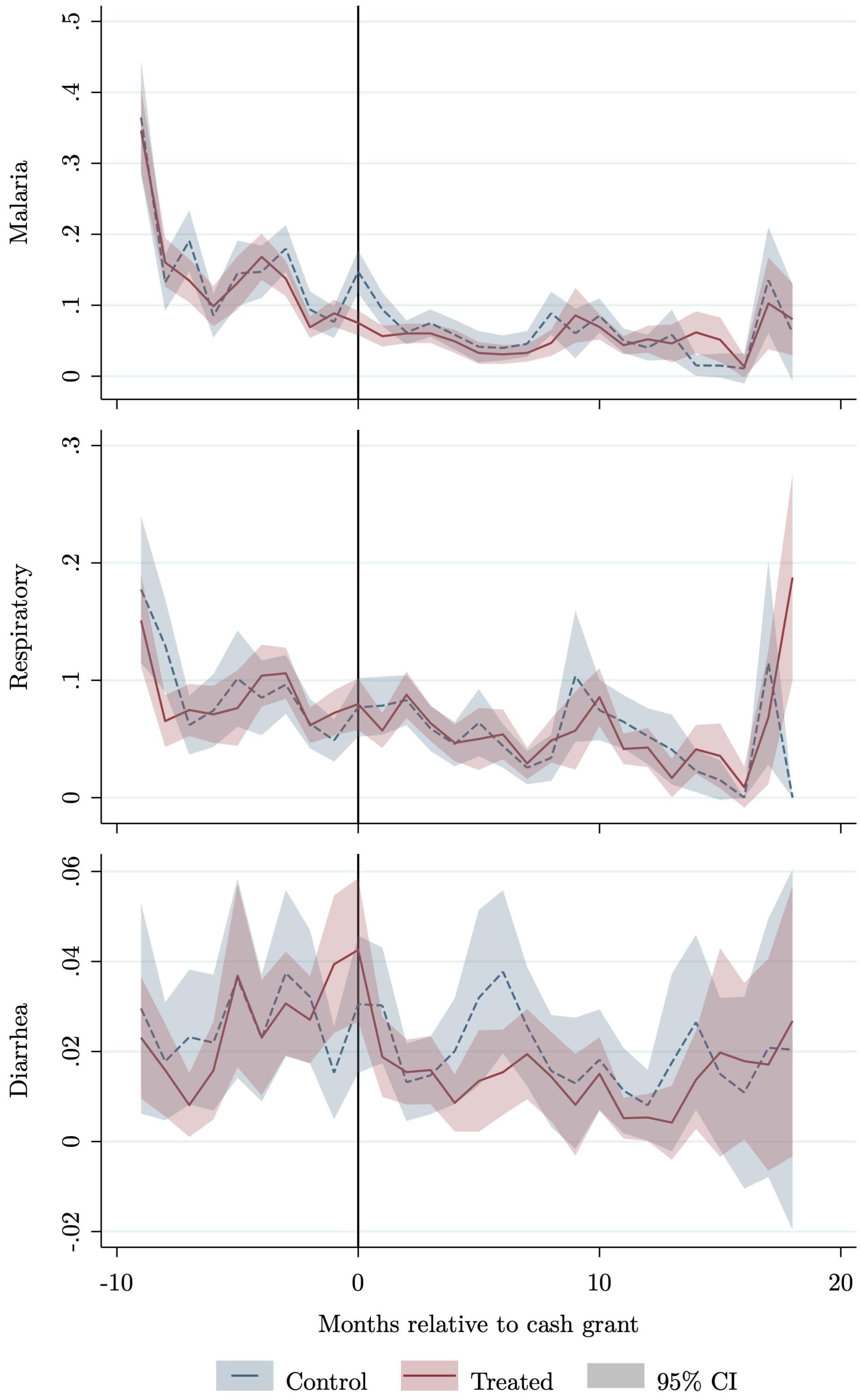


Figure A7: Effects on arrival rate of health shocks

Note: This figure shows the overall effects of the program on the arrival rate of malaria, respiratory diseases, and diarrhea – three diseases that the coaching intervention tries to specifically address.

Table A6: Kitagawa-Oaxaca-Blinder decomposition: Shielding vs resilience

	Coping severity (WFP, 0-3)	
	All shocks	Severe shocks
Overall treatment effect		
Control group	1.268***	1.268***
	(0.113)	(0.119)
Treatment group	0.629***	0.629***
	(0.0835)	(0.0871)
Difference	0.639***	0.639***
	(0.140)	(0.147)
Decomposition		
Shielding	0.0502	0.00467
	(0.0435)	(0.0247)
Resilience	0.631***	0.625***
	(0.134)	(0.145)
Interaction	-0.0418	0.00921
	(0.0632)	(0.0253)
Observations	1514	1514

Notes: This table presents a Kitagawa-Oaxaca-Blinder decomposition of the effect of treatment onto the observed severity of household coping behavior. It decomposes the difference into a component mediated by shock arrival rate (shielding) as well as a direct treatment effect (resilience). Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

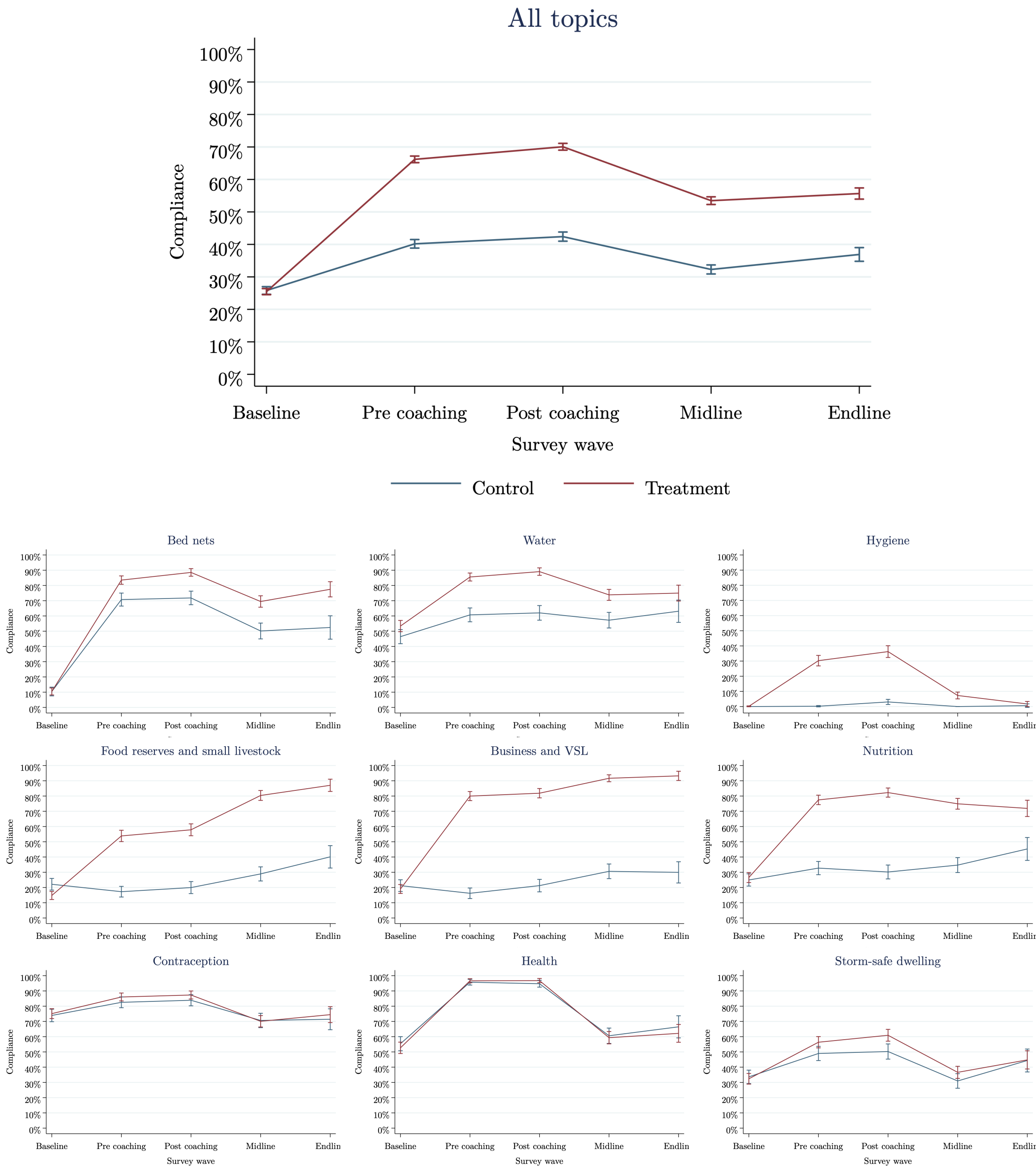


Figure A8: Coping compliance

Note: This figure shows participant compliance with recommended behaviors from the one-on-one coaching curriculum. Recommended behaviors inlude the use of bed nets, measures to make water safe for drinking, general hygiene recommendations, building up food reserves and keeping small livestock around the house, running a non-agricultural business and joining a village savings and loans group (VSL), a diversified nutrition for children, contraception, health check-ups and vaccination, as well as fortifying the dwelling against storms.

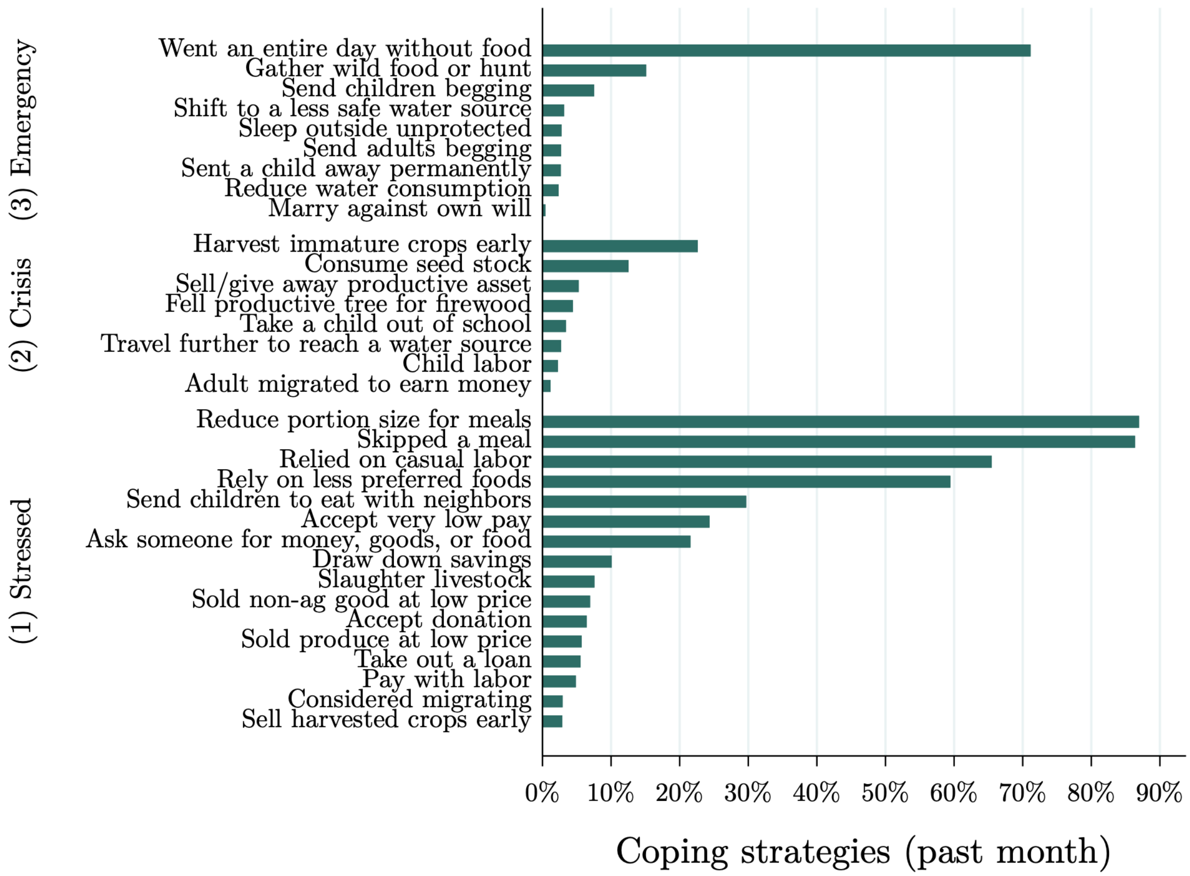


Figure A9: Coping behaviors practiced at baseline

Note: This figure shows baseline prevalence of a comprehensive set of severe coping behaviors as classified by the World Food Program.

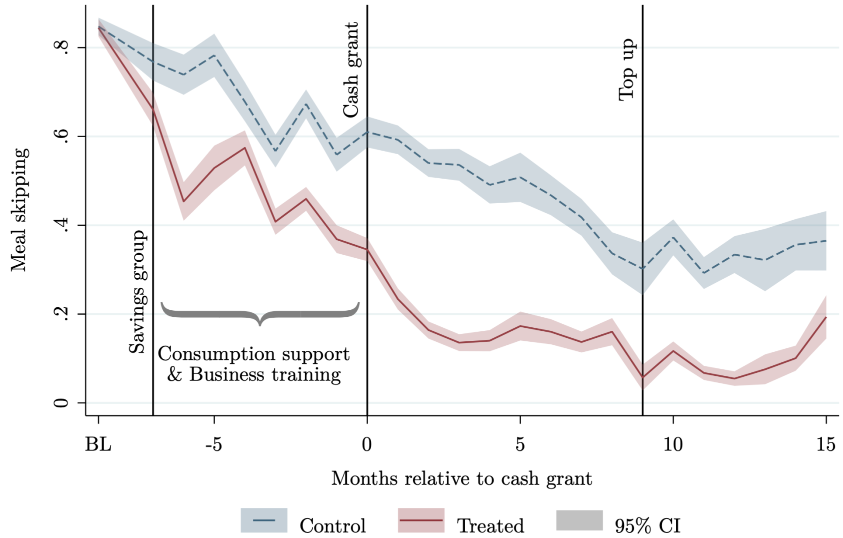


Figure A10: Effects on probability of meal skipping

Note: This figure shows the effects of the program on households' probability of having skipped a meal recently.

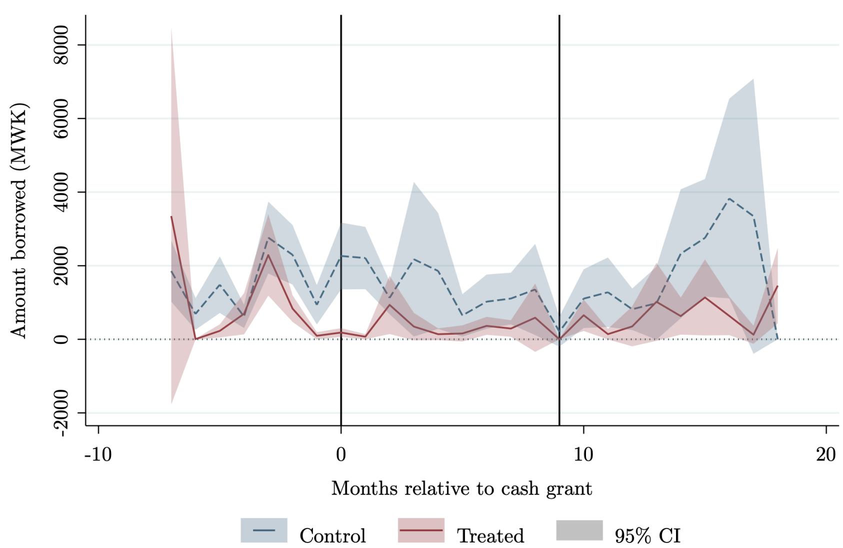


Figure A11: Effects on propensity to borrow

Note: This figure shows the effects of the program on households' probability of borrowing money. The large span in month -7 is caused by a single treated household taking out a very large loan.

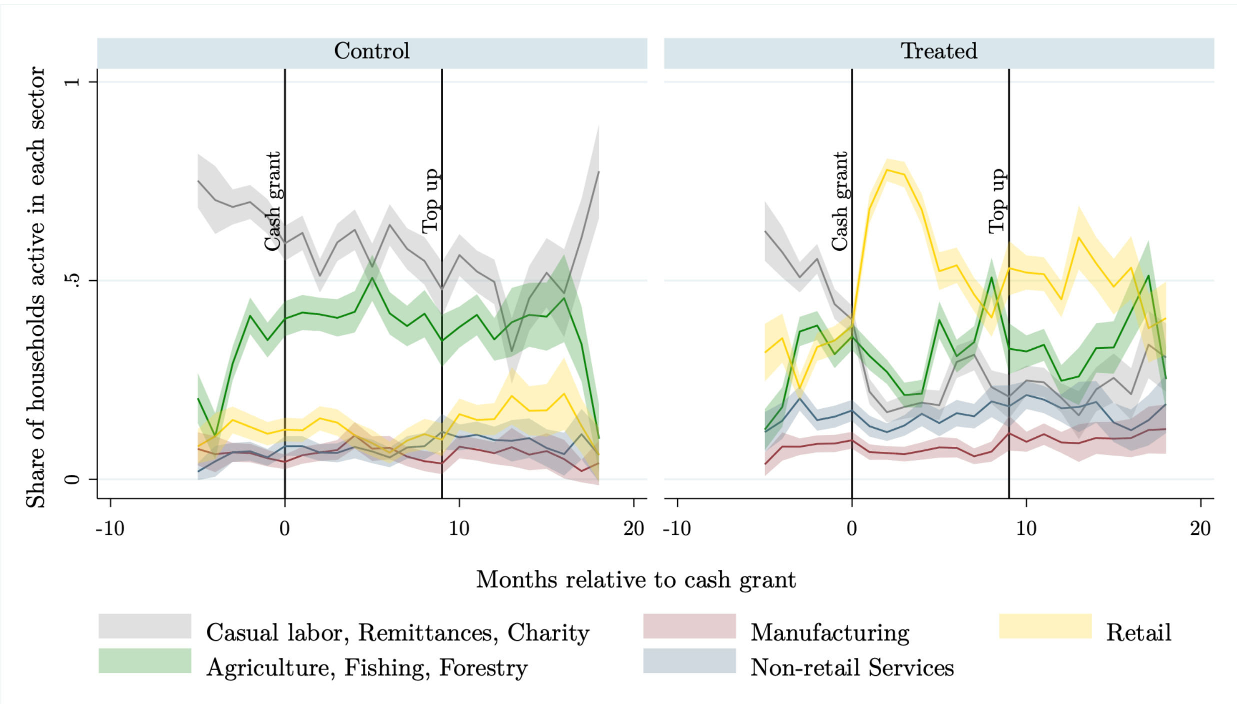


Figure A12: Livelihood churn at high frequency

Note: This figure shows which livelihoods households rely on in the control (left) and treatment group (right) over the duration of the study. The control group relies primarily on primary sector livelihoods (farming, fishing, and cutting firewood or making charcoal) as well as casual labor, remittances, and begging. The treatment group reduce their reliance on casual labor gradually in favor of retail, though even non-retail services gain popularity. Agriculture remains important and shows strong seasonality, while retail businesses are paused during agricultural peak seasons.

Table A7: Adoption of recommended behaviors by coaching intervention ordering

	(1) Control mean (SD)	(2) Treatment Effect	(3) Shielding Coaching first	(4) Interaction	(5) N (BL)
Bed Nets	.5126	.202***	.04092	-.04521	1214
	(.5003)	(.03276)	(.03468)	(.03805)	✓
Uses safe water source	.8409	.06157*	.02916	-.01315	1212
	(.3661)	(.0324)	(.03367)	(.03115)	✓
Makes water safe to drink	.6884	.1202***	.02312	-.0009525	1214
	(.4635)	(.03595)	(.04611)	(.05388)	✓
Handwashing station	.0586	.6188***	.003209	-.0126	1214
	(.2351)	(.04904)	(.0158)	(.03258)	✓
Hand washing station near latrine	.8286	.03422	-.1182	.1237	450
	(.3824)	(.08162)	(.1176)	(.1283)	(✓)
Handwashing station has soap	.2857	.1773**	.03889	-.007384	450
	(.4583)	(.06628)	(.1651)	(.1498)	(✓)
Has latrine	.9206	.0553***	-.002206	-.018	1175
	(.2707)	(.01485)	(.01672)	(.01881)	(✓)
Hygienic waste disposal	.6047	.1616***	-.07357	.0868	1214
	(.4893)	(.03883)	(.04406)	(.06035)	✓
Removes stagnant water	.928	.009191	.02062	.0006954	1214
	(.2587)	(.02618)	(.02419)	(.03245)	✓
Regularly bathing child	.7571	.06982**	-.04436	.09643***	1214
	(.4292)	(.02871)	(.02928)	(.02918)	✓
Home clean	.759	.1385***	.01881	-.009675	653
	(.4284)	(.03252)	(.04929)	(.06299)	✓
Cooking in separate room	.4228	.1512***	.04568	.009372	1213
	(.4944)	(.04298)	(.04331)	(.04588)	✓
Vaccinations	.6164	-.03389	.01586	.02068	1214
	(.4867)	(.04722)	(.03149)	(.0602)	✓
Storm-safe dwelling	.3317	.05495	.005862	-.02975	1214
	(.4712)	(.03633)	(.02625)	(.04554)	✓
Food reserves	.4439	.4241***	.01873	-.02649	1214
	(.4973)	(.03523)	(.04942)	(.05105)	✓
Has small livestock	.4941	.4079***	-.001219	.02723	1214
	(.5004)	(.04315)	(.02962)	(.03755)	✓
Contraception	.67	.01943	-.05274*	-.006288	1214
	(.4706)	(.033)	(.02629)	(.04355)	✓
Child has shoes	.5779	.2609***	.0718*	-.005184	1214
	(.4943)	(.03379)	(.03458)	(.04829)	✓
Child has blanket	.2278	.3596***	-.02261	.06484	1214
	(.4198)	(.043)	(.02792)	(.04112)	✓
Child has 3+ sets of clothes	.9715	.02073*	-.01107	.007106	1214
	(.1665)	(.01097)	(.01485)	(.01707)	✓
Child has toys	.2931	.3596***	.01514	-.01898	1214
	(.4556)	(.03263)	(.02693)	(.0403)	✓
Child attends kindergarten	.4506	.2514***	-.01842	.04354	1214
	(.498)	(.0606)	(.05625)	(.06331)	✓
Child enrolled in school	.8492	.008558	-.02688	.04575	1214
	(.3581)	(.02321)	(.02552)	(.03691)	✓
Child attends school	.4188	.02389	-.03489	.0285	1214
	(.4938)	(.04266)	(.03228)	(.05201)	✓
Second business	.7203	.2119***	-.00736	.02183	1214
	(.4492)	(.02636)	(.0253)	(.0229)	✓
Member of savings group	.3719	.6103***	.02784	-.02192	1214
	(.4837)	(.05002)	(.03921)	(.03909)	✓
Vegetable garden	.1374	.5325***	.004181	.02966	1214
	(.3445)	(.062)	(.0271)	(.04611)	✓
Diversified crops	.3836	.02788	-.03331	.06074	1214
	(.4867)	(.03597)	(.0367)	(.04411)	✓
Fallows land	.1273	.01931	.04399	-.02016	1214
	(.3336)	(.02823)	(.02567)	(.03315)	✓
Fruit trees	.5243	.07094*	.07175***	-.02423	1214
	(.4998)	(.03626)	(.01855)	(.04565)	✓
Monitors child (mal)nutrition	.9548	.005078	-.01644	.01831	1214
	(.208)	(.0176)	(.01619)	(.01976)	✓
3 meals per day	.3618	.4272***	.08072***	-.04671	1214
	(.4809)	(.03557)	(.02624)	(.03586)	✓

Notes: The last column indicates the number of observations. All regressions are implemented as an ANCOVA specification, controlling for strata fixed effects. A ✓ indicates that the outcome variable is available for all observations at baseline. If it was not collected at baseline, a ✗ is shown, and if it is partially available, missing observations are set to zero and dummied out (✓). Where available, regressions control for baseline levels of the outcome. Standard errors are clustered at the strata level. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

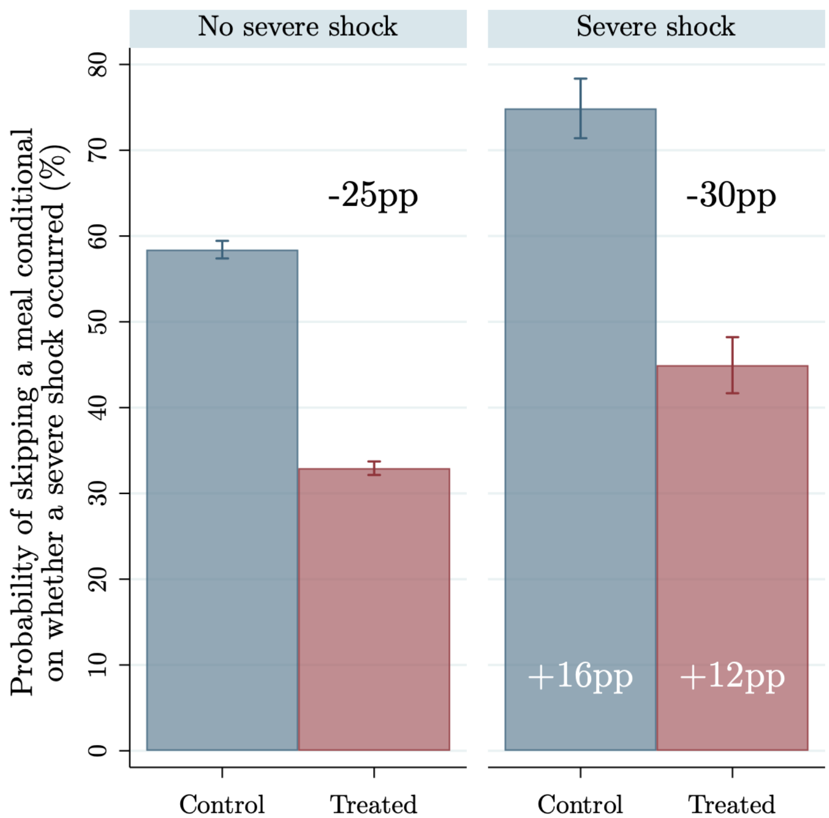


Figure A13: Probability of meal skipping and shocks

Note: This figure displays the probability of households skipping a meal, conditional on whether they experienced a health or livelihood shock in that period.

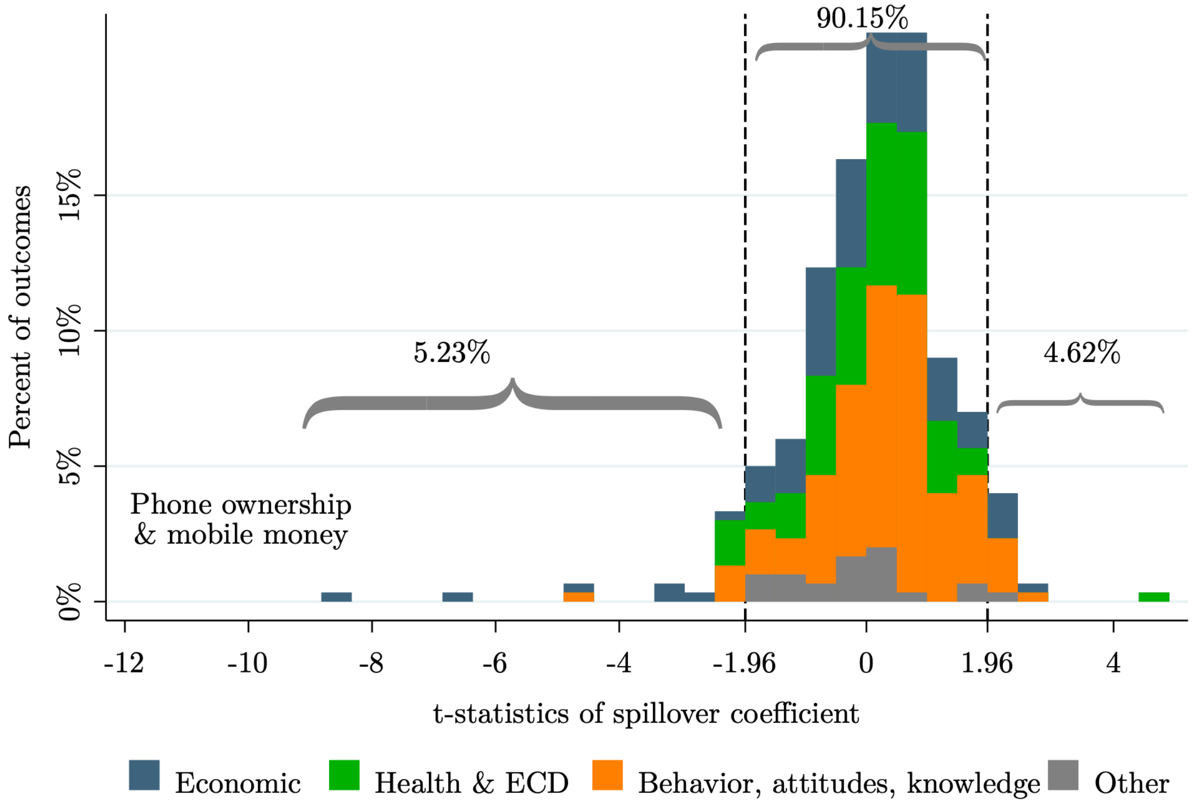


Figure A14: Spillover coefficient t-statistics at midline

Note: This figure shows the t-statistics on coefficient β_S in regression specification (4) for all outcome variables collected at midline. Variables that correlate strongly with other outcomes are removed. All variables are "flipped" as needed to ensure that positive t-statistics mean "good" outcomes. Under the null hypothesis of no spillovers, we would expect about 2.5% of the mass to lie on either side of the interval [-1.96,1.96].

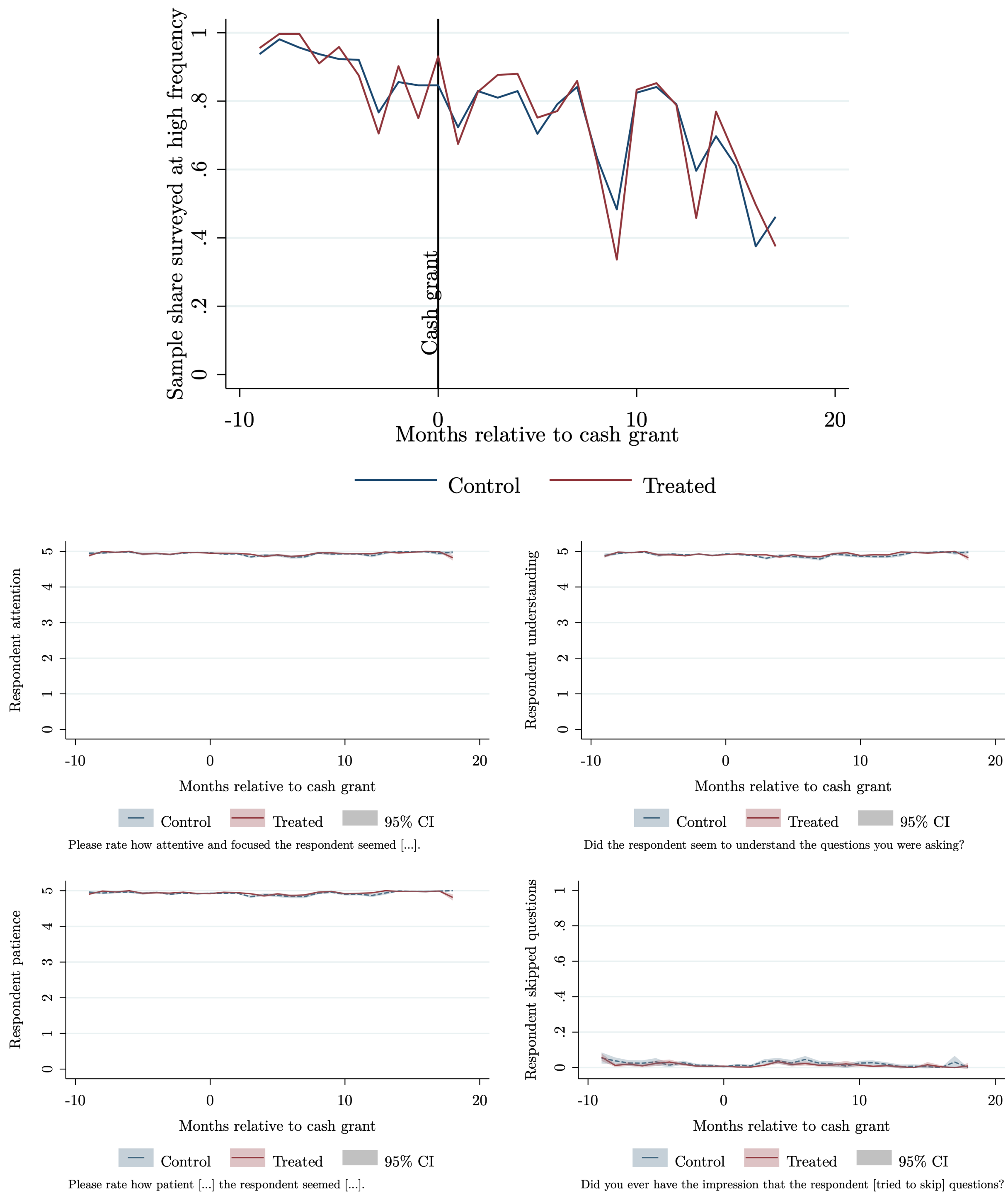


Figure A15: High-frequency data quality

Note: This figure describes data quality from high-frequency surveys, scored at the end of each interview.

B Sentiment analysis at high frequency

This appendix section provides details on the high-frequency sentiment analysis. Starting shortly after the launch of cohort 2 (and thus about one month after the cash grant for cohort 1 and before the start of cohort 3), I included an open-ended question to the start of each home visit survey asking: "First, I would like to hear from you in your own words what happened to you and your household since our last meeting? What are important events that affected you? What are you concerned or hopeful about?" Roughly half of the answers are detailed enough to allow sentiment analysis, while the remainder are brief comments informing the enumerator that nothing worthy of reporting had occurred. Answer lengths are unaffected by treatment.

Using the VADER lexicon, I analyze overall valence of the sentiment in the responses over time and standardize them by the control group mean and standard deviation. The top panel of Figure B16 shows this sentiment over the duration of the study. For the entire duration of the study, the treatment group remains about 0.25 standard deviations above the control group's overall sentiment score. I then split out four different domains of sentiment by manually scoring the most common 200 words into optimism ("hope", "excited", "progress", "happy", ...), worry ("worry/ied", "concern(ed)", ...), inner locus of control ("successful/ly", "start(ed)/ing", "manage(d)/ing", "build/t", ...), and outer locus of control ("receive(d)/ing", "affect(ed)/ing", "accept(ed)/ing", "need", "lack", ...). The middle panel shows the difference in the shares of surveys that express optimism relative to those expressing worry. I show that while optimism prevails in the treatment group at levels significantly above those in the control group (and with a spike around the time of the cash grant), the two groups reconverge about six months after the cash grant. It is remarkable how closely optimism follows the timing and amounts of money received from the NGO. External locus of control is the idea that one's fate is determined by forces outside of one's control, while internal locus of control expresses that one self can determine one's fate. Locus of control is similar between the two groups at first, while treatment recipients depend on NGO consumption support. After the cash grant, when participants start their own businesses, locus of control shifts inward and the gap remains for the remainder of the study period.

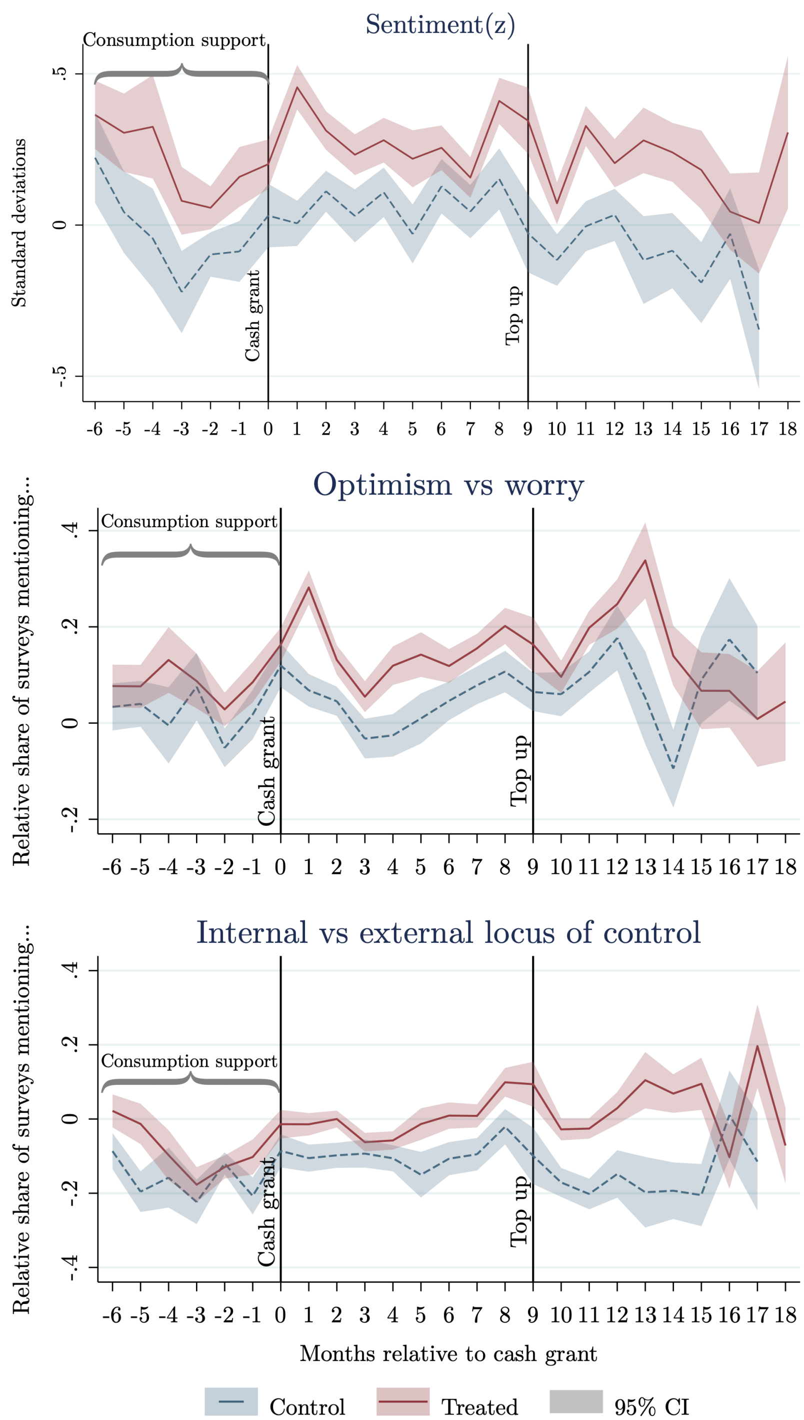


Figure B16: High-frequency sentiment analysis

Note: This figure describes the sentiment of households' life narratives.

Chapter II

Micro-enterprise saturation: Critical mass or overcrowding?

Abstract Should entrepreneurship support programs to extremely poor people be scaled by saturating a few villages, or maintaining low saturation and enrolling more villages at a higher cost? On the one hand, saturation may create a *critical mass* of entrepreneurs whose income generation and associated spending multipliers make each other viable. On the other hand, encouraging the start-up of many small firms in close proximity may be a recipe for *overcrowding* and artificially fierce competition in a small village economy. In a large randomized control trial in Malawi, I exogenously vary the saturation of an entrepreneurship support program to assess whether business performance increases or decreases in the share of treated neighbors. Business performance is tracked at fortnightly high frequency. Most participants start a retail business. Eighteen months after the start of the intervention, earnings take a 20% hit as saturation doubles and aggregate earnings are split among the larger number of supported entrepreneurs. About a third of the new retail businesses have gone out of business. Earnings losses and firm exits are of a comparable magnitude to implementer cost savings. The heightened competition does not lower consumer prices, but induces out-migration. I show that the sectoral concentration in retail is not driven by information frictions about other entrants' intended sectoral choice. Instead, respondents highlight a lack of hard skills.

1 Introduction

Poverty alleviation in the absence of stable employment opportunities from the private or public sector relies on individual entrepreneurship by poor people themselves (Blattman and Ralston, 2015; Banerjee et al., 2025). Policies that support entrepreneurship in developing countries are growing in popularity, but economic theory does not offer a clear-cut prediction of their impact at scale. On the one hand, the reason we are seeing few firms in contexts of extreme poverty might be that firms are not individually profitable as they depend on the weak demand of other similarly poor people. It is possible that a large-scale support program would create a critical mass of entrepreneurs whose income generation and associated spending multipliers make each other viable. The logic of a big-push individual-level intervention (Balboni et al., 2022) would then carry over to the aggregate (local) economy (Rosenstein-Rodan, 1957). On the other hand, encouraging the start-up of many small firms in close proximity may be a recipe for overcrowding and artificially fierce competition in a small village economy (Hardy and Kagy, 2020). Failure rates for micro-enterprises are high, especially soon after launch (De Mel et al., 2012; Balboni et al., 2022; Misha et al., 2019) and anecdotally too many UPG-beneficiary households start similar businesses in close proximity (Bauchet et al., 2015). This would suggest tight upper bounds to how far these programs can be scaled – and more broadly to the approach of alleviating poverty bottom-up from the household level (Goldberg and Reed, 2023).

These two scenarios (critical mass versus overcrowding) motivate my study. They imply contrasting policy prescriptions for whether entrepreneurship should be fostered by saturating a few locations or scaling over a larger geographic area at low saturation. In a large randomized control trial in Malawi, I exogenously vary the program saturation of an entrepreneurship support training to assess whether business performance increases or decreases in the share of treated neighbors. Saturation at the household cluster level is randomly assigned to be either 50% or 100% of eligible households, corresponding to 18.5% and 37% of the population, respectively. I also test a simple coordination intervention encouraging households to coordinate with each other when choosing their business venture, rather than duplicating existing businesses. The entrepreneurship support program under investigation is the graduation model, first popularized by the Bangladeshi NGO BRAC, which combines monthly cash transfers, productive assets or business start-up grants, entrepreneurship training, formation of savings groups, and one-on-one life skills coaching.

The coordination intervention did not succeed in shifting entrepreneurs-to-be out of the retail sector. While there are weak indications of sectoral shifts in both the business plans that program participants submit as well as the livelihoods households report very shortly after the cash grant, these differences dissipate within just a few months. My study demonstrates that the barrier is not simply an information friction, but I provide evidence that the lack of hard skills confines these entrepreneurs in the petty trading sector.

Eighteen months after the start of the intervention, the saturation results are nuanced: In terms of earnings, the businesses in high and low saturation clusters initially track each other's performance closely over time, but show signs of divergence, with low saturation cluster entrepreneurs earning at times up to twice as much. The cumulative earnings are not significantly different from one another, but the point estimate for the high-saturation group is about 20% lower than that of the low-saturation group. It is important to note that an implementer may prefer to implement at high saturation even if earnings were significantly lower as long as earnings losses do not offset savings from economies of scale. In the program I analyze, the cost savings at high saturation are about 30%. Firms in the high saturation clusters are significantly more likely to exit, with households reverting to casual labor. An even starker indication of overcrowding is that the higher saturation induces out-migration.

Ultra-poor graduation programs (UPGs), aimed at promoting entrepreneurial activity by people in extreme poverty, are the closest that development economics has come to a silver bullet for sustainably alleviating extreme poverty (Sulaiman et al., 2016; Wydick, 2020). While in-kind, conditional and unconditional cash transfers have been shown to provide effective poverty relief and invigorate micro-enterprises in the short run (Sulaiman et al., 2016; Baird et al., 2011; Blattman et al., 2014; Haushofer and Shapiro, 2016; Cahyadi et al., 2020; Brooks et al., 2022; Egger et al., 2022), in-kind transfers or strong-handed conditionalities may induce distortionary effects (Bryan et al., 2021), and the evidence on long-term effects is mixed at best and favors settings where households are above the international extreme poverty line to begin with (Gertler et al., 2012; De Mel et al., 2012; Aizer et al., 2016; Ikegami et al., 2017; Haushofer and Shapiro, 2018; Araujo et al., 2019; Millán et al., 2020; Aggarwal et al., 2022; Altındağ and O'Connell, 2023; Fiala et al., 2022). Skills training, coaching, or access to loans or savings facilities by themselves show few if any impacts (Blattman and Ralston, 2015; Berge et al., 2015; Valdivia, 2015; Bauchet et al., 2015; Chowdhury et al., 2017; Brooks et al., 2018; Sedlmayr et al., 2020; Giné and Mansuri, 2021; Leight et al., 2021, 2022). On the other hand, UPGs following the BRAC model (studied here) seem to be standing the test of time while providing results that are larger than the sum of their parts in response to the highly complex challenge that multidimensional poverty traps pose (Sulaiman et al., 2016; Banerjee et al., 2015; Bandiera et al., 2017; Banerjee et al., 2021; Balboni et al., 2022; Chowdhury et al., 2017; Sedlmayr et al., 2020; Bedoya et al., 2019). If the first Sustainable Development Goal of ending extreme poverty by 2030 is to be achieved, the development community needs to better understand how UPGs scale. My study highlights that a key barrier to scale will be the access to a broader set of sectors through the provision of diverse hard skills – a challenging barrier considering that providing a diverse set of hard skills within the same location will erode economies of scale.

Spillovers are a common concern in poverty alleviation programming. While Egger et al. (2022) find almost no inflationary pressure from cash transfers in Western Kenya, Filmer et al. (2021) argue that this may be linked to relatively low saturation in the study area while in their own study they find important price increases when cash transfer saturation is high and markets remote and Cunha et al. (2019) find differential price effects by village wealth. In the context of business training interventions, McKenzie and Puerto (2021) find no evidence of negative side effects, while Drexler et al. (2014) show that in their context, at least part of the benefits to the treated come from business stealing from the control group. Such spillovers may also extend to programs that support new entrepreneurial activity, the focus of this study. I do not find impacts of program saturation on the price of the key commodity, maize. Spillovers from the treatment group onto the control group are likewise modest. But I show that when taking an intervention to scale, it is important to consider spillovers from treated onto other treated people.

This paper proceeds as follows. Section 2 describes the setting, research design, and intervention. Section 3 discusses the analysis, while results can be found in Section 4.

2 Experimental design

Setting and census This study takes place in Mangochi district in the South of Malawi, one of the poorest districts in the country. To determine program eligibility and convey an accurate picture of the local economy, I conducted a household census of all 29,000 households in the area, covering a population of almost 140,000 individuals (Appendix Table A1). The majority of the population are subsistence or pre-commercial farmers (Appendix Figure A1). Outside the harvest season, few have any income sources other than ad hoc daily labor ("ganyu"). About 40% of aggregate production in the study area is consumed without being traded in a market. Three quarters of households report at least one member having skipped a meal in the past week (Appendix Table A3). This is in line with the fact that agricultural yields are low. The 2018 harvest lasted fewer than 20% of farmers longer than six months and four months after the 2019 harvest only half of registered households had any food reserves at all. It is six months after harvest that investments in the next planting, both in terms of seeds and fertilizers as well as in terms of labor need to be made, but when a majority of the rural poor are running precariously low on food and money, which they compensate for with casual labor, drawing them away from their own fields. Existing businesses are overwhelmingly one-person enterprises and a sizable share depend on demand within their own village.

Sample selection and treatment assignment The sampling process is described in Appendix Figure A2. Using the Malawi-validated Simple Poverty Score Card (Schreiner, 2019) and self-reported income, I identified about 40% of households as ultra-poor and eligible.[19] Using k-means geographic clustering, these roughly 11,000 households were assigned to 271 clusters. Due to how households tend to live nearby one another, clusters often roughly correspond to villages when these are small and remote. In larger towns, villages are usually comprised of several clusters or a cluster can combine households from more than one village. To determine the sample, I deployed a custom algorithm to select 72 clusters into the study in such a way as to maximize the distance between clusters and minimize the potential for spillovers with buffer clusters in between that are not part of the study.[20] The median household cluster is about 30 hectares in size and is home to about 70 households. The minimum distance between study clusters corresponds to the distance beyond which Egger et al. (2022) no longer find spillover effects of cash transfers.[21] In my analysis, I will therefore allow for spillovers within but not across clusters. Though this assumption may not exactly hold in practice (for example information is likely to "travel further" than cash), the assumption is that spillovers are substantially greater between neighbors within a cluster than they are across clusters. Clusters were grouped into sets of three that were assigned to a program implementer as well as being randomly assigned to one of three cohorts to be launched a few months apart from each other (early 2024, mid 2024, mid 2025, see Appendix Figure A3).

Within each implementer's area (stratum), the three clusters were assigned at random to either high treatment saturation (everybody treated), low treatment saturation (half treated,

[19]The implementing partner excluded households from participation that have no female; have no children under the age of 5 as of August 1, 2023; did not appear ultra-poor based on either their stated income or the Simple Poverty Scorecard; in which the female was unable to carry out physical labor; whose data appeared implausible to the research team or enumerator; or who were not willing to be registered or could not be found during several (mop up) visits. The latter two criteria were very rare.

[20]After cluster selection and baseline survey, but before the start of the intervention, the local government designated one of our clusters as a flood-prone area and ordered it to be vacated. As our study may have encouraged people to stay in an unsafe area, we shifted to a neighboring cluster in this one instance.

[21]If anything, my study area's road network is substantially less developed and personal ownership of bicycles or even motorbikes is lower in Southern Malawi than in Western Kenya and the terrain is sliced up by seasonal rivers that make effective distances larger than straight-line distance.

randomly selected at the individual level) or pure control group (nobody treated). The saturation levels were chosen following Baird et al. (2018).[22] Assigning each implementer one of each type of cluster is tantamount to geographic and implementer stratification (Bruhn and McKenzie, 2009; Bai, 2022).

In order to balance the sample on a number of key outcome and mechanism variables (see Table A4), I first run the treatment assignment algorithm 100,000 times (seeding each with the counter of its iteration). I then estimate baseline balance under each theoretical treatment assignment and discard assignments that lead to a statistically significant imbalance in any of the key mechanism variables at the 10% significance threshold (Morgan and Rubin, 2012). In other words, I enforce a "visually clean" balance table. I then randomly pick one of the permissible treatment assignments. As pointed out in both Bruhn and McKenzie (2009) and Morgan and Rubin (2012), classical inference that does not take this rerandomization into account would now on average (though not in every single case) be too conservative, leading to confidence intervals on the average treatment effect with overly wide coverage and too few rejections of null hypotheses. A Fisher (1935) exact test has the correct coverage, though it tests the sharp "randomization hypothesis" that the treatment had no effect on anyone rather than on average (Ritzwoller et al., 2025). Traditionally, researchers using rerandomization methods have done so somewhat *ad hoc*, rerandomizing until the balance table was satisfactory. My approach is more transparent. Given R simulated assignments and k baseline balance table variables, the likelihood of discarding an assignment at the 10% significance level is $q = 1 - (1 - 0.1)^k$ and I can therefore use the remaining $R \times (1 - q)$ simulated assignments for the randomization inference (Young, 2019).

I leverage the exogenous variation between high and low saturation to answer whether

[22] I am designing my study to be optimally powered for spillovers among treated households. Modifying equation (6) in Baird et al. (2018) to omit the standard error of the spillover effect on the control group, I restate their Theorem 2 to conclude that the optimal high saturation is 100% and the optimal low saturation would have been below 30% and dependent on intra-cluster correlation of outcomes. However, in conversation with the implementing partner, we decided to increase the lower saturation to 50% which would represent a more policy-relevant comparison. Less saturated clusters would also make program components like savings groups difficult to implement.

increased saturation leads to excess competition and overcrowding of small village economies or to critical mass and spending multipliers through which the budding entrepreneurs make each other viable.

Intervention The intervention is a standard BRAC-style UPG (Banerjee et al., 2015; Bandiera et al., 2017) with two key modifications. First, instead of livestock and related training, households receive a one-off labeled[23] cash grant of $300 intended for starting a business. Flexibility is given to participants for leveraging their own informational advantage over ability and preferences and they receive four months of weekly general-purpose business training covering financial literacy, market analysis, budgeting, record keeping, funding sources, and the development of business ideas. Second, program participants (mothers of children under five) receive an additional four-month training course on early childhood development. Otherwise, the program is comparable to BRAC's and other UPGs in the literature: Treated households receive six monthly cash transfers of $20 (labeled for investments in child health and wellbeing as well as business experimentation prior to the large grant). The structure of cash transfers is closely aligned with evidence on how poor households themselves prefer to receive such funds, combining a regular buffer with a large lump-sum that arrives after sufficient time for deliberation (Kansikas et al., 2023).

All households in the study receive fortnightly visits from program implementers which in the original program design are intended for tracking training progress and answering questions the participant may have. During the program evaluation, we additionally use these meetings for brief surveys on consumption and investment choices, livelihood activities, as well as shocks and coping behavior. Both the treatment and control group receive a basic phone for scheduling of surveys and for receipt of cash transfers or survey incentives through mobile money. At the community level, the formation of savings and loan groups is

[23]The grant is largely unconditional, but the NGO encourages that it be used for starting a business or supporting an existing one. Receipt requires attending the training and being able to adequately verbalize a concrete business plan, but participants can miss sessions without penalty and are coached until their plan is satisfactory and the actual implementation of said plan is not afterwards enforced.

encouraged. Local childcare centers are supported financially and with the same ECD training. Both treatment and control group are expected to benefit from this last aspect.

In half of the treated clusters, implementers do not guide which kinds of businesses participants start start, while the other clusters receive a business selection guidance treatment. In these clusters, participants first listen to a story of three women in a similar position as them, having participated in an entrepreneurship training program and ready to receive their cash grant when they discover that they all three had the same business idea. The women in the story come to the realization that there would probably not be enough demand for all three of them to pursue the same business and they instead choose different businesses. After reading the story, the implementers ask participants to share with each other what kinds of businesses they would like to start (as a ranked preference list). Participants are then asked to consider what would happen to their profitability if they all started similar businesses and encouraged (but not forced) to coordinate so that they do not all start the same businesses as one another. The goal is to first widen participants' pool of potential business ideas and second to prime them to think about the general equilibrium consequences of their business choice. The coordination intervention is motivated by the fact that when asked why they would like to pursue a particular research idea, the majority of participants mentioned that they had seen other people earn good money this way. For an individual prospective entrepreneur, imitation is of course a potentially sensible strategy. But the fact that this is by far the predominant motive points to a potential neglect of general equilibrium effects that may be detrimental at scale (Figure 1).

3 Empirical strategy

There are two natural benchmarks for program performance. One is to ask, whether more or fewer businesses become successful relative to how many were supported (and how well they perform) as more people are treated (Figure 2(a)). So for example, if five out of thirteen supported businesses are successful in a low saturation cluster, then if the program

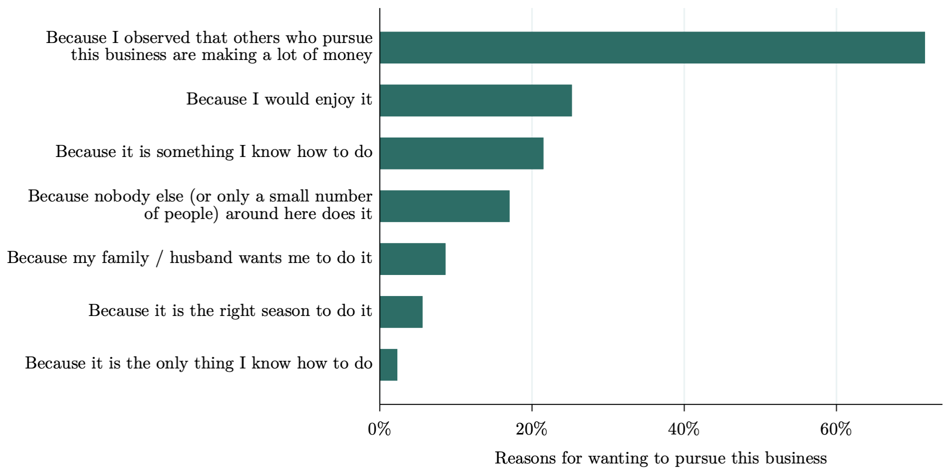
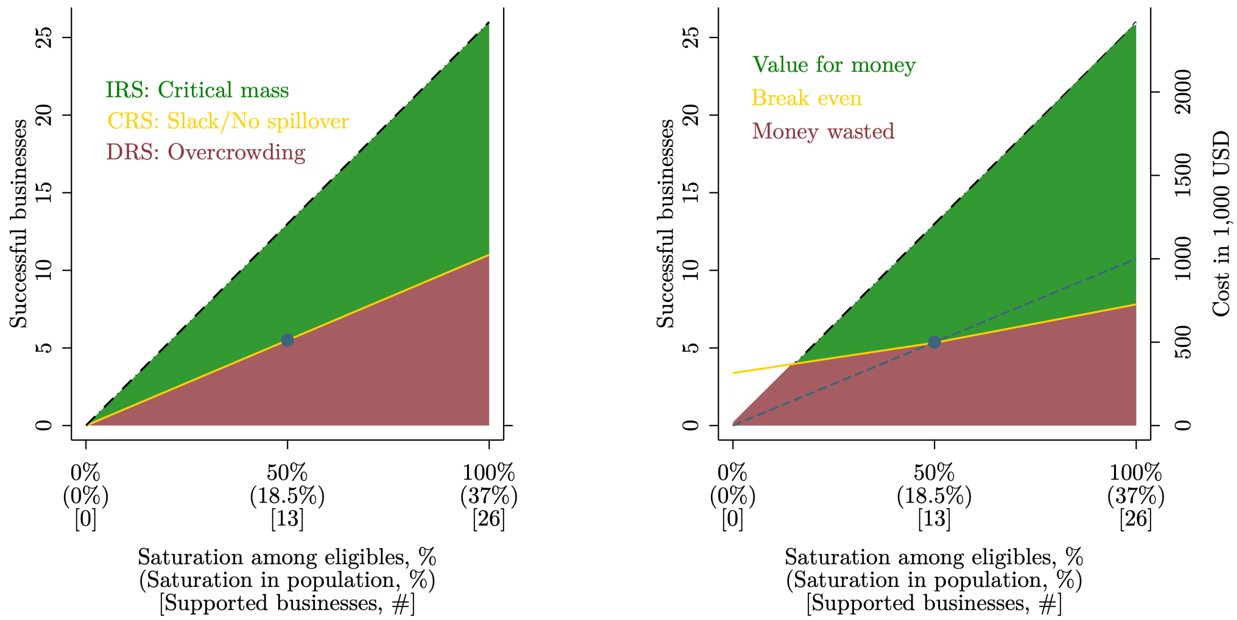


Figure 1: Evidence of general equilibrium neglect

Notes: This figure summarizes common motivations that participants state for why they would like to pursue a given business idea. The overwhelming majority expresses what I term general equilibrium neglect: That they would like to emulate the profits of others, perhaps without internalizing the effects of their own entry. I hypothesize that this will have detrimental consequences in a high-saturation implementation and it motivates devising a business coordination intervention.

scales with constant returns to saturation, we would expect ten to succeed when 26 are supported. Under increasing returns, it would be more than ten and vice versa. But in practice, some losses in program effectiveness are permissible considering the important economies scale in implementation cost (Figure 2(b)). High-saturation clusters, where all 26 households (100% of eligible households) are treated, correspond to a saturation level among the broader population surveyed during census of about 37% in the median cluster. In low-saturation clusters, 13 randomly chosen households (50% of the eligibles) are treated, corresponding to 18.5% of all households in the median cluster. Given my intimate familiarity with the cost structure of the program I evaluate, I am able to investigate not just whether there are *any* adverse saturation effects (which will be of interest to academic economists), but also how they relate to the cost savings (which will be of interest to policy makers and practitioners) when the goal is to maximize the number of beneficiaries with viable businesses *given* a fixed budget. The yellow line in the right panel, shows the program cost increasing less than one-for-one between the low and high-saturation scenario. The gap between the blue dashed and the yellow solid line represents the case where the program has decreasing returns to saturation, but they are outweighed by the economies of scale.



(a) Suppose at low saturation, the treatment produces about five successful businesses (blue dot). Relative to that, scaling with constant returns to scale (no spillovers) is indicated by the yellow line while increasing and decreasing returns to scale, are scenarios indicated by green (critical mass) and red (overcrowding) respectively.

(b) Implementation costs, however, increase less than proportionally with saturation and the area between the constant returns to scale line (blue, dashed) and the decreasing cost of scale curve (yellow) may still be attractive if the goal is the most successful businesses per dollar spent.

Figure 2: Conceptual framework

As pre-specified, I estimate the following linear regressions:[24]

$$y_i = \beta_0 + \beta_T \text{Treatment}_i + \beta_C \text{Coordination Intervention}_i + \beta_{\text{BL}} y_i^{\text{BL}} + \beta_{\text{M}} \mathbb{1}[y_i^{\text{BL}} \text{ missing}] + \beta_{\text{FE}} S_i + \varepsilon_i \tag{5}$$

$$y_i = \beta_0 + \beta_T \text{Treatment} + \beta_H \text{High saturation} + \beta_{\text{BL}} y_i^{\text{BL}} + \beta_{\text{M}} \mathbb{1}[y_i^{\text{BL}} \text{ missing}] + \beta_{\text{FE}} S_i + \varepsilon_i \tag{6}$$

Standard errors are clustered at the strata (cluster triplet) level (De Chaisemartin and

[24]The pre-analysis plan additionally contained an interaction of the saturation level and coordination intervention. Since the latter turned out to be ineffective, I concentrate the statistical power on the simple comparisons instead, which should therefore be interpreted as weighted averages of the treatment effect of the uninteracted treatment effects (Muralidharan et al., 2023).

Ramirez-Cuellar, 2024) which is the level of randomization and I am controlling for baseline levels of all outcome variables (ANCOVA). I investigate three key outcome variables, sector choice, business survival and profits.[25] Equation 5 investigates whether the addition of the coordination information intervention alters sector choices among treated households.

The sign of β_H in equation (6) is the main parameter of interest in this study, theoretically ambiguous, and merits discussion. If it is relatively close to zero for both outcomes, there would be no evidence of spillovers among the treated for business performance. This would indicate room in the local economy for more firms. If the coefficient is positive for either outcome, this would speak in favor of the spending multipliers and critical mass hypothesis. If either one or both of the coefficients are negative, this would support overcrowding. If one were positive and the other negative (e.g. more surviving firms but with lower profits), then the local economy may be at capacity, but the constraint operates only along one margin.

4 Results and mechanisms

4.1 Regression results

In Appendix Table A4, I show sample balance. As discussed above, the clean balance table follows mechanically from the randomization approach. Attrition at the one-year follow-up (midline) survey is 2.7% at midline and 9.5% at endline. Attrition is balanced between treatment and control group.

The coordination lesson did not meaningfully increase the number of distinct or unique businesses in a cluster. It appears to have marginally compelled program participants to *propose* non-retail service businesses ahead of receiving the cash grant (Table 1), but it did not meaningfully affect the *actual* sectors they are operating in after receiving the cash grant

[25] In what follows, I will use the term "profit" somewhat loosely. To be precise, I will focus on the gross margin, defined as revenue minus all ongoing expenses and abstract away from any amortization of a potential asset that households may already own or buy from intervention cash grant.

(Table 2). Since the coordination intervention did not affect business choices meaningfully, I henceforth pool their results to compare high and low saturation clusters.

Turning to the performance of the businesses, I compare high-saturation clusters to two benchmarks, as described in section 3. First, I compare them to low-saturation clusters to investigate whether program impacts scale linearly or display positive or negative spillovers that intensify with saturation. Second, I compare performance to a cost benchmark, since implementation at higher saturation saves the implementing partner about 30% of their per-household costs. I use two main metrics: One view of these entrepreneurship support programs is that not everybody is a good entrepreneur and it is very hard to know ex ante who will be talented (or lucky) enough to perform well (Hussam et al., 2022; Haushofer et al., 2022; Banerjee et al., 2025). I therefore count as successful entrepreneurs those who earn at least threefold the average control group earnings between after the cash grant and for the remainder of the study duration. I estimate that a little over half of households in both low and high-saturation clusters are successful by this metric, an increase by 25 percentage points over the control group (Figure 3, left). The estimate is somewhat but not significantly lower in the high-saturation group where it sits in between the low-saturation and the cost benchmark.

When it comes to earnings, the point estimates group are about 20% lower in high-saturation clusters than in the low-saturation clusters, close to the cost savings benchmark, though not statistically significantly different from either benchmark (Figure 3, right). I can confidently reject large positive spillovers as found in Egger et al. (2022). Figure 4 shows that earnings diverge from time to time between the two cluster saturation levels (Table 3). Earnings co-move within clusters, making standard error clustering very taxing.

With more firms and lower earnings per firm, it seems reasonable to ask whether the business environment in the treated areas has become more competitive, entailing spillovers outside the study sample. I track maize prices from August 2024 onwards in the entire study area. As Malawi's foremost staple crop, maize has a pivotal role as a barometer for price levels more broadly and many of the studied businesses engage in retail of food crops and may exert downward pressure on maize prices. Figure 5 suggests that maize prices track each other through heavy seasonal fluctuations in all cluster types for the first few months of the study, though during the lean season price peak, the treated clusters seem to have had slightly lower prices. This divergence, however, is driven if anything by the low saturation cluster, suggesting that higher saturation is not a driver of more competitive retail markets for maize.

Table 1: Business choices (planned)

	(1) Treatment mean (SD)	(2) Coordination Intervention	(3) N (BL)
Agriculture, Fishing, Forestry	.0733	-.009082	914
	(.2608)	(.01564)	✓
Manufacturing	.0339	-.0008599	914
	(.1811)	(.01664)	✓
Non-retail Services	.0481	.0351*	914
	(.2142)	(.01824)	✓
Retail	.8239	-.01487	914
	(.3812)	(.02813)	✓
Distinct businesses	7.604	.9583	48
	(2.06)	(.9639)	×
Unique businesses	3.75	.6667	48
	(1.952)	(.8102)	×
Cluster HHI	.066	-.0005026	48
	(.0238)	(.01337)	×

Notes: Analysis in this table is among treated households (who wrote a business plan) and the treatment coefficient is therefore omitted. The last column indicates the number of observations. All regressions are implemented as an ANCOVA specification, controlling for strata fixed effects. A ✓ indicates that the outcome variable is available for all observations at baseline. If it was not collected at baseline, a × is shown, and if it is partially available, missing observations are set to zero and dummied out (✓). Where available, regressions control for baseline levels of the outcome. Standard errors are clustered at the household cluster level. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 2: Business sector choices (actual)

	(1) Control mean (SD)	(2) Treatment Effect	(3) Coordination Intervention	(4) N (BL)
Casual labor, charity	.5733	-.3357***	.05437	997
	(.4952)	(.05922)	(.05456)	✓
		[0]	[.3477]	
Farming, fishing, forestry	.3033	.008952	.03826	997
	(.4603)	(.06333)	(.05231)	✓
		[.8991]	[.4955]	
Manufacturing	.0437	.06429**	-.03707	997
	(.2047)	(.02747)	(.03069)	✓
		[.04795]	[.2937]	
Non-retail services	.0797	.04801*	.05522**	997
	(.2712)	(.02447)	(.02293)	✓
		[.07093]	[.02298]	
Retail	.0925	.3128***	-.04897	997
	(.2902)	(.03084)	(.03629)	✓
		[0]	[.1958]	
Unique income source	.178	.06089*	.09704**	971
	(.383)	(.03171)	(.04239)	×
		[.08891]	[.02597]	
New income source (village)	.0445	.09141**	.05562	971
	(.2065)	(.03452)	(.04128)	×
		[.01698]	[.1968]	

Notes: The last column indicates the number of observations. All regressions are implemented as an ANCOVA specification, controlling for strata fixed effects. A ✓ indicates that the outcome variable is available for all observations at baseline. If it was not collected at baseline, a × is shown, and if it is partially available, missing observations are set to zero and dummied out (✓). Where available, regressions control for baseline levels of the outcome. Standard errors are clustered at the strata level. Fisher (1935) exact test p-values from 1000 simulated alternative treatment assignments are shown in brackets. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Business performance

	(1) Control mean (SD)	(2) Treatment Effect	(3) High Saturation	(4) N (BL)
Does business	.4723	.3744***	.01346	995
	(.3482)	(.04488)	(.01606)	✓
		[0]	[.4106]	
Investment in inventory	228,760	1,566,847***	-240,046	995
	(535,752)	(295,378)	(211,226)	×
		[0]	[.2757]	
∼ productive assets	12,273	27,294***	29,282*	995
	(105,741)	(8,618)	(15,018)	×
		[.000999]	[.06593]	
Hours worked	262.7	31.28	-22.13***	995
	(163.9)	(19.06)	(6.876)	×
		[.1329]	[.000999]	
Earnings	755,960	1,320,051***	-270,205	995
	(729,338)	(389,344)	(225,245)	×
		[.001998]	[.2817]	

Notes: The last column indicates the number of observations. All regressions are implemented as an ANCOVA specification, controlling for strata fixed effects. A ✓ indicates that the outcome variable is available for all observations at baseline. If it was not collected at baseline, a × is shown, and if it is partially available, missing observations are set to zero and dummied out (✓). Where available, regressions control for baseline levels of the outcome. Standard errors are clustered at the strata level. Fisher (1935) exact test p-values from 1000 simulated alternative treatment assignments are shown in brackets. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

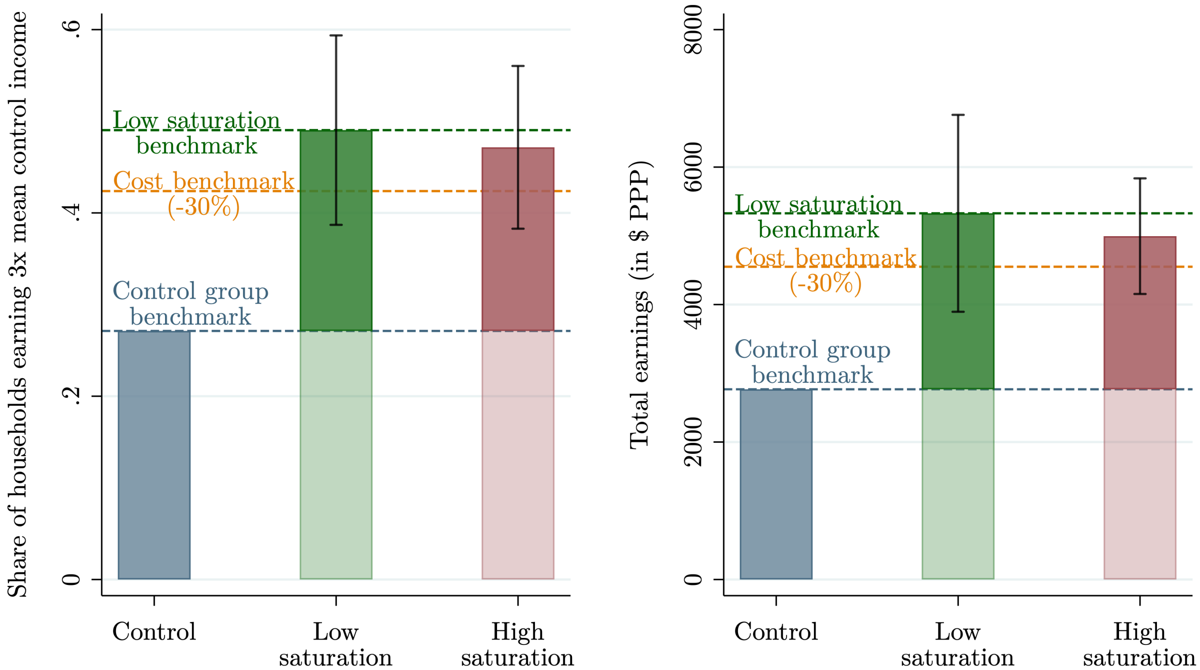


Figure 3: Earnings by saturation

Note: This graph compares regression results for indicators of being a successful entrepreneur (left) as well as average earnings (right) across cluster types.

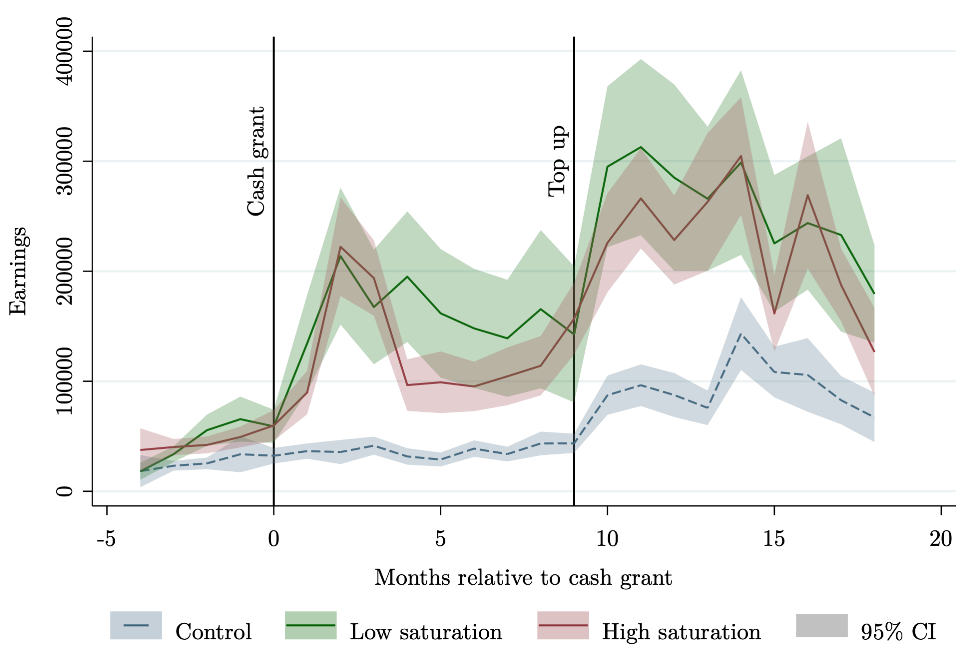


Figure 4: Earnings by saturation

Note: This graph plots monthly earnings over time relative to the cash grant. After an initial spike in the treatment group, earnings diverge and stay higher for the low-saturation clusters.

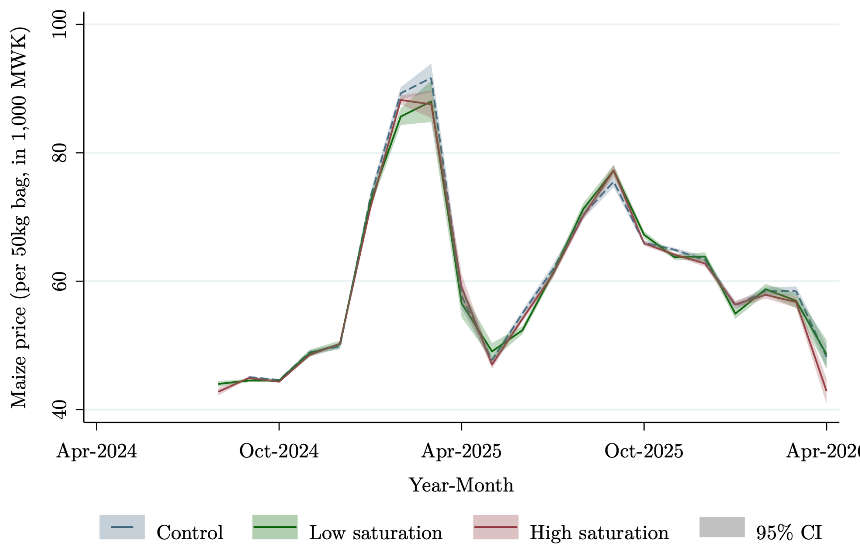


Figure 5: Maize price by saturation

Note: The price of maize, the dominant staple crop in the region, follows strong seasonal fluctuations, but does not appear affected by the high- or low-saturation treatment.

4.2 Mechanisms and robustness

The section discusses potential channels through which overcrowding might operate or manifest itself. Lower profits are not a mechanical byproduct of adverse selection into business as the 13 households selected into treatment among the 26 households that make up a low saturation cluster are chosen at random and Table A4 shows balance on ability and prior experience. I do not find any evidence that entrepreneurs in the different types of clusters spend more or less time working (Figure 6) or invest differentially in assets (Table 4).

All treated participants, regardless of saturation status, are more likely than the control group to *ever* sell goods at another town's market (Table 5) but the coefficient for the high-saturation clusters is sizable and negative though just shy of statistical significance. This may be a mediating response to increased competition in high saturation clusters and would explain why the spillover effects are not stronger. I find no effects on further away trips, longer travel in response to the treatment. The strongest indicator of competitive pressure is that households in high-saturation clusters are more likely to send an adult member permanently away or move with the household as a whole. The most common place for businesses to sell on a day-to-day basis, however, remains nearby (Figure 7).

A key concern for the validity of my results are spillovers across clusters. As discussed in section 2, household clusters are buffered in space in order to avoid spillovers across them. I now look for evidence of such spillovers. The presence of non-treated households in low-saturation clusters allows me to test for another type of spillovers from treated onto control households, by comparing them to pure control clusters. This is not a direct test of across-cluster spillovers. If one is willing to posit, however, that spillovers should generally decay with distance, then finding few spillovers onto immediate untreated neighbors from the treatment should be reassuring that there are no spillovers across clusters either. I test this using the pre-specified regression equation

$$y_i = \beta_0 + \beta_T \text{Treatment}_i + \beta_{\text{S}} \text{Spillover}_i + \beta_{\text{BL}} y_i^{\text{BL}} + \beta_{\text{M}} \mathbb{1}[y_i^{\text{BL}} \text{ missing}] + \varepsilon_i.$$

Here, Spillover is an indicator for being a control household in a low-saturation cluster and coefficient β_S compares these households directly to control group households in pure control clusters. Results in Table 6 show relatively little evidence of such spillovers onto non-treated households who are much less likely to run competitor firms to the treated.

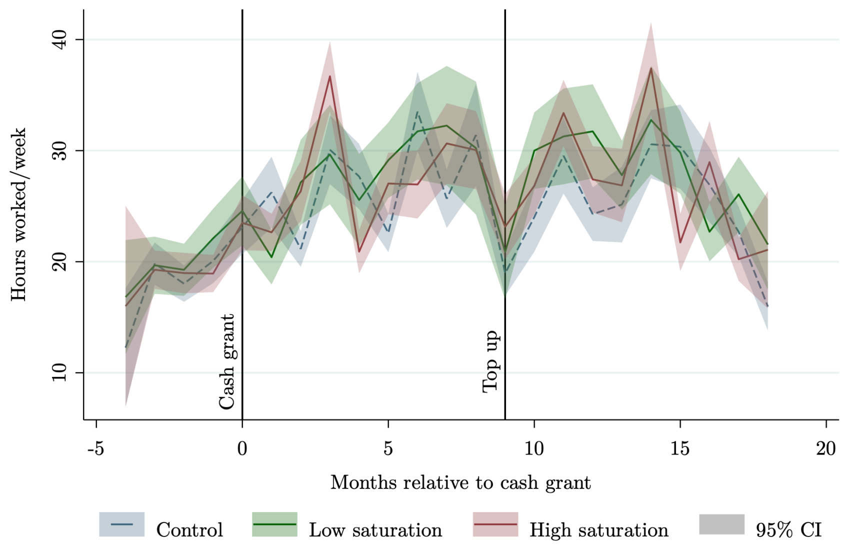


Figure 6: Hours worked by saturation

Note: Hours worked per week are not affected by treatment.

It does not appear as though the larger cluster size deteriorated training quality significantly. Table 7 shows that the high-saturation clusters score consistently (but not statistically significantly) lower than their low-saturation counterparts, though only by less than 1 point out of 100. I also verify whether the coordination intervention is reflected in the test scores by investigating whether coordination-related themes are mentioned by program participants. Those who received the coordination intervention were indeed more likely to mention being aware of one's competitor's weaknesses as factors that contribute to business success, though they are no more likely to say that one should find one's own niche and there is no difference in whether they highlight the presence of competitors (Table 8).

Table 9 splits the sample by whether households were engaged in retail businesses at baseline (column (3)) and those who were not (column (2)). It reveals that prior experience

Table 4: Assets owned

	(1) Control mean (SD)	(2) Treatment Effect	(3) High Saturation	(4) N (BL)
Total assets (MWK)	391,459	170,593***	22,104	1214
	(473,494)	(26,692)	(37,693)	✓
		[0]	[.5984]	
Business assets (MWK)	25,543	49,496**	38,951*	1214
	(114,443)	(17,258)	(21,047)	✓
		[.02098]	[.09091]	
Agricultural Assets (MWK)	14,960	8,907***	-5,802*	1214
	(19,353)	(2,539)	(2,909)	✓
		[.004995]	[.06793]	
Personal assets (MWK)	341,534	117,109***	-12,298	1214
	(387,610)	(25,032)	(26,639)	✓
		[.000999]	[.6683]	

Notes: The last column indicates the number of observations. All regressions are implemented as an ANCOVA specification, controlling for strata fixed effects. A ✓ indicates that the outcome variable is available for all observations at baseline. If it was not collected at baseline, a × is shown, and if it is partially available, missing observations are set to zero and dummied out (✓). Where available, regressions control for baseline levels of the outcome. Standard errors are clustered at the strata level. Fisher (1935) exact test p-values from 1000 simulated alternative treatment assignments are shown in brackets. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

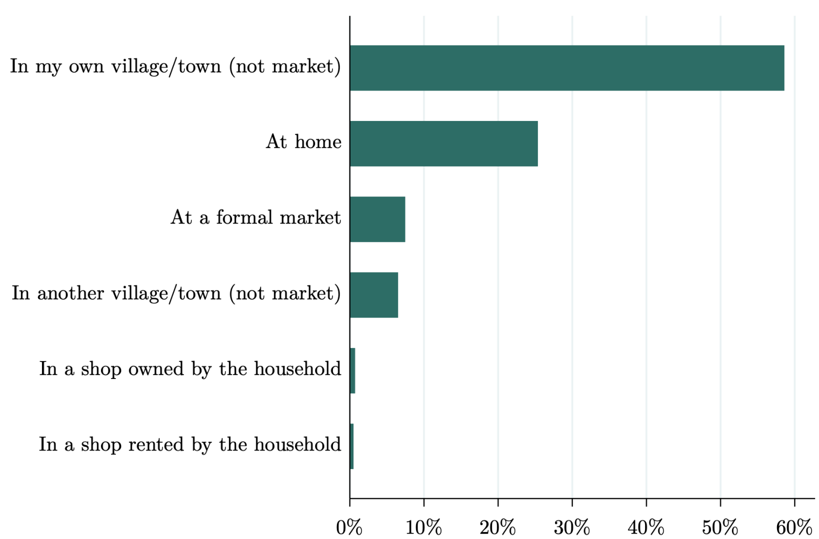


Figure 7: Location of most business sales

Note: The fact that most business sales happen within the same village or right in front of an entrepreneur's house reduces across-cluster spillover concerns through markets.

Table 5: Mobility

	(1) Control mean (SD)	(2) Treatment Effect	(3) High Saturation	(4) N (BL)
Sell at local market	.3183	.1853***	-.09576	1214
	(.4662)	(.05113)	(.06305)	✓
		[.001998]	[.1708]	
Sell elsewhere	.2663	.261***	-.0371	1214
	(.4424)	(.04984)	(.04665)	✓
		[0]	[.4396]	
Minutes to market	56.72	2.54	5.38	1213
	(52.91)	(4.875)	(7.104)	✓
		[.6374]	[.4885]	
Adults left HH	.2613	-.03173	.1113**	1214
	(.612)	(.03974)	(.04364)	✓
		[.4356]	[.02597]	
HH left study area	.0609	-.04647**	.02163**	1248
	(.2393)	(.01703)	(.008637)	×
		[.01598]	[.02897]	
Times left district	.4774	1.044	-.7362	1214
	(2.098)	(.6224)	(.6285)	✓
		[.1009]	[.2587]	
Traveled to Lilongwe	.1022	.06059	-.01138	1214
	(.9561)	(.09984)	(.1059)	✓
		[.6384]	[.9421]	
∼ Blantyre	.1139	-.07932	.09273	1214
	(1.276)	(.05907)	(.05423)	✓
		[.2388]	[.05594]	
∼ Mangochi town	.3903	.5313***	.3821	1214
	(1.381)	(.1576)	(.3918)	✓
		[.003996]	[.3916]	

Notes: The last column indicates the number of observations. All regressions are implemented as an ANCOVA specification, controlling for strata fixed effects. A ✓ indicates that the outcome variable is available for all observations at baseline. If it was not collected at baseline, a × is shown, and if it is partially available, missing observations are set to zero and dummied out (✓). Where available, regressions control for baseline levels of the outcome. Standard errors are clustered at the strata level. Fisher (1935) exact test p-values from 1000 simulated alternative treatment assignments are shown in brackets. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Spillovers onto untreated households

	(1) Control mean (SD)	(2) Treatment Effect	(3) Spillover Control	(4) N (BL)
Does business	.6851	.2096***	-.03621	1214
	(.4649)	(.03662)	(.04995)	✓
Casual labor, charity	.804	-.167***	.03153	1214
	(.3973)	(.03578)	(.03996)	✓
Farming, fishing, forestry	.6131	-.08689**	-.006033	1214
	(.4875)	(.04058)	(.05075)	✓
Manufacturing	.1424	-.01298	-.02965	1214
	(.3497)	(.02128)	(.02125)	✓
Non-retail services	.1457	.1022***	-.02338	1214
	(.3531)	(.02086)	(.02668)	✓
Retail	.1977	.4538***	.0491	1214
	(.3986)	(.03969)	(.03716)	✓
HH income (12 months, MWK, winsorized)	664,083	384,265***	-80,732	1213
	(647,484)	(62,354)	(57,500)	✓
Total assets (MWK)	391,459	169,153***	-48,596*	1214
	(473,494)	(33,447)	(22,897)	✓
Business assets (MWK)	25,543	75,036***	-836.2	1214
	(114,443)	(18,611)	(8,842)	✓
Agricultural Assets (MWK)	14,960	4,160**	-2,721	1214
	(19,353)	(1,691)	(2,055)	✓
Personal assets (MWK)	341,534	95,330***	-41,254*	1214
	(387,610)	(23,894)	(23,101)	✓
Any livestock	.4941	.4076***	-.04251	1214
	(.5004)	(.0466)	(.03304)	✓
Livestock value (MWK)	41,333	70,479***	-4,984	1214
	(91,642)	(10,588)	(5,338)	✓
Any savings	.3166	.4685***	-.0777*	1214
	(.4655)	(.0396)	(.03802)	✓
Total savings (MWK)	12,791	59,421***	-131.8	1214
	(32,717)	(8,584)	(2,936)	✓

Notes: The last column indicates the number of observations. All regressions are implemented as an ANCOVA specification, controlling for strata fixed effects. A ✓ indicates that the outcome variable is available for all observations at baseline. If it was not collected at baseline, a × is shown, and if it is partially available, missing observations are set to zero and dummied out (✓). Where available, regressions control for baseline levels of the outcome. Standard errors are clustered at the strata level. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Post-training evaluation results

	(1) Treatment mean (SD)	(2) High Saturation	(3) N (BL)
Business knowledge	.4562	-.005078	923
	(.16)	(.01237)	✓
ECD knowledge	.5212	-.008697	923
	(.1347)	(.00958)	✓
Preventive health knowledge	.6042	-.009652	923
	(.2078)	(.01519)	✓

Notes: Analysis in this table is among treated households (who wrote a business plan) and the treatment coefficient is therefore omitted. The last column indicates the number of observations. All regressions are implemented as an ANCOVA specification, controlling for strata fixed effects. A ✓ indicates that the outcome variable is available for all observations at baseline. If it was not collected at baseline, a × is shown, and if it is partially available, missing observations are set to zero and dummied out (✓). Where available, regressions control for baseline levels of the outcome. Standard errors are clustered at the household cluster level. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 8: "Name factors that contribute to business success."

	(1) Treatment mean (SD)	(2) Coordination Intervention	(3) N (BL)
Competitor weaknesses	.3424	.1081***	923
	(.4748)	(.03739)	✓
Business niche	.623	.0213	923
	(.4849)	(.03017)	✓
Aware of competitors	.4875	-.01049	923
	(.5001)	(.03557)	✓

Notes: Analysis in this table is among treated households (who wrote a business plan) and the treatment coefficient is therefore omitted. The last column indicates the number of observations. All regressions are implemented as an ANCOVA specification, controlling for strata fixed effects. A ✓ indicates that the outcome variable is available for all observations at baseline. If it was not collected at baseline, a × is shown, and if it is partially available, missing observations are set to zero and dummied out (✓). Where available, regressions control for baseline levels of the outcome. Standard errors are clustered at the strata level. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

correlates with improvements in almost every reported metric of business performance. While not all of the differences are statistically significant (at the current sample size), they predominantly point in the same direction.

Table 9: Business performance by prior experience

	(1) Control mean (SD)	(2) Treatment Effect (0)	(3) Treatment Effect (1)	(4) Difference p-value	(5) N (BL)
Does business	.4723	.3721***	.4349***	[.05248]*	995
	(.3482)	(.04488)	(.05048)		✓
		[0]	[0]		
Investment in inventory	228,760	1,272,590***	2,072,210***	[.006833]***	995
	(535,752)	(191,560)	(236,743)		×
		[0]	[0]		
∼ productive assets	12,273	44,997***	54,962***	[.2428]	995
	(105,741)	(12,305)	(16,008)		×
		[.001998]	[.003996]		
Hours worked	262.7	11.87	40.01	[.04555]**	995
	(163.9)	(19.74)	(24.29)		×
		[.5944]	[.3207]		
Earnings	755,960	1,025,797***	1,707,396***	[.01823]**	995
	(729,338)	(264,076)	(362,140)		×
		[0]	[0]		
Gross margin	526,793	-254,215	-378,290	[.4915]	995
	(507,894)	(176,662)	(249,046)		×
		[.1688]	[.1608]		
Profitable	.9588	-.3115***	-.353***	[.2894]	995
	(.1991)	(.07225)	(.07726)		×
		[0]	[0]		

Notes: The last column indicates the number of observations. All regressions are implemented as an ANCOVA specification, controlling for strata fixed effects. A ✓ indicates that the outcome variable is available for all observations at baseline. If it was not collected at baseline, a × is shown, and if it is partially available, missing observations are set to zero and dummied out (✓). Where available, regressions control for baseline levels of the outcome. Standard errors are clustered at the strata level. Fisher (1935) exact test p-values from 1000 simulated alternative treatment assignments are shown in brackets. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

4.3 Discussion

A broad but perhaps under-appreciated pattern in structural transformation is that the retail sector is the first livelihood outside of agriculture for many extremely poor people. My paper shows that while recipients express general equilibrium neglect in how they chose their businesses, this cannot have been the only barrier to broader sectoral diversification. The lack of hard skills emerges as a likely next bottleneck.

There are first signs of a divergence in earnings a few months after the cash grant, which come into sharper relief when focusing on business survival and exit. The point estimates on earnings reductions and exit are roughly in line with the cost savings for the implementer from scale economies. A particularly stark sign of overcrowding is that the higher saturation arm induces significant out-migration.

Prices do not appear to be systematically affected by saturation, suggesting that increased demand and supply either cancel each other out or the lower earnings come about by spreading a limited number of sales at fixed prices among more entrepreneurs who are already at a competitive benchmark.

On the whole, these findings cast doubt on the ability to scale an entrepreneurship-focused graduation model through saturation in contexts where a large share of the population is eligible, and without removing barriers to entry into a broader range of sectors. This, however, would require innovation in hard skills training delivery that can provide different skills to different people and thus may not attain scale economies. Apprenticeship programs are an avenue worth exploring in future research.

Appendix of chapter 2

A Additional figures and tables

Table A1: Study population

	(1) Count
Households	
Number of surveys conducted	28,853
Female primary adult present	27,684
U5 present	16,882
No children present	3,254
Individuals	
Population	138,548
Adults	65,162
Children	73,386
U5s	20,998
Number of villages	162
Number of CBOs	25
Number of ADCs	5

Census of all households in Mangochi district's subdistrict ("Traditional Authority") Nankumba. The urban center Monkey Bay and the tourist destination Cape Maclear were omitted from this census as well as an area in the South-West of Nankumba where another NGO is implementing a livelihood program. Area Development Committees (ADCs) are administrative subdivisions of a subdistrict, while Community-Based Organizations (CBOs) are associations of several villages organizing development initiatives or providing local public goods together.

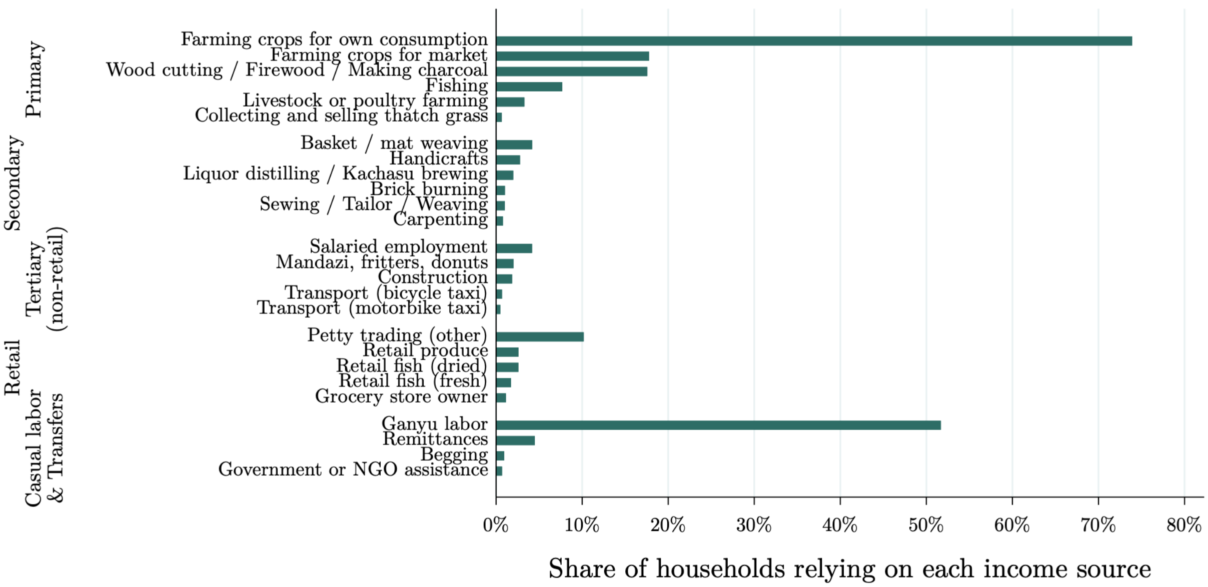


Figure A1: Most common income sources in the study area

Note: This figure describes the income sources of the population as stated during the area-wide census. People are predominantly farmers or casual "'ganyu" laborers. Outside of the primary sector, retail is another common income source.

Table A2: Livelihoods

	(1) Percent
Agriculture	
Crop farming	76.5
Subsistence crop farming	73.5
Marketed crop farming	17.8
Share of yield marketed	29.2
Share of yield marketed (conditional)	69.8
Livestock	3.3
Piece work	
Ganyu	51.7
Exclusively subsistence + ganyu	27.6
Business	
Business	58.0
Employees > 0	4.1
Local non-ag employment potential	3.4
Observations	29084

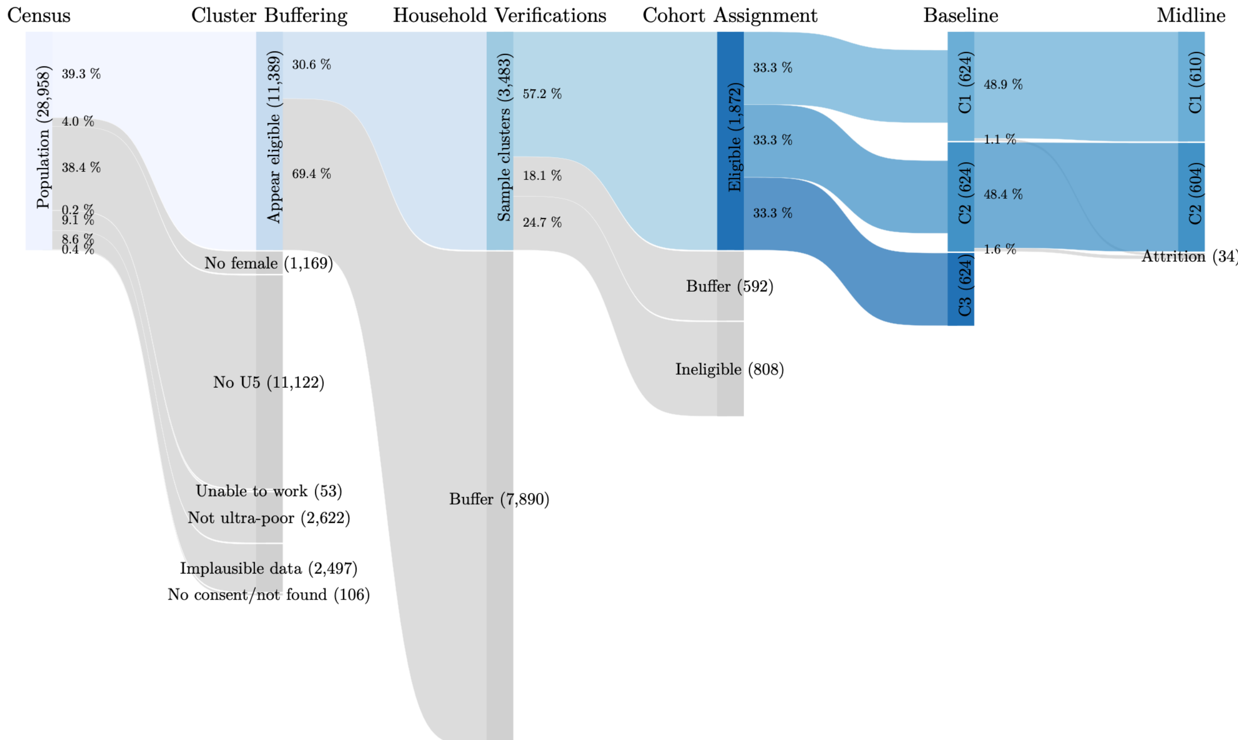


Figure A2: Sampling funnel

Note: This figure describes how the sample is selected from the census population. Starting from the census population of about 29,000 households, about 60% were deemed ineligible, mostly for lack of a child under age five. Among apparently eligible households, 271 clusters of 40-50 households were formed by proximity using *k*-means clustering. A second custom algorithm then selected 72 of these clusters in such a way as to maximize the distance between the nearest two clusters, in order to minimize spillover concerns (cf. Appendix Figure A3). The other clusters serve as buffer zones. Among the selected clusters, household verifications were conducted that resulted in some households being deemed ineligible under the same criteria as in the census, while selecting the 26 eligible households closest to the cluster centroid and excluding the remainder as additional buffer households on the periphery of the clusters. The 72 clusters were grouped into triplets by proximity and assigned an implementing staff. They were also assigned at random to three cohorts to undergo the intervention at different times in the agricultural season. The full sample was surveyed at baseline and the graph shows attrition in the subsequent survey rounds.

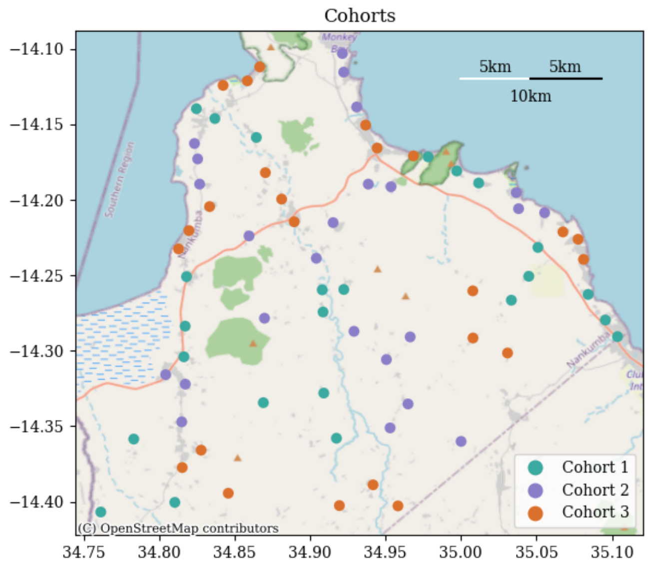


Figure A3: Clusters by cohort

Note: This map shows the 72 study clusters of 26 households each that make up the study sample, color-coded by the cohort they are assigned to.

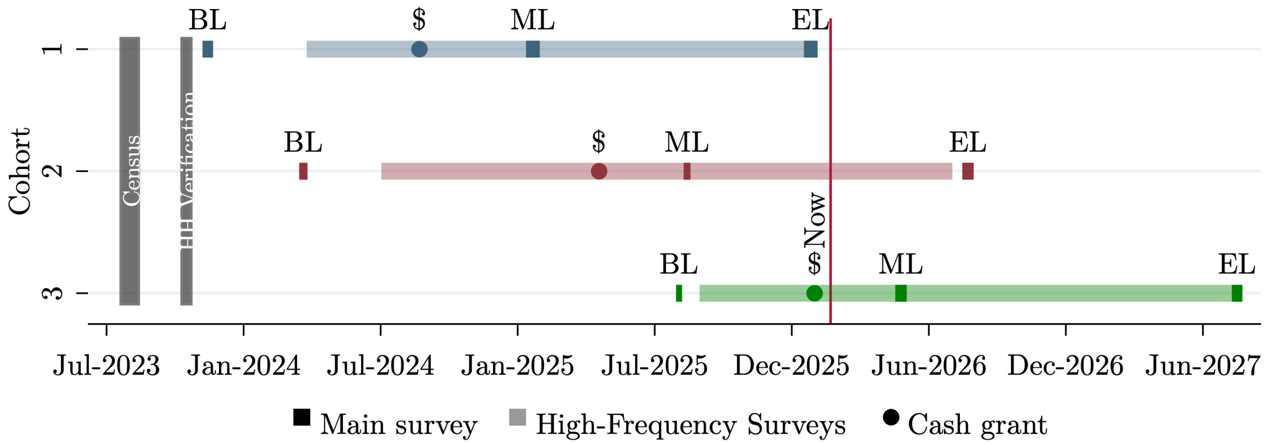


Figure A4: Study timeline

Note: This figure describes the study timeline. A census of the study area was conducted in August 2023. Household eligibility verifications followed in October 2023. The baseline survey for cohort 1 was conducted in December 2023 and high-frequency surveys in this cohort started in April 2024. The large one-off cash grant was paid out in August 2024. The midline survey was collected in January 2025, followed a year later by the endline survey.

Table A3: Poverty

	(1) Percent
Skip meal (past two weeks)	76.8
Female illiterate	37.7
Thatch roof	64.5
No furniture	80.3
No phone	51.9
No bedding	15.7
$<$2.15 PPP 2017 USD (self report)	92.7
$<$2.15 PPP 2017 USD (PMT)	85.1
# children / adult (if $>$ 0)	1.5
# income sources	1.5

Table A4: Balance

	(1) Control Mean (SD)	(2) Treatment (SE)
Household and participant characteristics		
Household size	5.4	-.1175
	(1.8)	(.09726)
Age	32	-.5641
	(8.4)	(.4244)
Years of education	5.7	-.1218
	(2.9)	(.1906)
Literate	.69	.005342
	(.46)	(.02665)
Livelihood and assets		
Hours worked (7 days)	22	-.3643
	(18)	(.7828)
HH income (12 months, MWK)	475,629	11,315
	(480,571)	(47,515)
Total savings, assets, livestock, minus debt (MWK)	256,874	-17,471
	(365,607)	(24,353)
Largest purchase (12 months, MWK)	38,607	-2,816
	(121,528)	(6,344)
Poverty and food security		
Thatch / tarp / no roof	.75	.02137
	(.44)	(.02314)
Food insecurity score (0-8)	7.5	.04915
	(1.3)	(.0641)
Health and non-health shocks		
Health episodes (12 months)	16	-.578
	(13)	(.9197)
Non-health shocks (12 months)	21	-.3771
	(18)	(.8531)
Not pre-specified		
Entrepreneurship knowledge score (0-100)	19	.1714
	(7)	(.4259)
Engaged in retail at baseline	.18	-.007479
	(.38)	(.0236)
Observations		1,872
F		0.58
$p(F = 0)$		.85

This table describes the baseline balance. The main panel reports the control group means and coefficients of the treatment indicator in a regression predicting each baseline outcome variable. At the bottom of the table, an F statistic is reported for a regression in which all outcomes jointly predict the treatment. Only the pre-specified variables are included in the F-test. Standard errors are clustered at the household cluster level and reported in parentheses. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Chapter III

Market Day Scheduling, Market Size, and Rural Economic Development

Abstract Market days aggregate thin demand and supply across space and time, making them the pulse of all agrarian societies. In a natural experiment in Western Kenya, weekly market schedules were set quasi-randomly over the past century. This induced exogenous variation in spatial competition between nearby markets. I find that the resulting differences in market catchment areas have persistent, causal effects on present-day market attendance, population, and nighttime luminosity. Markets in this setting proliferated at a time when attendees were walking and potential gains from schedule harmonization are concentrated among markets that were recently connected to a road.

1 Introduction

In a world with agglomeration gains on the one hand and transport costs on the other, there is a tradeoff between the access to markets and their size. Many small markets, scattered across a geographic area, have the advantage that buyers and sellers can reach markets at low transport cost. The probability of finding a trading partner and transacting at a favorable price, however, is greatly diminished in a small market. Reliably finding customers is vital to any business, but to smallholder farmers and entrepreneurs, it is existential.

I study how this tradeoff was brokered in the context of rural market days in Western Kenya and how that has affected local economic development. Villages in this context are too small to sustain daily trade. Large shares of the consumption basket are perishable, but inadequate availability of refrigeration limits bulk purchases or storage. On any given day, a share of consumers need to purchase, while sellers sit on post-harvest perishable stockpiles, or daily fish catch, and sellers of more long-lived products still rely on regular sales for their subsistence. Rotating, periodic market days[26] have emerged here and across the world as a solution to this challenge.

The periodic "pulsation of economic activity" that market days spark (Hill, 1966) has been described as "one of the basic life rhythms in all traditional agrarian societies" (Skinner, 1964). At their economic core, they solve a coordination problem: Bringing buyers and sellers together in the same place at same time. Rural periodic markets are also the final link in the food supply chain for millions of people. Market days are the site of myriad matching problems and the livelihoods of vulnerable market participants on both sides of the transaction depend on reliably finding someone to trade with at a predictable price. This process is challenging in thin markets with few buyers and sellers and can be eased through market integration that increases market attendance. In this paper, I study an age-old,

[26]Periodic markets either operate on a non-daily, typically weekly, schedule or on a daily schedule with substantially higher attendance on designated market days.

straightforward to implement, and essentially costless way of increasing the attendance to weekly markets: Scheduling market days in such a way that they do not clash between neighboring villages, thus allowing buyers and sellers to cross-attend. I investigate the impact of this strategy on present-day trade patterns: How many more sellers and buyers can a market attract if other markets on the same day are further away, thus increasing the catchment area? Conversely, if two nearby markets are splitting their surrounding area between each other on market day, how much does this hurt their attendance? To answer these questions causally, I identify a natural experiment in how market schedules were set in Western Kenya over the past century, which allows me to show that weekday scheduling in the distant past meaningfully impacted present-day market attendance by shaping market catchment areas.

I measure a market's catchment area as the area within which it is the nearest market to reach on its market day. A catchment boundary is the line along which two markets that share a market day are equidistant (Voronoï, 1908a,b). Catchments are thus fully determined by only two factors: The locations and schedules of all markets. Ignoring other determinants of market attendance, such as population density within the catchment or transport infrastructure is not a bug but a feature of my analysis. Over the course of a century, these are endogenous to market locations and schedules. By focusing on the spatio-temporal primitives of market catchment areas, I am restricting my analysis to a proxy of a potentially richer measure of market potential, but one for which I can pin down sources of potential bias, which I can remedy.

A simple regression analysis that relates outcomes of interest to market catchment area would not identify its causal effect. Geographic fundamentals or initial conditions (e.g. in population density) may both endow some markets with a larger catchment area while simultaneously affecting the outcomes we are interested in. I provide evidence of this threat in Section 5.3. My goal in this study is to resolve this bias and arrive at the causal effect of market catchment area.

An in-depth literature review in anthropology, sociology, geography, history, and economics, as well as over 1900 key informant interviews with village elders and market administrators in the study area allow me to reconstruct the market scheduling process for all 280 weekly markets in my market census. Market day scheduling was a function of the distance to existing markets at the time of market inception, region-wide weekday preferences, as well as idiosyncratic preference shocks, such as pre-existing village meetings or the weekday on which the butcher slaughtered. Using the method proposed in Borusyak and Hull (2023), I leverage this contextual knowledge to separate the observed variation in market areas into a component that was to be expected given the known geography and scheduling process, and a component of quasi-random deviations from this expectation which are the culmination of a cascade of idiosyncratic weekday preference shocks. In the analysis that follows, I use this latter component to identify the causal effect of market area and frictions on various outcomes of interest. In practice, rather than compare markets with small and large catchment areas, I compare those that we would have expected to have the same size of market area, was it not for their own and their neighbors' idiosyncratic weekday preferences on their part as well as those of all of their neighbors. The high present density of markets means that eventually a market can be reached anywhere in Western Kenya on any given day of the week and all markets face competition on the same day. The identification assumption here is not that markets chose their market day at random but rather that after accounting for the scheduling process, exactly who clashed with whom and how big a market catchment area that implied is as good as random.

I find that, indeed, distance to existing markets predicts market day scheduling decisions and that simulations from the estimated scheduling model can be used to purge out imbalances in predetermined market characteristics. The quasi-random component of the scheduling decision model is then used in the remainder of the analysis. I show that the variation in market area that can be attributed to idiosyncratic weekday preferences persistently and causally impacts market attendance by both buyers and sellers and that this is driven by outsiders cross-attending from other towns. I also show that market towns that are endowed with a larger market area than the scheduling model would have us expect have a larger population today and are brighter in nighttime satellite imagery, a proxy for economic activity.

Finally, I discuss the tradeoff between travel cost and agglomeration gain that these markets are brokering. Consider drawing market catchment areas with the objective to maximize agglomeration gains subject to travel cost. One would make catchment areas so large that those furthest from the market are just made indifferent about attending the market. Those even further away would ideally be part of another market's catchment. My results suggest that a sizable number of people live in boundary areas where more than one market is competing for their attendance in a context where we are worried about markets being too thin. What's more, market schedules have at this point been mostly set and highly persistent for over 50 years. Markets in this context largely predate transport infrastructure and I document that potential gains from schedule harmonization are driven by markets that have recently been connected to a tarmac road and now face lower transport costs. If market catchment areas were optimal before the construction of most road infrastructure, my results suggest that trade activity today is scattered across suboptimally many markets and potential agglomeration gains are being left on the table.

My paper contributes to four strands of literature on (a) periodic markets, (b) market integration, (c) the tradeoff between agglomeration and market access, as well as (d) economic persistence.

Periodic markets: Food markets in many developing countries are thin (Fafchamps, 1992) and thin markets are volatile (Desmet and Parente, 2010): With few buyers and sellers on either end of the transaction, supply, demand, and prices are likely to fluctuate as match rates are low. This makes it difficult for both buyers and sellers to reap the full potential of market participation. Market integration can improve the performance of price systems and the reliability of markets to all participants (Barrett, 2005; Jensen, 2007; Svensson and

Yanagizawa, 2009; Casaburi et al., 2013; Faber, 2014; Donaldson, 2015 and others) and ease matching (Agarwal et al., 2019; Akbarpour et al., 2020; Kapor et al., 2022). In this paper, I study a straightforward means of market integration aimed at increasing market attendance to periodic markets: Setting market day schedules with the goal of allowing buyers and sellers to cross-attend several markets is a widely recognized phenomenon in anthropology, sociology, geography, and history (Schurtz, 1900; Thomas, 1924; Fröhlich, 1941; Hill, 1966; Hodder, 1961; Fagerlund and Smith, 1970; Good, 1973; Wood, 1974b; Bromley et al., 1975; Scott, 1978; Smith, 1979, 1980; Geertz, 2000). Interest for them in the social sciences peaked in the 1970s, before the availability of modern satellite data, detailed maps of virtually every corner of the Earth, or the computing power required for many recent GIS and statistical methods. These earlier papers tend to be small-n case studies focusing on origins rather than impacts of market days. Despite their importance, periodic market days have not attracted broad attention from economists. I revive and build on this rich literature, bringing to bear newly available data and methods, while also departing from the purely descriptive and explanatory scope of earlier work with an attempt at assessing the causal impacts of periodic market schedules.

Market integration: A number of creative and insightful studies in economics have addressed agricultural markets. Fafchamps and Hill (2005) investigate the role of transport costs for coffee farmers in Uganda in deciding whether to travel to and sell at wholesale markets themselves or whether to accept the lower prices offered by aggregators who come to their farm gate. Jensen (2007) is the classic reference for market integration through information sharing using mobile phone technology and a lot of progress has been made since (Svensson and Yanagizawa, 2009; Aker, 2010; Goyal, 2010; Jensen and Miller, 2018). In a study of the road infrastructure connecting the major agricultural markets in Sierra Leone, Casaburi et al. (2013) study market integration through the reduction in travel times and what its effect on crop prices reveals about the competitive structure in these markets. None of the aforementioned studies discuss market attendance as an outcome. In a creative study on Kenyan markets, Bergquist and Dinerstein (2020) investigate the competitive structure of sellers and intermediaries by exogenously inducing firm entry. While large gains in market integration will certainly be reaped from modern information technology or physical infrastructure, it should be noted that none of these have existed nearly as long as the simple act of synchronizing schedules, which limits our ability to assess their effects in the very long run, and that the latter, if timed at the inception of a new market, can be done essentially costlessly. It should also be noted that the cited studies have focused on markets that tend to be around an order of magnitude larger than the ones I discuss in this paper. While these larger wholesale markets are crucial for the functioning of country-wide food security, the smaller, last-mile markets are the direct touch point of millions of people around the world with the food supply chain, either as consumers who rely on them for their daily sustenance, or as smallholder farmers in an early stage of graduating from subsistence agriculture to marketization – or in fact as both simultaneously (Smith, 1980). These smaller markets outnumber the large wholesale markets many dozen if not hundreds to one.[27] Market periodicity and schedule frictions have not, to my knowledge, been discussed in the economics literature, despite their presence in virtually every agrarian society and their primacy as a locus of exchange prior to industrialization.

Optimal market size: The optimal size and scope of markets, as well as whether optimality is reached in equilibrium, has been the subject of a longstanding debate in economics (Hotelling, 1929; Tóth, 1953; Lösch, 1954; Mills and Lav, 1964; Stern, 1972; Bollobás and Stern, 1972). Oberfield et al. (2024) study the optimal number, location, and distance between plants of a single firm as a trade-off between transport cost and plants cannibalizing one-another's customer base. In their model, the overarching firm chooses the locations of all its plants to maximize joint profits, while in my case – and perhaps even in some of their own examples of franchise branches – the decision is decentralized without a planner enforcing an optimal allocation. Recent advances by Vitali (2022) and the work cited

[27]Of the several hundred market places I sampled in Siaya County, only Ramba Market in Siaya Town would be considered a wholesale market and it in turn is substantially smaller than any of the markets for which Burke et al. (2019) present seasonal crop price fluctuation data.

therein have highlighted the advantages for firms, even those in similar industries, to spatially cluster, rendering it natural to recast the classic location tradeoff into one for markets rather than firms. In this context, the sequential nature of market entry in my paper may also be of interest to readers thinking about how classical static equilibrium concepts can be brought to bear in a dynamic world. Starting with Dixit and Stiglitz (1977), a plethora of models based on a CES functional form, notably among them Melitz (2003), obtain notions of static equilibrium in which differentiated products in product or quality space, or firms/towns in geographic space end up in some form of equidistance from one another. Lösch (1954) and later Stern (1972) warn that the sequential nature of entry is unlikely to always result in such an efficient equilibrium spacing and my paper provides an example of that for the case of marketplaces: Later entrants may undo the schedule optimality of existing markets. In addition, if fundamentals such as transport technology change, a frictionless market system would have markets gradually move further and further apart from one another, but in reality markets are tied to the amenities of the immobile towns that host them.

Persistence: There is important path dependence in economic activity, trade patterns and locational economic trajectories (Bleakley and Lin, 2012; Faber, 2014; Storeygard, 2016; Jedwab and Moradi, 2016; Donaldson, 2018; Hornbeck et al., 2024). Wahl (2016) documents a strong role for medieval trade centers in the emergence of modern European cities. Market days are no exception when it comes to persistence (Bromley et al., 1975) and I document that over the past century market schedules in my study area remained mostly unchanged. The clearly defined onset of periodic markets, their absence before the arrival of the British, and their persistent organization and effects contribute to the persistence literature.

This paper proceeds as follows. Section 2 reviews the literature about periodic markets in general and in Western Kenya in particular. This section makes the case for the study context being a suitable natural experiment. In section 3, I introduce the data collection exercise, describe the various outcomes of interest, and provide descriptive evidence to support conclusions from the literature review. Section 4 describes the identification argument and empirical strategy in more detail. Section 5 discusses results and section 6 concludes.

2 Background

2.1 Markets in Western Kenya

Western Kenya, which is the focus of this paper, did not have periodic markets until the early twentieth century. The region was "on the whole [...] outside of the major nineteenth century trade routes" and economic exchange was limited to sporadic barter trade sparked by encounters between pastoralist (Luo) and farming tribes (Luhya), dictated by specialization and seasonal fluctuations in livelihoods (Hay, 1975). Around the turn of the century, not merely weekly markets, but *the concept of the "week" itself* was introduced to the region (Wood, 1973a). The local tribes did not traditionally have an intermediary unit of timekeeping between the solar day and the lunar month. It was imposed jointly by a coalition of Christian missionaries and the British colonial administration[28] (Dietler and Herbich, 1993). In order to organize Christian worship, the missionaries wanted the local population to adopt the seven-day week so that they would know on what day of the week church service was to be attended. The British colonial administration had known periodic rural market days both in Europe – the remnants of which we know as farmers' markets – and in their colonies in West Africa, where periodic markets were substantially more common than in East Africa (Good, 1973) and where periodicities other than the seven-day week had emerged long before colonialism[29]. The colonial administration's goal of

[28]Present-day Kenya was allocated to the British Empire in the Berlin Conference of 1885. It was declared a British Protectorate in 1895 and the Uganda Railway, which allowed the White settlers to make inroads into Western Kenya, was completed in 1901. It connected Mombasa, an important stopover port for Indian Ocean trade routes and the location of the first Christian Mission since 1844, with Kisumu, the regional capital of Western Kenya on the shore of Lake Victoria and already a major settlement at the time. Kenya was declared a British Colony in 1920 and gained independence on 12 December 1963.

[29]Unlike the solar day and the lunar month, the week has no corresponding rhythm in nature and is a human-made concept to impose structure on activities that should occur with an intermediate frequency. Zerubavel (1989) is an excellent reference for its history. The seven-day week appears to be of Babylonian origin and, after entering Judaism through the book of Bereshit/Genesis, it was spread around the world by

marketization was to encourage division of labor and specialization to lift the local population out of subsistence agriculture and livestock rearing, turn them into more affluent trade partners, and to concentrate the resulting trade in a specific place and time in order to levy taxes. Only Central and Western Kenya appear to have had sufficient population density and economic activity to sustain markets (Wood, 1974a). Muslim influence, while present on the Swahili coast for over a millenium, only reached Western Kenya riding on European coattails in the 1880s (Trimingham, 1980 p. 32, 38). Both markets and their periodicity were thus introduced about a century ago by outside forces, thus marking an approximate start date for this natural experiment.

3 Data

3.1 Market mapping, history, and attendance

Markets, especially smaller ones, are not presently well mapped in Kenya. As part of my data collection, my implementing partner REMIT conducted a census of all markets in Siaya, Busia, and Kakamega County in Western Kenya[30]. From scoping activities, it became apparent that the most knowledgeable key informants on market and village history tend to be Village Elders, Market Chiefs/Chairpersons, and Market Incharges. The informational advantage of local leaders is increasingly recognized in the economics literature (Basurto et al., 2020; Balan et al., 2022) and their role has found renewed appreciation across the continent post-independence (Baldwin, 2016). The initial sampling frame comes from

the more expansionary abrahamic religions of Christianity and Islam. West Africa had and to some extent still has weeks of other periodicities. In contrast to Western Kenya, where the religiously motivated seven-day week dictates market periodicities, the week in West Africa is conjectured to be itself a product of the need to structure trade periodicities (Schurtz, 1900; Thomas, 1924). The outside impetus for market creation and the introduction of a previously unknown weekly periodicity in Western Kenya is crucial for my research design.

[30] The quality of the results of any study suffers when data are missing. This study measures a market's catchment area as a function of the schedules of all of its neighbors. Measurement error in one market therefore spills over into measurement error for all markets. Missing a market that clashes with a neighbor schedule makes surrounding markets appear better coordinated than they really are. Missing a market that does not clash with any neighbors but where two other neighbors clash with each other makes it look like there would have been a way to avoid a clash that was not being used. It is therefore critical to not just obtain a sample but a full census of markets for this study.

triangulating two listing exercises: Villages are part of two parallel administrative structures. Each village is part of a sublocation with an Assistant Chief, which is part of a location with a Chief, which is part of a subcounty with a Deputy Commissioner who reports to the County Commissioner. At the same time each village is part of a ward which has ward administrators and the wards are in turn part of the same subcounties. The location-sublocation structure is determined by Kenya's central government and structures executive power. Finding markets this way avoids missing small markets, but risks duplication if several villages share the same market. The ward structure is determined by the county government and structures taxation and public service delivery. Conducting a census of markets through this structure may miss small markets where no fees are being collected, but is unambiguous when it comes to uniquely identifying markets. Starting from the top of these two parallel hierarchies, we obtained contacts for every subcounty Deputy Commissioner from the County Commissioner. From those we worked our way down to the location Chiefs and from them to the sublocation Assistant Chiefs who gave us the contacts of Village Elders in each village in their sublocation. Starting from the County Revenue Inspector, we went to the Subcounty Revenue Inspectors who referred us to the Ward Market Incharges who shared contacts for Market Incharges and Market Chairpersons. The majority of our key informants are from these groups, although occasionally, we were referred on to other village members who offered additional insights. Every key informant was also asked to mention neighboring markets to us, ideally with contacts, and the team followed up on every one of those comments to achieve full coverage of all markets. The census was conducted over a period of 21 months in 2023 and 2024.

Geolocations for each market were obtained from Google Maps or Open Street Map (OSM) where available and otherwise an enumerator would visit the village and geocode it. Throughout, whenever we found markets that were not yet mapped, we documented them on OSM, while Google Maps appears to be automatically ingesting our OSM contributions, thus making this public good available much more broadly. In phone surveys, we collected key dates in each market's history, most importantly when it was started and on what day of the week it is being held. If markets operate on several days of the week, we ask whether there is a main market day. I document very strong persistence in weekdays to the present day, even for markets that have been around for a century. I also obtain an estimate of present-day market attendance as well as how this attendance is decomposed into sellers and buyers from within or outside the village and on the market's main versus any other day (cf. Appendix Figure A1). In order to guard the data collection against missing some markets entirely, every informant is also asked to list other markets in their area and to provide the contacts of knowledgeable informants. As Appendix Table A2 illustrates, the majority of respondents received their information either first-hand as they were alive at the time the market was started or second hand and only for very few markets does my analysis rely on information that was passed through more than one social network link. Village elders spend years in the company of previous elders, preparing for their role and learning about the local history of their town. Written records are rare in this context and practically non-existant for the pre-colonial era. In collecting these data, we thus also contribute to the preservation of the oral history of Western Kenya. These data form the basis for the empirical strategy described in section 4.

Two papers by von Carnap (2022, 2023) are closely related to mine. Using an innovative remote sensing methodology, he exploits the weekly pulsation in activity to detect weekly markets across sixteen Kenyan counties. He relates their location to historical records of marketplaces compiled by Wood (1973b). The natural experiment I describe was largely completed and market schedules established at that time.[31] Reassuringly – and using fundamentally different methodologies and data sources –, our two studies align closely on the finding that historic market centers are highly predictive of present-day population

[31]For about two thirds of these historical markets, von Carnap (2023) does not find activity that would be detectable in present-day satellite imagery, which I attribute to the remote sensing methodology being better suited to detect larger markets. Having a data collection team on the ground allows me to be somewhat more confident in having obtained a complete market census than either of these papers and in the counties where we overlap I find over twice as many weekly markets as von Carnap (2023), including virtually all of the 1970s markets (though some of them are small daily trading centers without a market day). The discrepancy lies mostly in the smaller markets being undetectable on satellite data.

patterns. I demonstrate that this link is causal.

3.2 Outcome data and pre-treatment controls

Ultimately, this study is interested in how market catchment area affected market attendance and rural economic development. Outcomes of interest for my study include village population size and night lights as a proxy for local economic activity. Appendix Table A1 provides more details.

Pre-colonial Western Kenya did not have a formalized administration that would have recorded social or economic data in writing. Pre-treatment controls are therefore restricted to geographic controls (location, agricultural suitability, terrain ruggedness, proximity to Lake Victoria, proximity to the regional capital Kisumu, see footnote 28) as well as an estimate by the key informant of the village population in 1920.

3.3 Descriptive evidence

Markets may be open every day of the week, but they tend to be substantially larger on market days (Fröhlich, 1941). In my study area, the majority of markets operate every day, as shown in Appendix Table A3. The first map in Appendix Figure A2 shows all markets that have a market day in the study area along with the days on which they operate, while the second map shows only the main market day. Market days are distributed across days of the week and there appears to be a preference for some days over others (Appendix Figure A3). A model of market scheduling should therefore allow for weekday fixed effects.

Markets in my sample sport slightly fewer sellers than the hosting village has households. Buyers outnumber sellers more than two to one. These averages mask important heterogeneity. Markets that have a market day are much larger than those that do not (Appendix Table A3) and market attendance is substantially higher on market days than on non-market days (Appendix Figure A1). The figure illustrates the pulsation that Polly Hill (1966) described: While attendance by locals to their own market roughly doubles or triples on market day, some markets see a five, ten, or even twenty-fold increase in outside attendees.

Anecdotally, from the conducted surveys and scoping interviews, market participants link lower attendance at their own market to the emergence of nearby markets on the same day of the week. I therefore inquire about the scheduling process for these market days. Historical sources suggest that while the Europeans pushed for the introduction of markets following a seven-day periodicity, they left the precise scheduling to local authorities (Wood, 1973a) – a fact I confirm for the study area. Market days were initially set by the Village Elders, local authorities, and local stakeholders (Appendix Table A4) with a *coordination motive* and with knowledge of already existing nearby markets, though in some cases local idiosyncrasies led to the adoption of particular days or they emerged organically (Appendix Table A5). Notably, these concern pre-existing village-internal community meetings ("chief barazas") or local customs such as the weekday on which the local butcher used to slaughter. There is no evidence that higher levels of government, the church, or another social planner-type actor held any sway over the schedule. We can therefore think of the scheduling process as decentralized, which introduces the possibility of inefficiencies.

Once set, market schedules and activities have been remarkably persistent (cf. Appendix Table A3). Only a small share of markets have ever seen a change to their schedule and of the few observed shifts, some are accounted for by adding a day or dropping one of several. Survey respondents also indicate that few if any markets have ever been interrupted in their activities for any meaningful amount of time, especially for reasons other than the COVID-19 pandemic, which does not appear to have caused lasting closures or schedule shifts.

Markets were first started in the colonial period with an acceleration around Kenyan Independence in 1963. The number of markets with a main market day has been largely stable since 1975 (cf. Appendix Figure A4). In evaluating the effects of market catchment area, this study thus takes a long-term view of the cumulative effects over the past century, where the regional market schedule was mostly set over half a century ago. Over the same period, the average distance from points in the study area to the nearest market fell sharply and stabilized over the past half century.

4 Identification and empirical strategy

The goal of this paper is to estimate the causal effect of market catchment area differences that arose from scheduling frictions over the past century on present-day market attendance and, in turn, on rural economic development today. If the area over which a market is the nearest operating one on its market day is larger, how many more sellers and buyers can it attract? If by contrast, two nearby markets are splitting their surrounding area between each other on market day, how much does this hurt their attendance? Are other measures of economic development affected by this? I will argue that scheduling frictions and therefore the market catchment areas on a given day are at least in part the product of chance. Yet, simply regressing outcomes of interest on the size of the market catchment area would be biased if omitted factors account for both a larger market catchment and a better market performance. This entails both initial conditions like geography or population density as well as selected entry if towns near an existing large market are deterred from starting their own market. I present evidence of such bias in Section 5.3.

The following regression equation illustrates how the bias arises. It relates market area A_l in a location l to some outcome Y_l where ε_l is an error term that contains other factors such as geography or lagged values of Y_l that are potentially correlated with A_l:

$$Y_l = \beta_0 + \beta_{\mathrm{A}}\mathrm{A}_l + \varepsilon_l, \qquad \text{where } \mathbb{E}\left[\sum_l \mathrm{A}_l\, \varepsilon_l\right] \neq 0. \tag{7}$$

The coefficient β_{A}, which I seek to estimate, is the causal effect of market catchment area differences that arose from scheduling frictions on outcomes of interest. Estimating equation (7) with simple OLS would yield a bias because of the omitted variables in ε_l that are correlated with market area.

The contextual insights from the key informant interviews and literature review in anthropology, sociology, history, and geography allow for a design-based approach to resolving this causal inference challenge (Abadie et al., 2020). It involves separating the scheduling process into its endogenous, fixed-by-design, as well as its exogenous, quasi-random component. As evidenced in Appendix Table A5, market scheduling (which produced today's market catchment areas A_l) was a function of the schedules and locations of only those nearby markets that already existed when a new market entered (the known geography $\Delta_{t(l)}$ that was in place when l entered at time $t(l)$) as well as some local idiosyncratic preferences by market stakeholders (ξ_l).

$$\mathrm{A}_l = f_l(\xi_l, \Delta_{t(l)})$$

The following three assumptions allow me to exploit the scheduling process as a natural experiment for causal identification of the effects of market catchment area, effectively holding $\Delta_{t(l)}$ fixed and pinning down changes in A_l only as a function of ξ_l:

Assumption A1: Market scheduling decisions were made without foresight over the entry or scheduling choices of future entrants. In other words, scheduling choices are made relative to the geography $\Delta_{t(l)}$ that existed at the time $t(l)$ at which market l was started.

Assumption A2: Preferences over weekdays ξ_l are independent of other factors contained in ε_l and therefore affect outcomes of interest only through the variation in market area that results from them: $\xi_l \perp\!\!\!\perp \varepsilon_l$. Concretely, I parameterize idiosyncratic weekday preferences as iid draws from a Type-1 Extreme Value distribution (one for each market and weekday): $\xi_l = (\xi_l^{\text{Monday}}, ..., \xi_l^{\text{Sunday}}) \overset{\text{iid}}{\sim} \text{T1EV}^7$

Assumption A3: Market entry order is fixed and does not depend on the schedule of existing ones. Suppose we label markets $l = 1, ..., L$ in the order in which we observe them start in the data, i.e. $t(1) < ... < t(L)$. **A3** says that $(t(1), ..., t(L)) \perp\!\!\!\perp (\xi_l, \Delta_{t(l)})$. In a counterfactual world in which market day scheduling may have played out differently, the order of entry would have been the same. If some markets have a large market area, because their presence deterred the entry of potential nearby competitors, this assumption ensures that this selection bias is held constant in any counterfactual world.

Taken together, these assumptions produce the following decision problem: Markets enter sequentially in their observed order (**A3**). Every market chooses one of seven possible weekdays to hold its market. Market stakeholders maximize the distance to other existing markets that are already happening on the same day (but have no foresight over the identity or actions of potential future entrants, **A1**) and on idiosyncratic local preferences (**A2**). Concretely, a market l chooses a day $d \in \{1, ..., 7\}$ at time $t(l)$ with idiosyncratic location-day-specific preference ξ_{ld} (**A2**) and a distance to the closest existing market on a given day d, $\Delta_{ldt(l)}$ and weighted by some utility weight δ_Δ. The term δ_d allows the T1EV distributions to have different means, should particular weekdays be overall more popular than others. The time subscript captures the fact that entry order matters and the establishment of new markets changes the distances and schedule clashes considered by later markets. Entrant markets maximize the utility from each day d

$$U_{lt(l)}(d) = \delta_d + \delta_\Delta \Delta_{ldt(l)} + \xi_{ld} \tag{8}$$

and present-day market area A_l is computed from the resulting schedule $\{\arg\max_d U_{lt(l)}(d)\}_{l=1}^{L}$. **A2** imposes logit substitution patterns, meaning that if a particular day of the week were no longer available for a market, those markets would substitute to other days according to the overall market day distribution. I test this with hypothetical questions to respondents in Appendix Figure A3 and cannot reject that the observed and substitution distribution are identical.

Borusyak and Hull (2023) show that knowledge of $f(\xi_l, \Delta_{t(l)})$ suffices to decompose $\mathrm{A}_l = \mathbb{E}[\mathrm{A}_l] + \widetilde{\mathrm{A}}_l$ – the expectation is taken over ξ_l – in such a way that the expected market area has the exact same correlation with the error term ε_l as A_l:

$$\mathbb{E}\left[\sum_l \mathbb{E}[\mathrm{A}_l]\ \varepsilon_l\right] = \mathbb{E}\left[\sum_l \mathrm{A}_l\ \varepsilon_l\right]$$

This result implies that adding $\mathbb{E}[\mathrm{A}_l]$ to equation (7) resolves the omitted variable bias. Unlike in the lead case discussed by (Borusyak and Hull, 2023), I need to first estimate the distribution from which the ξ_l are drawn by estimating the hedonic utility function (8).

Figure 1 builds intuition for the identification strategy in this natural experiment. Suppose eight villages, depicted by nodes, start markets sequentially ($t = 1, ..., 8$) and decide on what day of the week (indicated by colors) to hold their market day. In the top four panels, we observe the first seven villages choosing their market days in a certain way and the eighth village finds it optimal to host its market on Wednesday, since it is relatively further away from the next Wednesday market than from any other market. In a world where attendees from surrounding villages attend their closest market, the eighth market would capture the attendees from most of the other villages on its Wednesday market day. The shading indicates the market catchment area of each market. In the bottom four panels, the same scheduling process unfolds with one small deviation: Village 5 chooses Wednesday instead of Tuesday, perhaps because the local chief was traditionally holding a village gathering on this day, and village 8 therefore subsequently finds it optimal to host its market on a Tuesday. The resulting cross-attendance patterns on Wednesdays look very different in this counterfactual world. Attendance to the market in village 4 would be larger, driven by a larger catchment. In this paper, I leverage variation in trade patterns caused by small and arguably quasi-random perturbations like this and I provide evidence of their role for scheduling in Appendix Table A5. If most counterfactual scheduling outcomes look more like the top panel, then under the realized scheduling pattern in the bottom panel, a place like village 4 suffers less strongly under the spatial competition from neighboring villages than the fundamentals of geography and the scheduling process would have us expect. This difference between how large we would have expected the catchment area to be versus how large it turned out to be is the treatment in this natural experiment. The effects of that (if any) should then be detectable in market attendance and other measures of rural economic development.

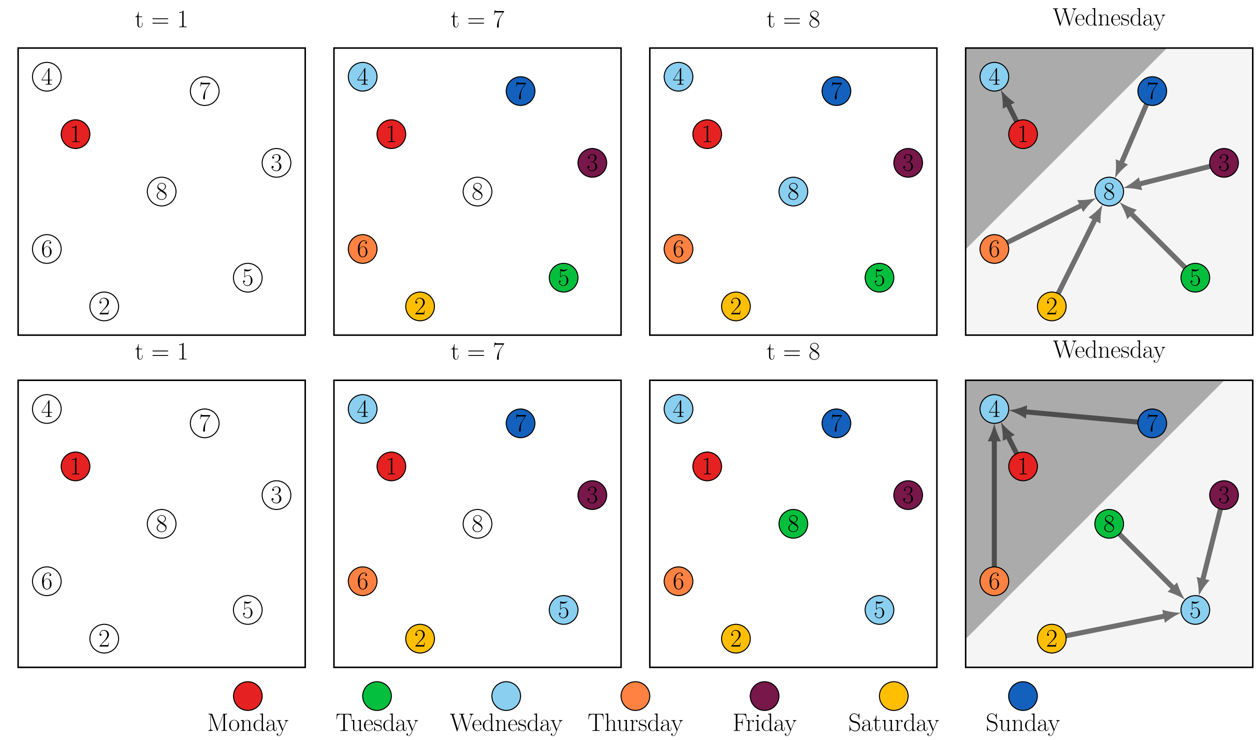


Figure 1: Counterfactual scheduling

Each node is a town. They sequentially choose market days in the order of their numbering ($t = 1, ..., 8$). In the top four panels the last town to pick a market chooses to host its market on a Wednesday in order to avoid clashing with its close neighbors. This makes it the closest Wednesday market for the majority of towns. The bottom four panels show that a small local deviation in the scheduling process (town 5 choosing Wednesday over Tuesday) changes the market cross-attendance pattern, indicated by arrows. The shading indicates each market's catchment area.

My empirical strategy proceeds in six steps, described in more detail below and simulated in Appendix B:

1. Measure present-day market areas.

2. Reconstruct the decision problem that each market town faced when starting its market (Figure 1, bottom).

3. Estimate the hedonic utility function in equation (8).

4. Simulate many counterfactual scheduling processes, measure, and average their resulting market catchment areas (Figure 1, top).

5. Following Borusyak and Hull (2023), control for each market's *expected* market catchment area. The remaining variation in market catchment area becomes the treatment assignment.

6. Estimate the effects of market catchment areas on present-day market attendance and rural development more broadly after purging out bias by controlling for expected market catchment areas.

4.1 Measure market catchment areas

To operationalize the notion that a market schedule might affect attendance, I formalize what the geography literature calls market area (Skinner, 1964; Hodder, 1965; Wood, 1974a,b). I compute the area of the map for which a particular market is the closest market *on the weekday on which this market holds its market day.* This can be thought of as a geographic measure of market potential. Formally, call Q the set that contains all points q_k in the two-dimensional plane that constitute the study area. The set $P_d \subset Q$ contains the locations of all market places p_{ld} that have their main market day on day d. A market area A_l is defined as the set of points q_k that are closer to p_{ld} in Euclidean distance than to any other market p_{dm} that holds its market day on the same day d:

$$A_l = \{q_k \in Q \ \mid \ ||q_k - p_{ld}||_2 \leq ||q_k, p_{dm}||_2 \ \forall \ p_{dm} \in P_d, m \neq l\}$$

In practice, it is calculated as a day-specific Voronoi tesselation (referred to in different contexts as Descartes, 1664; Lejeune Dirichlet, 1850; Voronoï, 1908a,b ; or Thiessen, 1911 polygons), up to a distance of 20km, and removing the area covered by Lake Victoria and Uganda.

Arguably, a more economically meaningful measure of the potential trade activity that a market can capture would be the population of the catchment, not its area. However, there are two important reasons I prefer this area measure: First, there are no reliable estimates of population density prior to market establishment that would be granular enough for this exercise and present-day population density patterns may themselves be a product of the treatment I am studying, i.e. they are endogenous to the process under study through migration toward market towns over the past century or varying fertility and child survival. Second, the purely spatial definition of catchment areas makes it straightforward to identify and deal with sources of bias. The catchment area of each market depends purely on the location of each of the other markets (who entered) and the day of the week they entered on. My understanding of scheduling process allows me to reconstruct alternative, plausible schedules under the same entry process which thus share the same selected entry and omitted variable bias. This constitutes an important advantage over more complex measures of market catchment area.

4.2 Reconstruct scheduling process

Following the distributional assumption for the error term in **A2**, I implement the decision problem of maximizing the above utility as a logit across the seven days of the week, accounting for alternative-specific characteristics (distance to other markets vary by the day of the week). The day-specific intercept allows for the possibility that some days of the week might be inherently more appealing for markets than others. I estimate seven parameters of the hedonic utility function in equation (8): Six day-specific intercepts $\hat{\delta}_d$ (relative to the seventh) as well as $\hat{\delta}_\Delta$ which translates distance to a competing market on the same day relative to those on other days into utils. Only markets within a travel distance of 20km are taken into consideration here.[32]

[32] Otherwise, as the map of markets grows unboundedly, very distant market schedules (which would be highly leveraged in this estimation) would have arbitrary sway on the estimated relationship.

4.3 Simulate counterfactual scheduling outcomes

Using the parameters $\hat{\delta}_d$ and $\hat{\delta}_\Delta$ estimated above, I simulate how towns *could have scheduled* their markets, by redrawing iid Type-1 EV preference shocks (**A2**) for each town and facing them in their same order (**A3**) with their respective decision problems under the new idiosyncratic shocks. After each decision τ, the distance measures $\Delta_{ldt(l)}$ for all future entrants l for whom $t(l) > \tau$ are updated based on the last decision. Decisions take the current distances as given without foresight of future entrants or their scheduling decisions (**A1**). When all market days are assigned I calculate the catchment area size from section 4.1. I repeat this process 1,000 times.

4.4 Recentered market area, counterfactual validity, and baseline balance

Following Borusyak and Hull (2023), I average the market area I obtain across the 1,000 simulations described in section 4.3 at the market level. These can be interpreted as the expected market area for each market $\mathbb{E}[\mathrm{A}_l]$. Any omitted variable and selection bias that may have been present in the observed market locations and schedules is held constant across all simulations, so deviations from the average are driven purely by the variation stemming from location-specific week day preference shocks. Borusyak and Hull (2023) suggest that under a correctly specified shock assignment process, the slope of a regression of A_l on $\mathbb{E}[\mathrm{A}_l]$ should be close to 1 and the intercept close to 0, which I confirm in Appendix Figure A5. In order to scrutinize whether the residual A_l is indeed quasi-random conditional on $\mathbb{E}[\mathrm{A}_l]$, I further check baseline balance of pre-treatment control variables, discussed in section 3.2, using the below OLS specification:

$$X_l = \theta_0 + \theta_{\mathrm{A}}\mathrm{A}_l + \theta_E \mathbb{E}[\mathrm{A}_l] + \varepsilon_l$$

4.5 Effect of market catchment area on attendance and economic development

After controlling for the expected market area $\mathbb{E}[\mathrm{A}_l]$ and pre-treatment control variables $\mathbf{X}$, the coefficient γ_A in equation (9) captures the effect of only the observed market area's random component in how the market scheduling actually played out on our outcome of interest. In other words, rather than compare markets with small and large catchment areas, I compare those that we would have expected to have the same size of market area was it not for idiosyncratic weekday preferences.

$$Y_l = \beta_0 + \beta_\mathrm{A}\mathrm{A}_l + \beta_{E|a}\mathbb{E}[\mathrm{A}_l \mid \text{area}] + \beta_{E|d}\mathbb{E}[\mathrm{A}_l \mid \text{distance}] + \beta_X X_l + \varepsilon_l \tag{9}$$

Conditional on $\mathbb{E}[\mathrm{A}_l]$, the treatment in this natural experiment is assigned at the market level, which is therefore the level at which I cluster standard errors (Abadie et al., 2017). The concern of outcomes being spatially correlated suggests the use of Conley (1999) standard errors.[33]

5 Results

5.1 Reconstructing the scheduling process

Non-parametric, descriptive evidence of the coordination motive is presented in the binscatter in Figure 2. Weekday clashes are less likely in the near vicinity than at longer distances.

Table 1 presents the results from the discrete choice scheduling problem in Section 4.2, where markets decide on their market day based on the distance to other already existing

[33] Conley (1999) standard error adjustments follow the code implementation by Hsiang (2010).

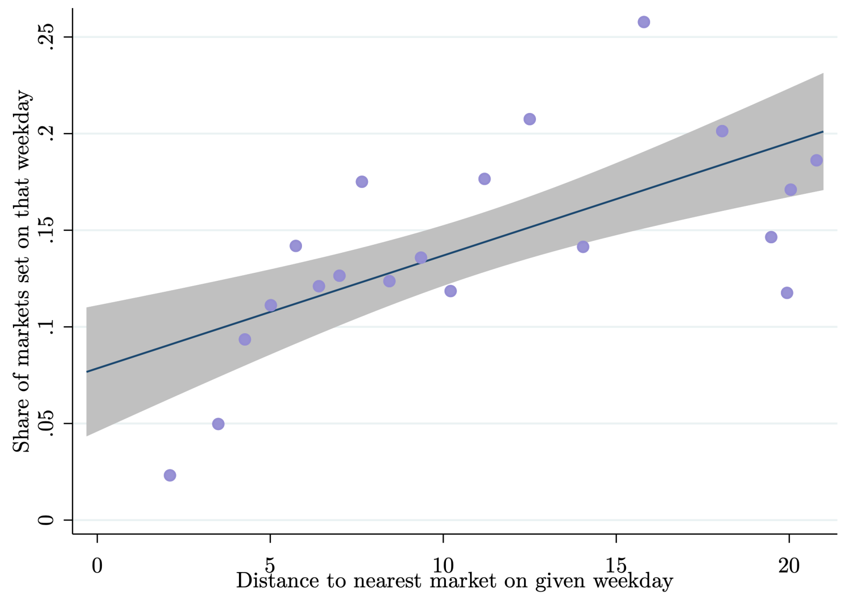


Figure 2: Binscatter of the share of markets with clashing weekday schedules at different distances.

markets at the time they start their own markets, as well as day fixed effects. The distance parameter enter significantly positively, indicating that markets are more likely to be scheduled on a particular day of the week if the nearest existing market with the same market day is further away. At the sample-wide average distance between two markets happening on the same day of 9km (Appendix Table A5), the probability of two markets that are exactly this far apart to share the same market day is 13.1% (Figure 2).[34] The estimated coefficient on distance $\hat{\delta}_{\Delta}$ implies that halving the distance between the two markets would reduce that probability to 9.1%. The coefficients on the weekday fixed effects mirror Appendix Figure A3.

It is worth recalling Assumption **A3** here, that after accounting for distance and weekday fixed effects, the remaining variation in weekday choices can indeed be attributed to idiosyncratic preference shocks ξ_{ld} that may have materialized differently. Concretely, they cannot be related to the error term ε_l for outcomes in equation (7). This assumption would be violated if, for example, village elders chose to purposefully compete with certain markets rather than others based on knowledge unobserved by me, the econometrician. If they had

[34]For reference, under completely random weekday assignment we would expect the probability of a clash at any distance to be about $1/7 = 14.3\%$.

accurate foresight over which market would pose a lesser degree of competition to their own market, that would reintroduce omitted variable bias I am trying to purge out. While readers may be reassured to know that none of the key informants mentioned such scheduling motives, it may still be more convincing to test this assumption empirically. Fortunately, I have some means of testing this assumption. In Appendix Table A6, I add three candidate variables of the kind of information village elders may have had in addition to the location and schedules of existing markets that should be correlated with any measures of competitiveness that we may be worried about: The age of the neighboring market at the time, the population of its hosting town when that market started, as well as whether it was at the time connected to a paved road. Reassuringly, including these controls barely changes the coefficients of interest and only one of the controls appears significant and not in the direction one might expect: Village elders appear slightly more likely to compete with the market of a larger town holding location fixed. I interpret this as evidence that the two error terms are indeed orthogonal.

5.2 Expected market area

After parameterizing the scheduling process in Table 1, I simulate 1,000 new runs of it under different Type-1 EV draws. Borusyak and Hull (2023) suggest that under a correctly specified shock assignment process, the slope of a regression of A_l on $\mathbb{E}[\mathrm{A}_l]$ should be close to 1 and the intercept close to 0. I provide evidence of this in the top panel of Appendix Figure A5. The bottom panel shows the distribution of residualized market area, which will be useful in interpreting the results. Holding geography and market entry order (and therefore the expected market area) fixed and moving a village from the 10^{th} to the 50^{th} percentile of the realized market area distribution (which can be thought of as preventing schedule clashes with immediate neighbors) coincides with adding about 86 km^2 of market area and from the 50^{th} to the 90^{th} percentile (akin to assigning a market a spatio-temporal monopoly) about an additional 157 km^2. For comparison, the area of Manhattan is about

Table 1: Weekday choice

	(1)
Distance	.0911***
	(.0158)
Day FEs	
Tuesday	-.162
	(.205)
Wednesday	-.00674
	(.203)
Thursday	-.392*
	(.219)
Friday	-.563**
	(.23)
Saturday	-.961***
	(.253)
Sunday	-.321
	(.213)
Observations	1960

Note: This table shows the estimates of the hedonic utility function (8). Standard errors in parentheses. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

59.1 km^2, Paris inside the Boulevard Périphérique measures about 87 km^2, while the Bronx measures about 110 km^2, and the District of Columbia is about 158 km^2. Reassuringly, the simulations match other moments of the observed data well. For example, similar to the finding in Appendix Table A5, I find an average distance to the nearest market with a clashing market day of 8.88 km.

5.3 Balance

I now proceed to testing for baseline balance on the pre-treatment control variables I described in section 3.2. Column (1) of Table 2 shows strong imbalances in the pre-determined control variables with respect to the market area in a simple regression. Markets further North and East and further away from Lake Victoria and in more rugged terrain appear to have larger market areas. In a regression in which all of the controls are used to predict market area, I strongly reject the hypothesis that the controls are irrelevant (see the F-test at the bottom). Imbalances with respect to treatment would indicate effects estimated by simply comparing markets with large and small catchment areas are not causal and potentially systematically biased, confounding the effect of market area with that of the imbalanced pre-determined variables and geographic fundamentals. The Borusyak and Hull (2023) methodology provides a solution to this problem: By purging out the non-random component of market scheduling, it allows me to obtain market catchment area deviations from the expectation that are driven purely by random idiosyncratic weekday preference. Under my model, controlling for the expectation of market schedule market area purges out potential omitted variable bias, not only of the variables presented in the sample balance analysis but any other unobservable variable as well. I provide evidence that the sample imbalance is indeed ameliorated after accounting for expected market area in column (2) of Table A4. Reassuringly, the loss in significance appears to be driven by a reduction in coefficient size rather than an inflation of standard errors.

Table 2: Sample balance

	(1) Mean (SD)	(2) Market area	(3) Recentered Market area
Latitude	0.3	.00047***	.00004
	(0.2)	(.0000899)	(.000145)
Longitude	34.4	.000807***	-3.00e-06
	(0.3)	(.0001)	(.000138)
Distance: Lake Victoria	36,580.2	64.6***	1.12
	(24,365.8)	(8.84)	(14.1)
Distance: Kisumu	65,318.6	4.58	-4.48
	(19,011.8)	(7.66)	(11.9)
Population 1920 (HHs)	3.8	-.00354	.01**
	(8.3)	(.00369)	(.00496)
Agricultural suitability	-14.1	.0531	.14
	(848.5)	(.0539)	(.428)
Ruggedness	30,965.3	32.2**	-14.1
	(38,004.9)	(15.9)	(21.5)
Observations		280	280
F		9.26	1.04
$p(F = 0)$		2.8e-10***	.41

Note: This table describes the sample balance of the measure of market area with respect to several pre-determined variables. The main panel reports the sample means and coefficients of the area measure in a regression predicting each control variable. At the bottom of the table, an F statistic is reported for a regression in which all controls jointly predict the treatment. The first column shows sample means and standard deviations. In the second column, the coordination measure is used at face value while in the third column I partial out the expectation of the coordination measure. Standard errors are reported in parentheses. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

5.4 Market attendance

I estimate equation (9) for analyzing the effects of market catchment areas on seller and buyer attendance in Tables 3 and 4. An additional square kilometer of market catchment area (beyond what the reconstructed scheduling process would have us expect) induces the market size to increase by about 1.636 sellers and 2.596 buyers. Moving a village from the 10^{th} to the 50^{th} percentile of the market area distribution translates into an increased market attendance of 141 sellers and 223 buyers and moving it from the 50^{th} to the 90^{th} percentile leads to an additional 257 sellers and 408 buyers respectively, relative to an average main market day attendance of 649 sellers and 1171 buyers (after accounting for rounding precision loss). Decomposing attendance in local and outside attendance, both seem to mainly be driven by cross-attendees from outside, although both increase significantly in the case of buyers. All of these results are robust to the inclusion of baseline controls.

5.5 Economic development

In Table 5, I estimate equation (9) for the outcomes discussed in Section 3.2. Even after adjusting for its expectation, market area has an effect on present-day population of the villages that host markets. An additional square kilometer of market area (relative to its expectation) induces a population increase in the hosting village of .4137 households. Moving a village from the 10^{th} to the 50^{th} percentile of the market area distribution translates into an effect size of 36 additional households and moving it from the 50^{th} to the 90^{th} percentile leads to 65 additional households respectively, relative to an average village size of 213 households. Assuming an average household size of 7 people, this corresponds to headcount increases of 249 and 455, respectively. Note that this is not a purely mechanical result, as the outcome variable in question is not the population of the market area, but that of the village hosting the market. Even after adjusting for its expectation, market area has an effect on present-day nighttime luminosity, as Table 5 shows. This suggests that market day scheduling has impacts not just on market attendance but more broadly on economic activity in the area. All of these results are robust to including baseline controls into the regression.

5.6 A comment on debiasing

Before discussing the substantive results of this paper, it is worth briefly reviewing how the balance Table 2 differs from the results in Tables 3 – 5. While the shifts in coefficients in geographic fundamentals from including the expected market catchment area are large and meaningfully change their interpretation in the balance Table 2 (and somewhat in the broader development outcomes in Table 5), the attendance results in Tables 3 and 4 barely change. This suggests that they do not appear to have induced omitted variable bias in these outcomes. Of course, while it may be ex post reassuring that the same bias removal technique that cleans up the balance table has relatively little effects on the naive OLS regression results of market attendance on catchment area, this was not ex ante obvious.

6 Discussion

As Appendix Figure A4 shows, the number of markets with a main market day has been fairly stable for fifty years and their schedules and with those their catchment areas have been largely unchanged (Appendix Table A3), suggesting that the local markets are in equilibrium. The literature review suggests that this is a context in which markets are likely too thin. Consider drawing market catchment areas with the objective of maximizing attendance subject to participants' travel costs. A catchment area would be defined in such a way that those living on its boundary are indifferent about attending the market when weighing their transport cost. Considering the long-term horizon of this quasi-experiment, finding significantly positive average treatment effects may be surprising as it suggests agglomeration gains in equilibrium: The areas within which participants can gainfully attend various markets on the same day appear to be overlapping and the coefficients I estimate suggest that the population over which adjacent markets are competing is sizable relative to these markets' attendance. One might imagine resolving this excess competition by moving markets slightly further apart, but since villages cannot readily move to achieve gradual gains, agglomeration would have to happen on the extensive margin. However, I find little evidence of markets disappearing or merging, suggesting that the situation is not sufficiently suboptimal for such a large-scale restructuring either.

In an equilibrium that maximizes agglomeration, market catchment areas should be so large that further expansion of market catchment areas should not induce higher attendance, but this does not appear to be the equilibrium that we are in. While the previous literature on market integration through roads or information technology does not provide reference points for the effect on market attendance, the share of the population "stuck between two markets" that I find in this study seem large. I rationalize this finding through a combination of misaligned incentives and uninternalized externalities, leading to persistence in an equilibrium made suboptimal by technological advance: As documented in Appendix Table A3, schedule persistence is very high. The absence of a social planner implies that two kinds of externalities were potentially not internalized: First, markets choose their optimal day sequentially, given the market schedule that prevailed at the time of market inception. This may undo the optimality of the schedule for the markets that came before them. Why do markets not change their schedules or cease to exist? Market attendants likely do not internalize the market thickness they contribute to the market they shop at (and withhold from the one they do not attend). They might privately optimally decide to attend a nearer-by market with lower transport cost (thus keeping that market alive), disregarding that if they went to a bigger, further-away market, that would motivate others to follow suit and in turn make that market worth the longer travel. A political economy version of this is the question: If two markets are in suboptimal spatial competition with each other, which of them should give up their market in favor of the other? Markets rely on the amenities of its local town (safety, transport, sanitary facilities, short commute for local participants), so merging two markets into a location halfway between two towns may not be practical, so it would have to be in one or the other. Even if the joint market were appealing enough to attend even for the villagers who no longer have their own village market, the immediate increased travel cost may outweigh or at least be more salient than the agglomeration gains. The gains from market consolidation accrue to the attendants of the surviving market gaining market thickness while the costs are borne by the attendants of the dying market that are now traveling further, leading to misaligned incentives. It may be locally optimal to keep a small market operational. Even schedule adjustments are hampered by this logic: When asked what they would do if holding a market on their current market day were no longer possible, a non-negligible share of market stakeholders indicate that shifting their market day would have such adverse effects on market attendance that they might as well close the market altogether (Appendix Table A3).

Further, consider that optimal catchment area size increases as transport costs fall and the introduction of markets in this area predates the mass adoption of even the bicycle, while motorized public transport on paved roads is a relatively recent phenomenon. In 1975, when the market schedule had mostly settled to what it is today, only about 12% of the markets in my census were connected to a paved road while the share has increased to 48% today. Splitting the sample by recent (endogenous!) road infrastructure improvements, it is precisely the markets with road access that would benefit from schedule harmonization that increases their catchment areas (Table 6). Markets in this area proliferated when participants were walking and transport costs were prohibitive and this finding suggests that trade activity today may be scattered across suboptimally many markets, leaving agglomeration gains on the table.

In this paper, I have described and evaluated a natural experiment in how market days in Western Kenya were scheduled. I found that while a simple regression analysis of outcomes of interest on present-day market catchment area may be biased, this bias can be purged out by leveraging detailed contextual knowledge of the scheduling process. Doing so allows me to conclude that market catchment areas indeed have a sizable and lasting causal impact on market attendance, particularly cross-attendance from other towns and that it can also partly explain the population size and nighttime luminosity of the towns that host the market. Market areas in Western Kenya, while perhaps optimal in an age when its participants were exclusively walking, appear to be too fragmented in light of advances in transport technology. Yet, strong persistence in market locations and their schedules as well as misaligned incentives on which town should give up its market, appear to be preventing a consolidation. I hasten to add, however, that policy intervention in these immensely fragile food and livelihood systems requires caution and local stakeholder involvement. While my results suggest agglomeration gains from market consolidation *on average*, they do not imply that there wouldn't be both winners and losers. This is especially true if we believe that transport cost is heterogeneous and poorer market participants would disproportionately be excluded from market participation if their local market closed in favor of a larger, more distant commercial center.

Table 3: Market attendance: Sellers

	Seller attendance			Seller attendance (local)			Seller attendance (outside)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Market area	1.451***	1.636***	1.582***	0.0733*	0.0907	0.0752	1.376***	1.535***	1.495***
	(0.351)	(0.487)	(0.501)	(0.0406)	(0.0604)	(0.0571)	(0.329)	(0.454)	(0.471)
$\mathbb{E}$[Market area | Distance]		-0.317	-1.077		-0.0298	0.0617		-0.272	-1.076*
		(0.474)	(0.662)		(0.0558)	(0.0584)		(0.450)	(0.629)
Baseline controls	×	×	✓	×	×	✓	×	×	✓
Observations	273	273	273	271	271	271	271	271	271

Note: This table describes the effects of market area induced by scheduling on the outcome variable indicated at the top. Columns (1), (4), and (7) describe the results of the naive OLS regression of the outcome on market area, while columns (2), (5), and (8) remove omitted variable bias by controling for the expected market area. Columns (3), (6), and (9) add baseline controls to the regression. Conley (1999) standard errors with a distance cutoff of 5km are reported in parentheses. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Market attendance: Buyers

	Buyer attendance			Buyer attendance (local)			Buyer attendance (outside)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Market area	2.548***	2.596***	2.436***	0.208***	0.265***	0.248***	2.334***	2.312***	2.169***
	(0.561)	(0.750)	(0.765)	(0.0724)	(0.0954)	(0.0879)	(0.520)	(0.696)	(0.709)
$\mathbb{E}$[Market area | Distance]		-0.0825	-1.173		-0.0969	0.0858		0.0375	-1.170
		(0.777)	(1.068)		(0.0821)	(0.0912)		(0.739)	(1.003)
Baseline controls	×	×	✓	×	×	✓	×	×	✓
Observations	273	273	273	271	271	271	271	271	271

Note: This table describes the effects of market area induced by scheduling on the outcome variable indicated at the top. Columns (1), (4), and (7) describe the results of the naive OLS regression of the outcome on market area, while columns (2), (5), and (8) remove omitted variable bias by controling for the expected market area. Columns (3), (6), and (9) add baseline controls to the regression. Conley (1999) standard errors with a distance cutoff of 5km are reported in parentheses. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Economic development

	Population (HHs)			Nightlights			Log nightlights		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Market area	0.155	0.414***	0.370***	0.106**	0.215***	0.217***	0.0114***	0.0146***	0.0144***
	(0.105)	(0.152)	(0.142)	(0.0423)	(0.0781)	(0.0811)	(0.00366)	(0.00517)	(0.00537)
$\mathbb{E}$[Market area | Distance]		-0.441*	-0.226		-0.185	-0.239*		-0.00556	-0.00704
		(0.234)	(0.213)		(0.118)	(0.139)		(0.00731)	(0.00873)
Baseline controls	×	×	✓	×	×	✓	×	×	✓
Observations	278	278	278	280	280	280	280	280	280

Note: This table describes the effects of market area induced by scheduling on the outcome variable indicated at the top. Columns (1), (4), and (7) describe the results of the naive OLS regression of the outcome on market area, while columns (2), (5), and (8) remove omitted variable bias by controling for the expected market area. Columns (3), (6), and (9) add baseline controls to the regression. Conley (1999) standard errors with a distance cutoff of 5km are reported in parentheses. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Heterogeneous effects by road access

	Sellers in markets w/o road			Sellers in markets w/ road			Buyers in markets w/o road			Buyers in markets w/ road		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Market area	0.510***	0.0674	0.0428	1.459***	1.900**	1.949***	1.000***	0.272	0.215	2.576***	2.865**	2.915**
	(0.168)	(0.277)	(0.261)	(0.552)	(0.727)	(0.717)	(0.285)	(0.485)	(0.459)	(0.887)	(1.129)	(1.133)
$\mathbb{E}$[Market area | Distance]		0.629**	0.368		-0.863	-3.906**		1.034*	0.857		-0.567	-5.302*
		(0.293)	(0.344)		(0.864)	(1.548)		(0.563)	(0.611)		(1.380)	(2.313)
Baseline controls	×	×	✓	×	×	✓	×	×	✓	×	×	✓
Observations	140	140	140	133	133	133	140	140	140	133	133	133

Note: This table describes the effects of market area induced by scheduling on the outcome variable indicated at the top. Columns (1), (4), and (7) describe the results of the naive OLS regression of the outcome on market area, while columns (2), (5), and (8) remove omitted variable bias by controling for the expected market area. Columns (3), (6), and (9) add baseline controls to the regression. Conley (1999) standard errors with a distance cutoff of 5km are reported in parentheses. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Appendix of chapter 3

A Additional tables and figures

Table A1: Outcome Variables

Name	**Definition**	**Source**	**Spatial resolution**	**Temporal resolution**	**Citation**
Attendance	Attendance estimates of sellers and buyers on the main market day	Phone survey	Market	2023–2024	This paper
Population	Number of households presently living in the village	Phone survey	Market	2023–2024	This paper
Nighttime lights	Lunar-adjusted top of atmosphere radiance	NASA BlackMarble	∼500m	January 2023 (composite)	Román et al. (2018)

Note: This table summarizes the outcome data sources used in this paper, alongside with definitions, sources, applicable resolution and useful citations.

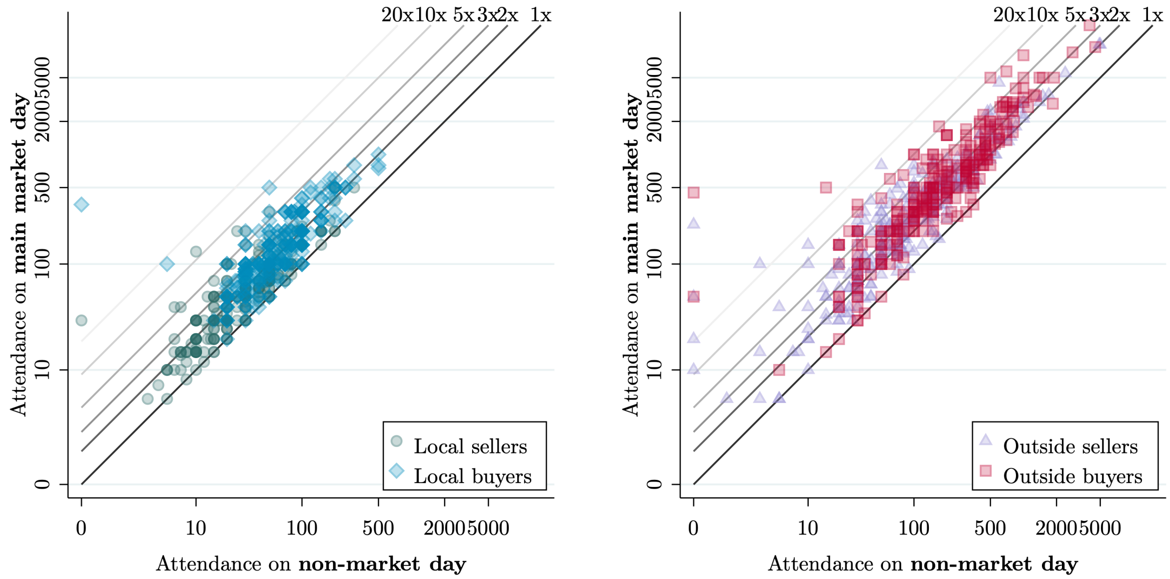


Figure A1: Market attendance

Market attendance, shown here on a $\log(x+1)$ scale by market, broken down by attendant category, is substantially higher on market day than on other days of the week, especially among participants from outside the town or village (right panel). But even local attendance is higher on market day.

Table A2: Source of information

	Count
I was alive at the time the market started	95
From a village elder who **was** ”	29
From a village elder who **was not** ”	3
From another village member who **was** ”	70
From another village member who **was not** ”	16
From a family member who **was** ”	64
From a family member who **was not** ”	1
From a written record	2
Total	280

Note: This table describes how close respondents are to the source of information about the process of market scheduling. Most respondents either experienced the start of their market first hand or learned about it from someone who did. Only a small share of responses relies on data that is more than one network link remote. Written records are rare.

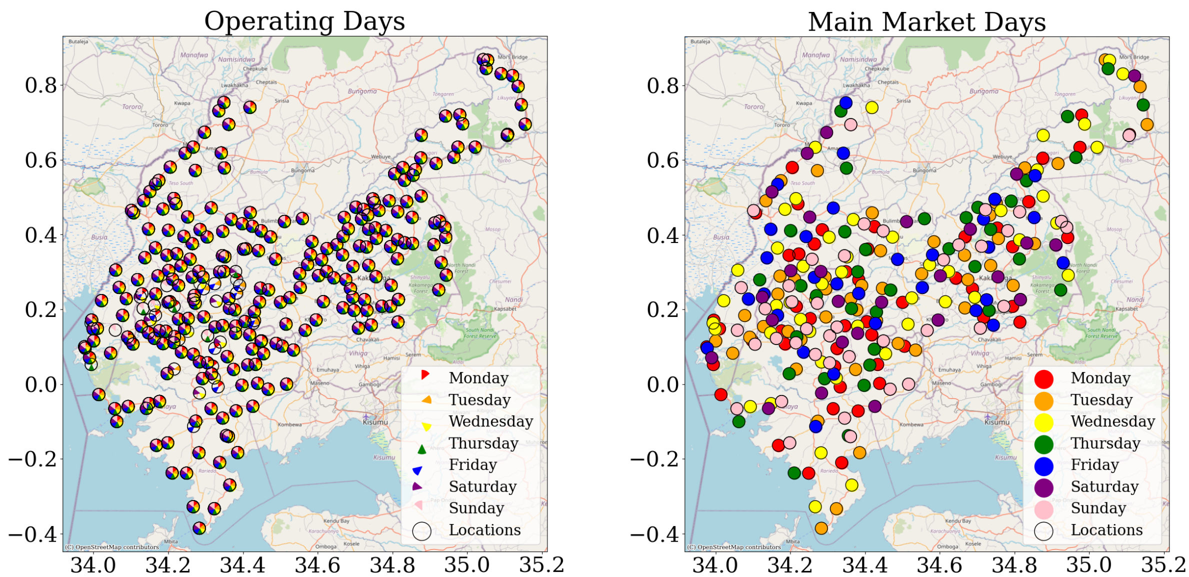


Figure A2: Map of markets

Map of markets in Siaya, Busia, and Kakamega County. The left panel shows all markets along with the days on which they operate (often daily), while the right panel shows the main market days with substantially higher attendance.

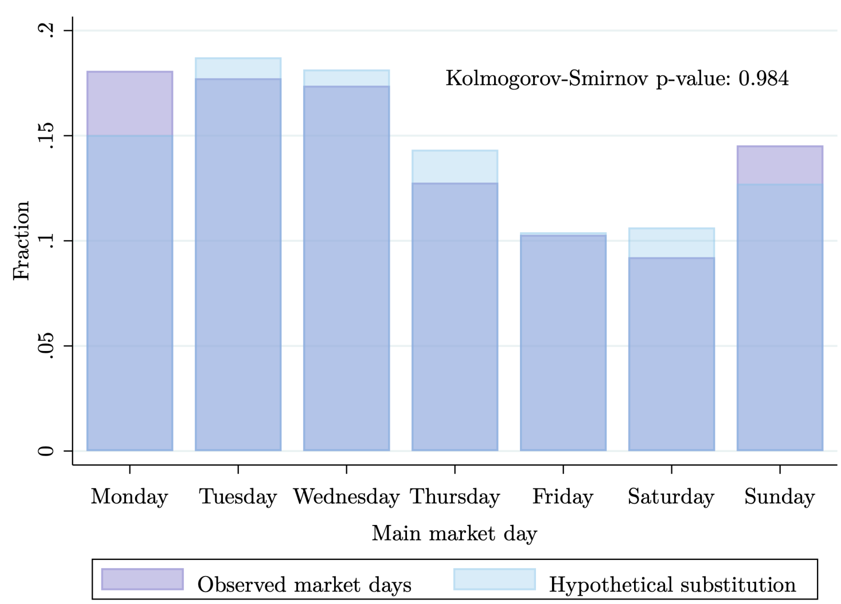


Figure A3: Market days hypothetically chosen if current day were no longer feasible.

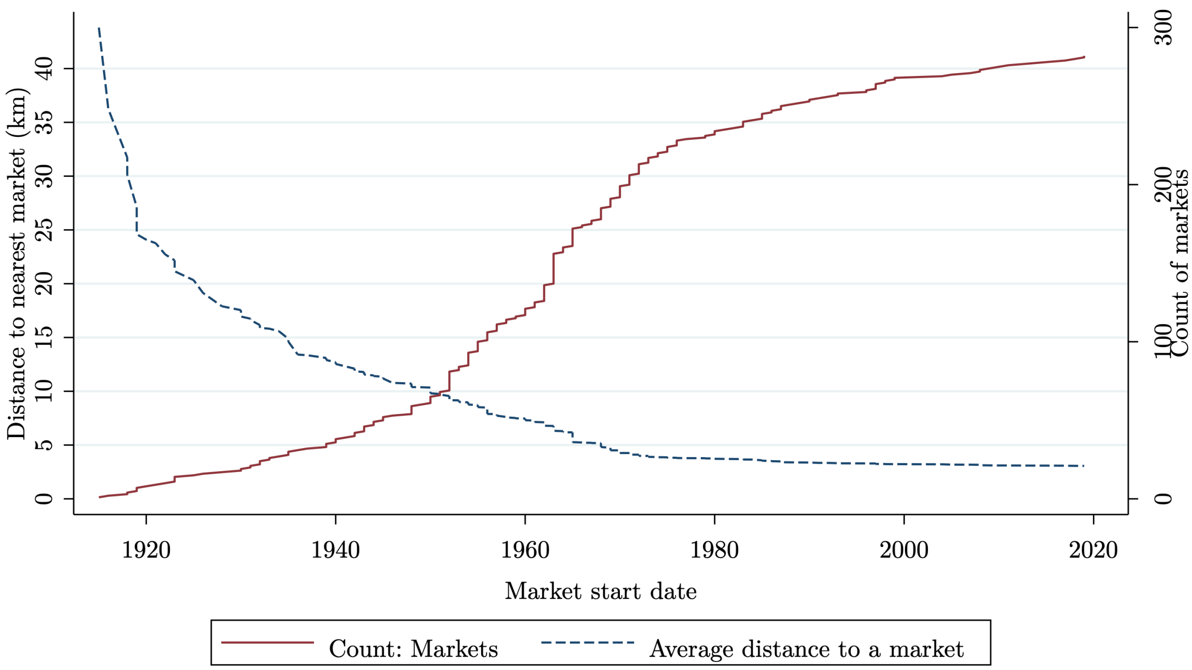
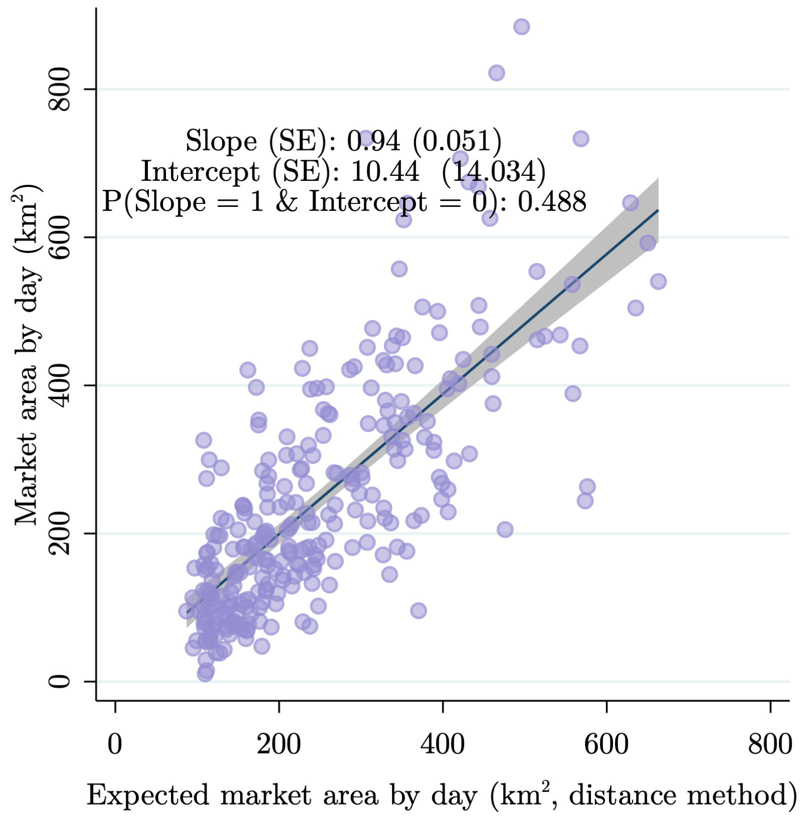


Figure A4: Market creation over time

Markets were first started in the colonial period with an acceleration around Kenyan Independence in 1963 (left axis). Most markets with a main market day had been established by 1975 (left axis). The average distance to a market with a main market day similarly fell rapidly until around 1975 (right axis) to a walkable distance.



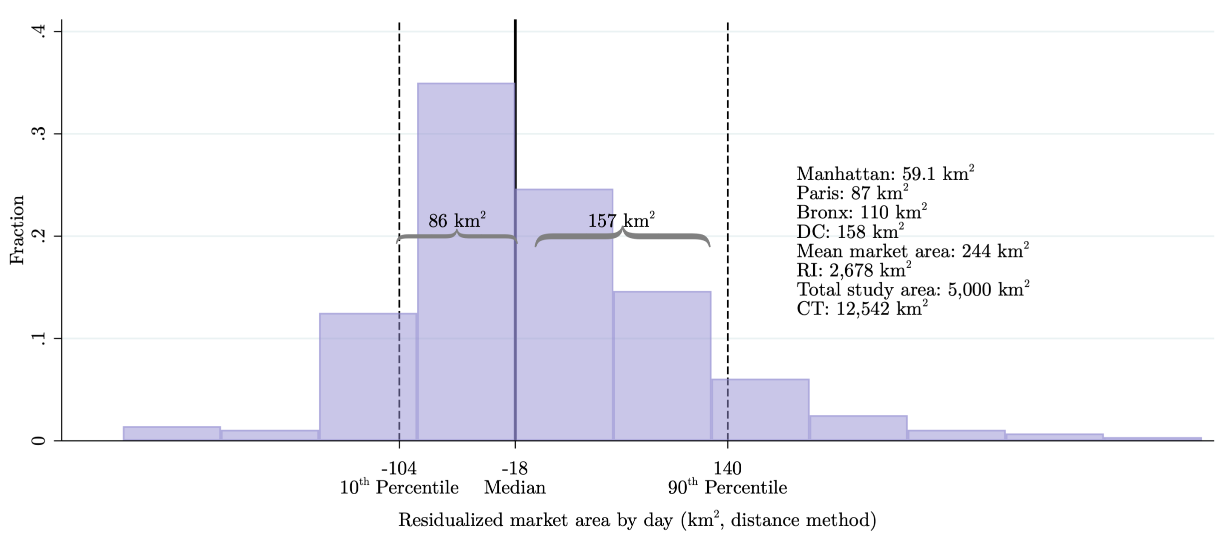


Figure A5: Expected market area predicts realized market area.

Table A3: Market descriptives and persistence

	Percent
Markets descriptives	
Number of sellers	648.7
Number of buyers	1171.3
Population (HHs) around 1920	4.7
Population (HHs) when market was first started	53.8
Population (HHs) today	263.1
Market operates daily	87.1
Road access in 1975	12.5
Road access today	48.2
Schedule persistence	
Market schedule never changed	89.2
Shifted to different market day	9.3
Added or dropped a market day	1.5
Rather close than change market day	60.4
Operational persistence	
Market never interrupted	77.5
Interrupted: COVID	18.9
Interrupted: Other reason	3.6
Observations	280

Note: This table describes the sample of markets, their schedule and operational persistence.

Table A4: Decision makers

	Percent
Village elder(s)	20.5
Local council	43.3
Traders from the village	66.4
Traders from outside the village	37.7
Emerged organically	7.5
Sublocation chief	9.7
Location chief	17.2
Ward administration	0.7
County government	23.3
Central government	0.0
Colonial government	2.6
Don't know	1.9
Other	0.4
Observations	280

Note: This table describes who was (and who wasn't) involved in the schedule setting of the main market day. It highlights that market scheduling was a decentralized coordination process that was not implemented by a social planner.

Table A5: Coordination motive

	Percent
Scheduling motives	
To not clash with the schedule of another market	69.8
Outside traders visited regularly on this day	13.8
Idiosyncratic day preference	33.9
To align with the church schedule	8.2
Don't know	9.3
Other	0.7
Obervations	280
Hypothetical coordination of new markets	
Mention schedule coordination	86.3
Mention idiosyncratic day preference	12.5
Observations	328

Note: This table describes what factors determined schedule setting of the main market day. It highlights that market scheduling was driven by a coordination motive as well as local idiosyncratic preferences over week days. The bottom panel shows that the majority of towns without a market but that entertain the possibility of starting one also mention the need for schedule coordination. These data are manually coded from open-ended answers.

Table A6: Scheduling: Robustness given other local knowledge

	(1)	(2)
Distance	.0911***	.0896***
	(.0158)	(.0165)
Age of neighboring market at the time		-.00689
		(.00479)
Population of ∼		.002*
		(.00106)
Road access to ∼		-.133
		(.154)
Day FEs		
Tuesday	-.162	-.191
	(.205)	(.207)
Wednesday	-.00674	.0317
	(.203)	(.204)
Thursday	-.392*	-.324
	(.219)	(.22)
Friday	-.563**	-.532**
	(.23)	(.235)
Saturday	-.961***	-.957***
	(.253)	(.254)
Sunday	-.321	-.288
	(.213)	(.216)
Observations	1960	1960

Note: This table shows the estimates of the hedonic utility function (8), including other potential explanatory variables for market day choice. Standard errors in parentheses. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

B Bias simulation

In this section, I simulate the bias that this paper is concerned with, to build intuition for how the Borusyak and Hull (2023) methodology corrects it in the context of market catchment areas. Directly comparing markets with large catchments to those with small catchments (equation (7)) would potentially be confounded. Omitted variables might either directly make a place more suitable than its surroundings for market exchange *and* improve the outcome of interest (perhaps fertile, non-rugged terrain) or causality might be reversed in that places with stronger initial conditions (e.g. larger towns in the early twentieth century) started markets first and their surrounding areas were thus discouraged from starting their own markets, leaving the early markets with a larger catchment area while their present-day outcomes are simply a product of persistence of the initial conditions, not of the market catchment area. This would bias the simple OLS coefficient in equation (7) upward.

To simulate such a situation, I start with a regular grid of potential market locations l, randomly select a subset as being particularly favorable sites, and prevent any competing markets from entering nearby. The simulation is set up so that the overall number of markets and the average distances between sites approximately match my data. The left panel of Figure B1 displays a map with one such bias scenario in which a small share of sites are endowed with favorable initial conditions that both afford them low spatial competition as no markets are started nearby, but I also add some arbitrary bonus B to their outcome variable y_l that is large enough to make me overestimate the effect of catchment area when estimating equation (7). For simplicity, the outcome variable is a sum of this bonus (for favorable market sites) and the causal impact of market area (in km^2), which I will assume enters one-for-one, and some noise:

$$y_l = 1 \times \text{Catchment area}_l + B \times \mathbb{1}[\text{Favorable location}_l] + \varepsilon_l$$

where ε_l is normally distributed. Following the scheduling process I describe in the paper, I then allow all markets in some random sequential order to enter. If they are a favorable location, they deter all potential neighbors from entering, creating a catchment area advantage. For simplicity, I assume that δ_{Δ} in equation (8), which describes how strongly market places seek to avoid schedule clashes with another market depending on distance to that market relative to the variance of idiosyncratic weekday preferences ξ_l, be equal to 1. These parameters are shown in column (1) of Table B1. The resulting market schedules are shown in the right panel of Figure B1.

Estimating equation (7) produces the expected upward bias, and I show the average coefficient from 1,000 bias scenarios in the top panel of column (2) in Table B1. The bottom panel rejects the hypothesis that the estimate for β is equal to the true value. To recover the true effect, I start by backing out the parameters that guide scheduling behavior (equation (8)) and the middle panel shows the average estimate of δ_{Δ} across all 1,000 scenarios. Following Borusyak and Hull (2023), I use the estimated δ_{Δ} for each bias scenario to compute the expected market catchment area for all markets across 1,000 scheduling simulations and include it in the regression. This purges out the bias from deterred entry and recovers the true β parameter in the top panel of column (3).

Table B1: Bias simulation

	(1) True value	(2) Naive estimate	(3) BH estimate
Outcome regression equations (7) & (9)			
Market Area (β_A)	1.000	1.763	1.002
		(0.214)	(0.446)
		[0.000]	[0.103]
Expected Market area			0.992
			(0.510)
			[0.155]
Scheduling equation (8)			
Distance (δ_Δ)	1.000		1.019
			(0.099)
			[0.000]
$P(\delta = 1)$			0.501
$P(\beta = 1)$		0.018	0.503
Observations		273	273

Note: This table shows results from simulating and then removing bias using the Borusyak and Hull (2023) method. A regular grid of 1,024 potential market locations is generated with distance metrics matching those observed in the real data. A share of 10% of them is randomly selected to be intrinsically more suitable for market activities and economic development more broadly. This deters markets within a 10km distance from entering which mechanically endows suitable markets with a larger market area. Simultaneously these markets find themselves with better present-day outcomes beyond the imposed true effect β of market area in column (1). Column (2) shows that this would bias the coefficient in equation (7) upward. Market scheduling was governed by the parameter δ in column (1). In the Scheduling panel, using 1,000 simulations, I show that I can recover this parameter. Employing this information to compute expected market catchment areas for the markets in each bias scenario over 1,000 simulations, I show that controlling for expected market catchment area can purge out the introduced bias from deterred entry and recover the true β parameter. All coefficients, standard errors (in parentheses), and p-values (in brackets) in this table are averages over all simulations. In the bottom panel I test, whether they deviate from the assumed true values and list the number of observations used in the simulations.

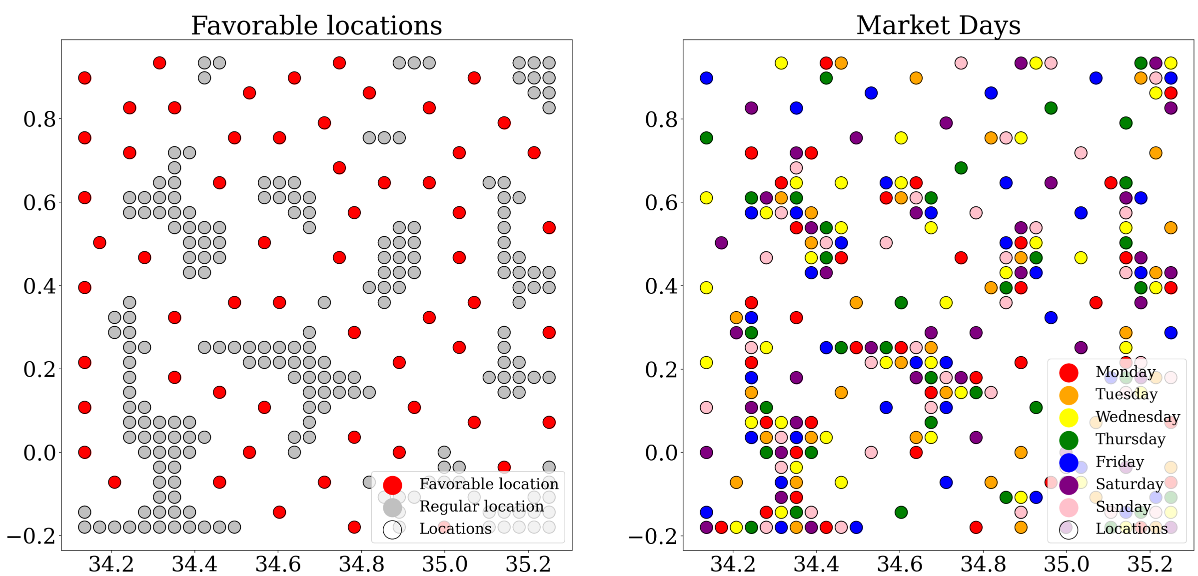


Figure B1: Simulated map

Map of simulated (favorable) markets locations (left) and simulated schedules (right)

Bibliography

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey M Wooldridge, "Sampling-based versus design-based uncertainty in regression analysis," *Econometrica*, 2020, *88* (1), 265–296.

_ , _ , _ , and Jeffrey Wooldridge, "When should you adjust standard errors for clustering?," Technical Report, National Bureau of Economic Research 2017.

Agarwal, Nikhil, Itai Ashlagi, Eduardo Azevedo, Clayton R Featherstone, and Ömer Karaduman, "Market failure in kidney exchange," *American Economic Review*, 2019, *109* (11), 4026–4070.

Aggarwal, Shilpa, Jenny Aker, Dahyeon Jeong, Naresh Kumar, David Sungho Park, Jonathan Robinson, and Alan Spearot, "The dynamic effects of cash transfers: Evidence from Rural Liberia and Malawi," 2022.

Aizer, Anna, Shari Eli, Joseph Ferrie, and Adriana Lleras-Muney, "The long-run impact of cash transfers to poor families," *American Economic Review*, 2016, *106* (4), 935–71.

Akbarpour, Mohammad, Shengwu Li, and Shayan Oveis Gharan, "Thickness and information in dynamic matching markets," *Journal of Political Economy*, 2020, *128* (3), 783–815.

Aker, Jenny C, "Information from markets near and far: Mobile phones and agricultural markets in Niger," *American Economic Journal: Applied Economics*, 2010, *2* (3), 46–59.

Altındağ, Onur and Stephen D O'Connell, "The short-lived effects of unconditional cash transfers to refugees," *Journal of Development Economics*, 2023, *160*, 102942.

Andrews, Isaiah, Toru Kitagawa, and Adam McCloskey, "Inference on winners," *The Quarterly Journal of Economics*, 2024, *139* (1), 305–358.

Araujo, Maria Caridad, Mariano Bosch, and Norbert Schady, "Can Cash Transfers Help Households Escape an Inter-Generational Poverty Trap?," *The Economics of Poverty Traps*, 2019, pp. 357–382.

Athey, Susan, Raj Chetty, Guido W Imbens, and Hyunseung Kang, "The surrogate index: Combining short-term proxies to estimate long-term treatment effects more rapidly and precisely," Technical Report, National Bureau of Economic Research 2019.

Bai, Yuehao, "Optimality of matched-pair designs in randomized controlled trials," *American Economic Review*, 2022, *112* (12), 3911–3940.

Baird, Sarah, Craig McIntosh, and Berk Özler, "Cash or condition? Evidence from a cash transfer experiment," *The Quarterly Journal of Economics*, 2011, *126* (4), 1709–1753.

_ **, J Aislinn Bohren, Craig McIntosh, and Berk Özler**, "Optimal design of experiments in the presence of interference," *Review of Economics and Statistics*, 2018, *100* (5), 844–860.

Balan, Pablo, Augustin Bergeron, Gabriel Tourek, and Jonathan L Weigel, "Local Elites as State Capacity: How City Chiefs Use Local Information to Increase Tax Compliance in the Democratic Republic of the Congo," *American Economic Review*, 2022, *112* (3), 762–97.

Balboni, Clare, Oriana Bandiera, Robin Burgess, Maitreesh Ghatak, and Anton Heil, "Why do people stay poor?," *The Quarterly Journal of Economics*, 2022, *137* (2), 785–844.

Baldwin, Kate, *The paradox of traditional chiefs in democratic Africa*, Cambridge University Press, 2016.

Bandiera, Oriana, Robin Burgess, Narayan Das, Selim Gulesci, Imran Rasul, and Munshi Sulaiman, "Labor markets and poverty in village economies," *The Quarterly Journal of Economics*, 2017, *132* (2), 811–870.

Banerjee, Abhijit, Dean Karlan, Robert Osei, Hannah Trachtman, and Christopher Udry, "Unpacking a multi-faceted program to build sustainable income for the very poor," *Journal of Development Economics*, 2022, *155*, 102781.

_ **, Emily Breza, Esther Duflo, and Cynthia Kinnan**, "Can microfinance unlock a poverty trap for some entrepreneurs?," Technical Report, National Bureau of Economic Research 2025.

— , Esther Duflo, and Garima Sharma, "Long-term effects of the targeting the ultra poor program," *American Economic Review: Insights*, 2021, *3* (4), 471–86.

— , — , Nathanael Goldberg, Dean Karlan, Robert Osei, William Parienté, Jeremy Shapiro, Bram Thuysbaert, and Christopher Udry, "A multifaceted program causes lasting progress for the very poor: Evidence from six countries," *Science*, 2015, *348* (6236), 1260799.

Barrett, Christopher B., "Spatial Market Integration," in Steven N. Durlauf Lawrence E. Blume, ed., *The New Palgrave Dictionary of Economics*, 2nd ed., London: Palgrave Macmillan, 2005, pp. 115–168.

Basurto, Maria Pia, Pascaline Dupas, and Jonathan Robinson, "Decentralization and efficiency of subsidy targeting: Evidence from chiefs in rural Malawi," *Journal of Public Economics*, 2020, *185*, 104047.

Bauchet, Jonathan, Jonathan Morduch, and Shamika Ravi, "Failure vs. displacement: Why an innovative anti-poverty program showed no net impact in South India," *Journal of Development Economics*, 2015, *116*, 1–16.

Beam, Emily A, Lasse Brune, Narayan Das, Stefan Dercon, Nathanael Goldberg, Rozina Haque, Dean Karlan, Maliha Noshin Khan, Doug Parkerson, Ashley Pople et al., "Group versus Individual Coaching for Rural Social Protection Programs: Evidence from Uganda, Philippines, and Bangladesh," 2025.

Bedoya, Guadalupe, Aidan Coville, Johannes Haushofer, Mohammad Isaqzadeh, and Jeremy P Shapiro, "No household left behind: Afghanistan targeting the ultra poor impact evaluation," Technical Report, National Bureau of Economic Research 2019.

Benhassine, Najy, Florencia Devoto, Esther Duflo, Pascaline Dupas, and Victor Pouliquen, "Turning a shove into a nudge? A "labeled cash transfer" for education," *American Economic Journal: Economic Policy*, 2015, *7* (3), 86–125.

Berge, Lars Ivar Oppedal, Kjetil Bjorvatn, and Bertil Tungodden, "Human and financial capital for microenterprise development: Evidence from a field and lab experiment," *Management Science*, 2015, *61* (4), 707–722.

Bergquist, Lauren Falcao and Michael Dinerstein, "Competition and entry in agricultural markets: Experimental evidence from Kenya," *American Economic Review*, 2020, *110* (12), 3705–3747.

Blattman, Christopher and Laura Ralston, "Generating employment in poor and fragile states: Evidence from labor market and entrepreneurship programs," *Available at SSRN 2622220*, 2015.

_ , Eric P Green, Julian Jamison, M Christian Lehmann, and Jeannie Annan, "The returns to microenterprise support among the ultrapoor: A field experiment in postwar Uganda," *American Economic Journal: Applied Economics*, 2016, *8* (2), 35–64.

_ , Nathan Fiala, and Sebastian Martinez, "Generating skilled self-employment in developing countries: Experimental evidence from Uganda," *The Quarterly Journal of Economics*, 2014, *129* (2), 697–752.

Bleakley, Hoyt and Jeffrey Lin, "Portage and path dependence," *The Quarterly Journal of Economics*, 2012, *127* (2), 587–644.

Bollobás, Bela and Nicholas Stern, "The optimal structure of market areas," *Journal of Economic Theory*, 1972, *4* (2), 174–179.

Borusyak, Kirill and Peter Hull, "Nonrandom Exposure to Exogenous Shocks," *Econometrica*, 2023, *91* (6), 2155–2185.

Bossuroy, Thomas, Markus Goldstein, Bassirou Karimou, Dean Karlan, Harounan Kazianga, William Parienté, Patrick Premand, Catherine C Thomas, Christopher Udry, Julia Vaillant et al., "Tackling psychosocial and capital constraints to alleviate poverty," *Nature*, 2022, *605* (7909), 291–297.

Botea, Ioana, Markus Goldstein, Corinne Low, and Gareth Roberts, "Supporting women's livelihoods at scale: RCT evidence from a nationwide graduation program," *World Bank Working Paper*, 2021, *175851.*

Bromley, Rose J, Richard Symanski, and Charles M Good, "The rationale of periodic markets," *Annals of the Association of American Geographers*, 1975, *65* (4), 530–537.

Brooks, Wyatt, Kevin Donovan, and Terence R Johnson, "Mentors or teachers? Microenterprise training in Kenya," *American Economic Journal: Applied Economics*, 2018, *10* (4), 196–221.

_ , _ , _ , and Jackline Oluoch-Aridi, "Cash transfers as a response to COVID-19: Experimental evidence from Kenya," *Journal of Development Economics*, 2022, *158*, 102929.

Bruhn, Miriam and David McKenzie, "In pursuit of balance: Randomization in practice in development field experiments," *American Economic Journal: Applied Economics*, 2009, *1* (4), 200–232.

Bryan, Gharad, Shyamal Chowdhury, Ahmed Mushfiq Mobarak, Melanie Morten, and Joeri Smits, "Encouragement and distortionary effects of conditional cash transfers," 2021.

Burchi, Francesco and Christoph Strupat, "Unbundling the impacts of economic empowerment programmes: Evidence from Malawi," 2018.

Burke, Marshall, Lauren Falcao Bergquist, and Edward Miguel, "Sell low and buy high: Arbitrage and local price effects in Kenyan markets," *The Quarterly Journal of Economics*, 2019, *134* (2), 785–842.

Cahyadi, Nur, Rema Hanna, Benjamin A Olken, Rizal Adi Prima, Elan Satriawan, and Ekki Syamsulhakim, "Cumulative impacts of conditional cash transfer programs: Experimental evidence from Indonesia," *American Economic Journal: Economic Policy*, 2020, *12* (4), 88–110.

Callaway, Brantly and Pedro HC Sant'Anna, "Difference-in-differences with multiple time periods," *Journal of Econometrics*, 2021, *225* (2), 200–230.

Casaburi, Lorenzo, Rachel Glennerster, and Tavneet Suri, "Rural roads and intermediated trade: Regression discontinuity evidence from Sierra Leone," *Available at SSRN 2161643*, 2013.

Chaisemartin, Clément De and Jaime Ramirez-Cuellar, "At what level should one cluster standard errors in paired and small-strata experiments?," *American Economic Journal: Applied Economics*, 2024, *16* (1), 193–212.

Chowdhury, Reajul, Elliott Collins, Ethan Ligon, and Munshi Sulaiman, "Valuing Assets Provided to Low-Income Households in South Sudan," 2017.

Collins, Daryl, Jonathan Morduch, Stuart Rutherford, and Orlanda Ruthven, *Portfolios of the poor: How the world's poor live on $2 a day*, Princeton University Press, 2009.

Conley, Timothy G, "GMM estimation with cross-sectional dependence," *Journal of Econometrics*, 1999, *92* (1), 1–45.

Crosta, Tommaso, Dean Karlan, Finley Ong, Julius Rüschenpöhler, and Christopher R Udry, "Unconditional cash transfers: A Bayesian meta-analysis of randomized evaluations in low and middle income countries," Technical Report, National Bureau of Economic Research 2025.

Cunha, Jesse M, Giacomo De Giorgi, and Seema Jayachandran, "The price effects of cash versus in-kind transfers," *The Review of Economic Studies*, 2019, *86* (1), 240–281.

Descartes, René, *Principia Philosophiae* 1664.

Desmet, Klaus and Stephen L Parente, "Bigger is better: Market size, demand elasticity, and innovation," *International Economic Review*, 2010, *51* (2), 319–333.

Dietler, Michael and Ingrid Herbich, "Living on Luo time: Reckoning sequence, duration, history and biography in a rural African society," *World Archaeology*, 1993, *25* (2), 248–260.

Dirichlet, Gustav Lejeune, "Über die Reduction der positiven quadratischen Formen mit drei unbestimmten ganzen Zahlen.," *Journal für die reine und angewandte Mathematik (Crelles Journal)*, 1850, *1850* (40), 209–227.

Dixit, Avinash K and Joseph E Stiglitz, "Monopolistic competition and optimum product diversity," *The American economic review*, 1977, *67* (3), 297–308.

Donaldson, Dave, "The gains from market integration," *economics*, 2015, *7* (1), 619–647.

_ , "Railroads of the Raj: Estimating the impact of transportation infrastructure," *American Economic Review*, 2018, *108* (4-5), 899–934.

Drexler, Alejandro, Greg Fischer, and Antoinette Schoar, "Keeping it simple: Financial literacy and rules of thumb," *American Economic Journal: Applied Economics*, 2014, *6* (2), 1–31.

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

Faber, Benjamin, "Trade integration, market size, and industrialization: Evidence from China's National Trunk Highway System," *Review of Economic Studies*, 2014, *81* (3), 1046–1070.

Fafchamps, Marcel, "Cash crop production, food price volatility, and rural market integration in the third world," *American Journal of Agricultural Economics*, 1992, *74* (1), 90–99.

_ **and Ruth Vargas Hill**, "Selling at the farmgate or traveling to market," *American Journal of Agricultural Economics*, 2005, *87* (3), 717–734.

Fagerlund, Vernon G and Robert HT Smith, "A preliminary map of market periodicities in Ghana," *The Journal of Developing Areas*, 1970, pp. 333–348.

Fiala, Nathan, Julian Rose, Filder Aryemo, and Jörg Peters, "The (very) long-run impacts of cash grants during a crisis," 2022, (961).

Filmer, Deon, Jed Friedman, Eeshani Kandpal, and Junko Onishi, "Cash transfers, food prices, and nutrition impacts on ineligible children," *The Review of Economics and Statistics*, 2021, pp. 1–45.

Fisher, Ronald Aylmer, *The design of experiments*, Edinburgh: Oliver and Boyd, 1935.

Foster, Elizabeth, Dean Jolliffe, Gabriel Lara Ibarra, Christoph Lakner, and Samuel Tetteh-Baah, "Global Poverty Revisited Using 2021 PPPs and New Data on Consumption," Technical Report, The World Bank 2025.

Fröhlich, Willy, *Das afrikanische Marktwesen*, Springer, 1941.

Geertz, Clifford, *Deep play: Notes on the Balinese cockfight*, Springer, 2000.

Gertler, Paul J, Sebastian W Martinez, and Marta Rubio-Codina, "Investing cash transfers to raise long-term living standards," *American Economic Journal: Applied Economics*, 2012, *4* (1), 164–92.

Giné, Xavier and Ghazala Mansuri, "Money or management? A field experiment on constraints to entrepreneurship in rural Pakistan," *Economic Development and Cultural Change*, 2021, *70* (1), 41–86.

Goldberg, Pinelopi Koujianou and Tristan Reed, "Presidential Address: Demand-Side Constraints in Development. The Role of Market Size, Trade, and (In)Equality," *Econometrica*, 2023, *91* (6), 1915–1950.

Good, Charles M, "Markets in Africa: A Review of Research Themes and the Question of Market Origins (Marchés africains: recension des thèmes de recherche et de la question de l'origine des marchés)," *Cahiers d'études africaines*, 1973, pp. 769–780.

Goyal, Aparajita, "Information, direct access to farmers, and rural market performance in central India," *American Economic Journal: Applied Economics*, 2010, *2* (3), 22–45.

Hardy, Morgan and Gisella Kagy, "It's getting crowded in here: experimental evidence of demand constraints in the gender profit gap," *The Economic Journal*, 2020, *130* (631), 2272–2290.

Haushofer, Johannes and Jeremy Shapiro, "The short-term impact of unconditional cash transfers to the poor: Experimental evidence from Kenya," *The Quarterly Journal of Economics*, 2016, *131* (4), 1973–2042.

_ and _, "The long-term impact of unconditional cash transfers: Experimental evidence from Kenya," *Busara Center for Behavioral Economics, Nairobi, Kenya*, 2018.

_, Paul Niehaus, Carlos Paramo, Edward Miguel, and Michael W Walker, "Targeting impact versus deprivation," Technical Report, National Bureau of Economic Research 2022.

Hay, Margaret, "Local trade and ethnicity in western Kenya," *African Economic History Review*, 1975, *2* (1), 7–12.

Heil, Anton, Munshi Sulaiman, Oriana Bandiera, and Robin Burgess, "Graduation," Technical Report, CEPR Discussion Papers 2024.

Hill, Polly, "Notes on Traditional Market Authority and Market Periodicity in West Africa," *The Journal of African History*, 1966, *7* (2), 295–311.

Hodder, Bromwell W, "Rural periodic day markets in part of Yorubaland," *Transactions and Papers (Institute of British Geographers)*, 1961, (29), 149–159.

_, "Distribution of markets in Yorubaland," *Scottish Geographical Magazine*, 1965, *81* (1), 48–58.

Hornbeck, Richard, Guy Michaels, and Ferdinand Rauch, "Identifying agglomeration shadows: Long-run evidence from ancient ports," Technical Report, National Bureau of Economic Research 2024.

Hotelling, Harold, "Stability in competition," *The Economic Journal*, 1929, *39* (153), 41–57.

Hsiang, Solomon M, "Temperatures and cyclones strongly associated with economic production in the Caribbean and Central America," *Proceedings of the National Academy of sciences*, 2010, *107* (35), 15367–15372.

Hussam, Reshmaan, Natalia Rigol, and Benjamin N Roth, "Targeting high ability entrepreneurs using community information: Mechanism design in the field," *American Economic Review*, 2022, *112* (3), 861–898.

Ikegami, Munenobu, Michael R Carter, Christopher B Barrett, and Sarah Janzen, "Poverty traps and the social protection paradox," *The economics of poverty traps*, 2017, pp. 223–256.

Jedwab, Remi and Alexander Moradi, "The permanent effects of transportation revolutions in poor countries: Evidence from Africa," *Review of Economics and Statistics*, 2016, *98* (2), 268–284.

Jensen, Robert, "The digital provide: Information (technology), market performance, and welfare in the South Indian fisheries sector," *The Quarterly Journal of Economics*, 2007, *122* (3), 879–924.

_ **and Nolan H Miller**, "Market Integration, Demand, and the Growth of Firms: Evidence from a Natural Experiment in India," *American Economic Review*, 2018, *108* (12), 3583–3625.

Kansikas, Carolina, Anandi Mani, and Paul Niehaus, "Customized cash transfers: Financial lives and cash-flow preferences in rural Kenya," 2023.

Kapor, Adam, Mohit Karnani, and Christopher Neilson, "Aftermarket Frictions and the Cost of Off-Platform Options in Centralized Assignment Mechanisms," Working Paper 30257, National Bureau of Economic Research July 2022.

Karlan, Dean, Beniamino Savonitto, Bram Thuysbaert, and Christopher Udry, "Impact of savings groups on the lives of the poor," *Proceedings of the National Academy of Sciences*, 2017, *114* (12), 3079–3084.

Kinnan, Cynthia, Krislert Samphantharak, Robert Townsend, and Diego Vera-Cossio, "Propagation and insurance in village networks," *American Economic Review*, 2024, *114* (1), 252–284.

Leight, Jessica, Daniel O Gilligan, Michael Mulford, Alemayehu Seyoum Taffesse, and Heleene Tambet, "Aspiring to more? New evidence on the effect of a light-touch aspirations intervention in rural Ethiopia," Technical Report 2021.

_ **, Josué Awonon, Abdoulaye Pedehombga, Rasmané Ganaba, and Aulo Gelli**, "How light is too light touch: The effect of a short training-based intervention on household poultry production in Burkina Faso," *Journal of Development Economics*, 2022, *155*, 102776.

List, John A, Azeem M Shaikh, and Yang Xu, "Multiple hypothesis testing in experimental economics," *Experimental Economics*, 2019, *22* (4), 773–793.

Lösch, August, *The Economics of Location*, Yale University Press, New Haven, Conn., 1954.

McKenzie, David, “Beyond baseline and follow-up: The case for more T in experiments,” *Journal of Development Economics*, 2012, *99* (2), 210–221.

— **and Susana Puerto**, “Growing markets through business training for female entrepreneurs: A market-level randomized experiment in Kenya,” *American Economic Journal: Applied Economics*, 2021, *13* (2), 297–332.

Mel, Suresh De, David McKenzie, and Christopher Woodruff, “One-time transfers of cash or capital have long-lasting effects on microenterprises in Sri Lanka,” *Science*, 2012, *335* (6071), 962–966.

Melitz, Marc J, “The impact of trade on intra-industry reallocations and aggregate industry productivity,” *Econometrica*, 2003, *71* (6), 1695–1725.

Merfeld, Joshua D and Jonathan Morduch, “Poverty at Higher Frequency,” 2022.

Millán, Teresa Molina, Karen Macours, John A Maluccio, and Luis Tejerina, “Experimental long-term effects of early-childhood and school-age exposure to a conditional cash transfer program,” *Journal of Development Economics*, 2020, *143*, 102385.

Mills, Edwin S and Michael R Lav, “A model of market areas with free entry,” *Journal of Political Economy*, 1964, *72* (3), 278–288.

Misha, Farzana A, Wameq A Raza, Jinnat Ara, and Ellen Van de Poel, “How far does a big push really push? Long-term effects of an asset transfer program on employment trajectories,” *Economic Development and Cultural Change*, 2019, *68* (1), 41–62.

Morgan, Kari Lock and Donald B Rubin, “Rerandomization to improve covariate balance in experiments,” *The Annals of Statistics*, 2012, *40* (2), 1263–1282.

Mullainathan, Sendhil and Eldar Shafir, *Scarcity: Why having too little means so much*, Macmillan, 2013.

Muralidharan, Karthik, Mauricio Romero, and Kaspar Wüthrich, “Factorial designs, model selection, and (incorrect) inference in randomized experiments,” *Review of Economics and Statistics*, 2023, pp. 1–44.

Oberfield, Ezra, Esteban Rossi-Hansberg, Pierre-Daniel Sarte, and Nicholas Trachter, “Plants in space,” *Journal of Political Economy*, 2024, *132* (3), 867–909.

Poll, Moritz, “Micro-enterprise saturation: Critical mass or overcrowding?,” Technical Report 2026.

— , “Seasons of change: Experimental evidence on the optimal timing for poverty alleviation,” Technical Report 2026.

Pryce, Joseph, Marty Richardson, and Christian Lengeler, “Insecticide-treated nets for preventing malaria,” *Cochrane Database of Systematic Reviews*, 2018, (11).

Ritzwoller, David M, Joseph P Romano, and Azeem M Shaikh, “Randomization inference: Theory and applications,” *Journal of Political Economy: Microeconomics*, 2025.

Román, Miguel O, Zhuosen Wang, Qingsong Sun, Virginia Kalb, Steven D Miller, Andrew Molthan, Lori Schultz, Jordan Bell, Eleanor C Stokes, Bhartendu Pandey et al., “NASA’s Black Marble nighttime lights product suite,” *Remote Sensing of Environment*, 2018, *210*, 113–143.

Rosenstein-Rodan, Paul N., “Notes on the theory of the ’big push’,” in “Economic Development for Latin America: Proceedings of a conference held by the International Economic Association” Springer 1957, pp. 57–81.

Schreiner, Mark, “Simple Poverty Scorecard® Tool Malawi,” 2019.

Schurtz, H, “Das afrikanische Gewerbe. 35,” *Preisschrift der Fürstlich Jablonowskischen Gesellschaft zu Leipzig*, 1900.

Scott, Earl P, “Subsistence, markets, and rural development in Hausaland,” *The Journal of Developing Areas*, 1978, *12* (4), 449–469.

Sedlmayr, Richard, Anuj Shah, and Munshi Sulaiman, “Cash-plus: Poverty impacts of alternative transfer-based approaches,” *Journal of Development Economics*, 2020, *144*, 102418.

Skinner, G William, “Marketing and social structure in rural China, Part I,” *The Journal of Asian Studies*, 1964, *24* (1), 3–43.

Smith, Robert HT, “Periodic market-places and periodic marketing: Review and prospect—I,” *Progress in Human Geography*, 1979, *3* (4), 471–505.

_ , “Periodic market-places and periodic marketing: Review and prospect—II,” *Progress in Human Geography*, 1980, *4* (1), 1–31.

Stern, Nicholas, “The optimal size of market areas,” *Journal of Economic Theory*, 1972, *4* (2), 154–173.

Storeygard, Adam, “Farther on down the road: Transport costs, trade and urban growth in sub-Saharan Africa,” *The Review of Economic Studies*, 2016, *83* (3), 1263–1295.

Sulaiman, Munshi, Nathanael Goldberg, Dean Karlan, and Aude de Montesquiou, “Eliminating extreme poverty: Comparing the cost-effectiveness of livelihood, cash transfer, and graduation approaches,” 2016.

Svensson, Jakob and David Yanagizawa, "Getting prices right: The impact of the market information service in Uganda," *Journal of the European Economic Association*, 2009, *7* (2-3), 435–445.

Thiessen, Alfred H, "Precipitation averages for large areas," *Monthly Weather Review*, 1911, *39* (7), 1082–1089.

Thomas, Northcote W, "The Week in West Africa," *The Journal of the Royal Anthropological Institute of Great Britain and Ireland*, 1924, *54*, 183–209.

Tóth, L Fejes, *Lagerungen in der Ebene auf der Kugel und im Raum*, Vol. 65, Springer Verlag, 1953.

Townsend, Robert M, "Village and larger economies: The theory and measurement of the Townsend Thai project," *Journal of Economic Perspectives*, 2016, *30* (4), 199–220.

Trimingham, J. Spencer, *The Influence of Islam upon Africa*, 2nd ed., Longman, 1980.

Valdivia, Martin, "Business training plus for female entrepreneurship? Short and medium-term experimental evidence from Peru," *Journal of Development Economics*, 2015, *113*, 33–51.

Vitali, Anna, "Consumer search and firm location: Theory and evidence from the garment sector in Uganda," *Working Paper*, 2022.

von Carnap, Tillmann, "Remotely-sensed market activity as a short-run economic indicator in rural areas of developing countries," *Available at SSRN 3980969*, 2022.

_ , "Rural marketplaces and local development," 2023.

Voronoï, Georges, "Nouvelles applications des paramètres continus à la théorie des formes quadratiques. 1$^{\text{er}}$ mémoire. Sur quelques propriétés des formes quadratiques positives parfaites.," *Journal für die reine und angewandte Mathematik (Crelles Journal)*, 1908, (133), 97–102.

_ , "Nouvelles applications des paramètres continus à la théorie des formes quadratiques. 2$^{\text{ème}}$ mémoire. Recherches sur les parallélloèdres primitifs.," *Journal für die reine und angewandte Mathematik (Crelles Journal)*, 1908, (134), 198–287.

Wahl, Fabian, "Does medieval trade still matter? Historical trade centers, agglomeration and contemporary economic development," *Regional Science and Urban Economics*, 2016, *60*, 50–60.

Weiss, Daniel J, Tim CD Lucas, Michele Nguyen, Anita K Nandi, Donal Bisanzio, Katherine E Battle, Ewan Cameron, Katherine A Twohig, Daniel A

Pfeffer, Jennifer A Rozier et al., "Mapping the global prevalence, incidence, and mortality of Plasmodium falciparum, 2000–17: A spatial and temporal modelling study," *The Lancet*, 2019, *394* (10195), 322–331.

Wood, Leslie J, "The temporal efficiency of the rural market system in Kenya," *East African Geographical Review*, 1973, *1973* (11), 65–70.

_ , "Unpublished Ph.D. thesis," 1973.

_ , "Population Density and Rural Market Provision (Densité de population et marchés ruraux)," *Cahiers d'études Africaines*, 1974, pp. 715–726.

_ , "Spatial interaction and partitions of rural market space," *Tijdschrift voor economische en sociale geografie*, 1974, *65* (1), 23–34.

World Bank, "The State of Economic Inclusion Report," Technical Report, Partnership for Economic Inclusion 2021.

_ , "The State of Economic Inclusion Report," Technical Report, Partnership for Economic Inclusion 2024.

Wydick, Bruce, "When Are Cash Transfers Transformative?," 2020.

Young, Alwyn, "Channeling Fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results," *The Quarterly Journal of Economics*, 2019, *134* (2), 557–598.

Zerubavel, Eviatar, *The seven day circle: The history and meaning of the week*, University of Chicago Press, 1989.