Essays in Development Economics: technology adoption and behavior change

by

Thomas Woldu Assefa

(Under the Direction of Ellen McCullough)

Abstract

In the first chapter, we evaluated a large scale government fertilizer promotion effort in Ethiopia. In the last decade, the Government of Ethiopia has implemented several major investments and policies seeking to improve fertilizer use by farmers. After creating a national, high resolution digital soil map, the government identified specific soil nutrient deficiencies and then established five fertilizer blending facilities to produce fertilizer products that were tailored to the nutrient profile of surrounding soils. In parallel, it established more than 30,000 demonstration plots nationwide to educate farmers about the returns to blended fertilizer and thus bolster demand. We evaluate the impact of these large-scale government investments on fertilizer adoption using a difference – in – difference (DID) approach and detailed panel data at the plot, farm and community level.

In the second chapter, we conducted a randomized control trial in Ethiopia to test if phone calls made to deliver repeated reinforcement messages affect gender norms and change men's behavior in terms of their participation in household chores. Our intervention lasted for three months, following a face-to-face gender norms training, reinforcing training messages through six phone calls made on a biweekly basis. Based on detailed baseline and endline data on division of labor in household chores, we found that men's participation increased with no effect on women's burden.

Index words: Fertilizer Adoption, Market Access, Agricultural Extension, Demonstration Plots, Behavior Change Communication, Difference in Difference (DID), Gender Norms, Household Chores,

Essays in Development Economics: technology adoption and behavior change

by

Thomas Woldu Assefa

BA, Mekelle University, Ethiopia, 2006
M.Sc., Norwegian University of Life Sciences (UMB), Norway, 2010

A Dissertation Submitted to the Graduate Faculty of the
University of Georgia in Partial Fulfillment of the Requirements for the Degree.

Doctor of Philosophy

Athens, Georgia

2022

©2022
Thomas Woldu Assefa
All Rights Reserved

Essays in Development Economics: technology adoption and behavior change

by

Thomas Woldu Assefa

Major Professor: Ellen McCullough

Committee: Jeffrey H. Dorfman
Mateusz Filipski
Bart Minten

Electronic Version Approved:

Ron Walcott
Dean of the Graduate School
The University of Georgia
May 2022

Dedication

I dedicate this dissertation to the Almighty God for his guidance throughout my life; to the memory of my father (Woldu Assefa) and my mother (Berhane Woldemariam), who taught me my ABCs and sent me to school and always believed in me and thought me to always achieve more; to my boys Elnathan Thomas Woldu and Nolawi Thomas Woldu; and most importantly to my loving wife, Bancheamlak Haile Gebregiyorgis, for standing beside me, for her unreserved support, and words of encouragement; and to my big brother (Mitiku Woldu), who told me not to take off my eyes from the books; to my unique grandmother Weletemariam Etansa , for everything she did for me and most importantly for teaching me how to multiply numbers; to the entire Woldemariam family (the Woldemariams) for their support and encouragement.

Acknowledgments

My sincere gratitude goes to my major advisor, Dr.Ellen McCullough, for her time, support, advice, and encouragement throughout my study. I am very grateful for all the guidance she provided to ensure that I successfully complete the PhD program. I would also like to acknowledge Prof. Jeffrey H. Dorfman, Dr. Mateusz Filipski and Dr. Bart Minten for their invaluable advice and comments which were very useful in strengthening this research. Thank you for serving on the committee and for your contribution to the success of my graduate school. I would also like to express my sincere gratitude to Dr. Guush Berhane, who guided me throughout my career and supported me during my study. I would like also to express my sincere gratitude to the Agricultural and Applied Economics Department of UGA, for the financial support and other benefits I received through graduate assistantship.

I would like to express my special thanks to my uncle, Dr. Tekalign Woldemariam, for his motivation and advice throughout my graduate school and even beyond. I am lucky to have you, Tekish! I am also grateful to the support I received from Dr. Mulugeta Demiss, who provided me data on more than 30000 demonstration plots. I also would like to thank the Standing Panel on Impact Assessment (SPIA) of the Consultative Group for International Agricultural Research (CGIAR) for partially funding this study.

Contents

Acknowledgments **v**

List of Figures **vii**

List of Tables **ix**

1 Evaluation of large-scale government effort to improve fertilizer use in Ethiopia **1**

1.1 Introduction . . . 1
1.2 Background . . . 3
1.3 Theoretical Model . . . 5
1.4 Data . . . 10
1.5 Empirical strategy . . . 10
1.6 Results . . . 17
1.7 Robustness checks . . . 28
1.8 Conclusions . . . 31

Appendices **33**

A Figures **33**

B Tables **43**

C Goodman – Bacon Decomposition **46**

Bibliography **52**

2 Engaging men in household chores and child care through mobile phone-delivered Behavior Change Communications (BCC) **54**

2.1 Introduction . . . 54
2.2 Sample . . . 56
2.3 Experimental design and randomization . . . 57
2.4 Outcome variables, Hypotheses, and estimation strategy . . . 60

2.5 Results 64
2.6 Heterogeneity analysis 75
2.7 Discussions and conclusion 80

Bibliography **82**

List of Figures

1.1 Household Indirect Utility with fixed transaction costs . . . 8
1.2 Timing of the datasets used and the establishment of fertilizer blending facilities . . . 11
1.3 Pre-trend test: Overall fertilizer adoption: Without and with controls . . . 17
1.4 Pre-trend test: DAP fertilizer adoption: Without and with controls . . . 17

A.1 Fertilizer use application rate and number of extension package users . . . 33
A.2 DAP Vs NPS fertilizer uses (CSA Data) . . . 34
A.3 Adoption rate by fertilizer type (IFPRI-FtF Data) . . . 34
A.4 Fertilizer imports and apparent consumption in Ethiopia from 2010-2018. . . . 35
A.5 Locations of five fertilizer blending plants in Ethiopia. . . . 36
A.6 Blended fertilizer adoption 2014 . . . 37
A.7 Blended Fertilizer adoption 2017 . . . 38
A.8 Cost Surface: the darker the cells the slower the speed . . . 39
A.9 Cost distance map after all blending facilities are established (2016): the time it takes to the nearest blending facility . . . 40
A.10 Cost Path: the least cost path from households to blending facilities. . . . 41
A.11 Cost distance map in 2014: the time it takes to the nearest blending facility . . . 42

C.1 Goodman-Bacon decomposition of the Effects of Fertilizer Blending Facilities on NPS Fertilizer Adoption: AGSS and FtF . . . 47
C.2 Goodman-Bacon decomposition of the Effects of Fertilizer Blending Facilities on DAP Fertilizer Adoption: AGSS and FtF . . . 48
C.3 Goodman-Bacon decomposition of the Effects of Fertilizer Blending Facilities on Overall Fertilizer Adoption: AGSS and FtF . . . 49
C.4 Difference-in-Differences with Variation in Treatment Timing: Three Groups . . . 50
C.5 The Four Simple (2x2) Difference-in-Differences Estimates from the Three Group Case 51

2.1 Treatment effects on the number of tasks performed by the woman, the man, and others 66
2.2 Treatment effects on 'performing at least one task' by the woman, the man, and others . 67
2.3 Treatment effects on 'performing at least one task' by types of members . . . 68
2.4 Treatment effects on the number of tasks performed by types of members . . . 68
2.5 Treatment effects on the probability of participating in different types of household chores 70

2.6 Treatment effects on children's probability of participating in different types of household chores . . . 71
2.7 Treatment effects on adult men's and non-members' probability of participating in different types of household chores . . . 72
2.8 The effect of treatment on the Women's Level of satisfaction in the amount of time that Men contribute towards household chores . . . 73
2.9 The effect of treatment on women's level of satisfaction in the amount of time that men contribute towards household chores, by activity . . . 74
2.10 Treatment effect heterogeneity by age: Number of tasks performed by the Man and Sons 75
2.11 Treatment effect heterogeneity by age: Performing at least one task by the Man and Sons 76
2.12 Treatment effect heterogeneity by age: Probability of Performing tasks: the man and the woman . . . 77
2.13 Treatment effect heterogeneity by age: Probability of Performing tasks: Daughters and Sons . . . 77
2.14 Treatment effect heterogeneity by the man's participation at baseline . . . 78
2.15 Treatment effect heterogeneity by the man's participation at baseline . . . 79

List of Tables

1.1 Effects of Fertilizer Blending Facilities on NPS Fertilizer Adoption: DID Models-LPM 19
1.2 Effects of Fertilizer Blending Facilities on DAP Fertilizer Adoption: DID Models-LM . 20
1.3 Effects of Fertilizer Blending Facilities on Overall Fertilizer Adoption: DID Models-LM 21
1.4 The Effects of Fertilizer Blending Facilities on Fertilizer Adoption: Callaway and Sant'Anna Estimator 22
1.5 Effects of Demonstration Plots on NPS Fertilizer Adoption: DID Models-LPM 23
1.6 Effects of Demonstration Plots on DAP Fertilizer Adoption: DID Models-LM 24
1.7 Effects of Demonstration Plots on Overall Fertilizer Adoption: DID Models 25
1.8 Effects of Extension Services on Fertilizer Adoption: Fixed Effect Models 27
1.9 Effects of Fertilizer Blending Facilities on Fertilizer Adoption: by alternative radius for services areas of blending facilities 29
1.10 Effects of Fertilizer Blending Facilities on Fertilizer Adoption: Continuous DID model 31

B.1 Effects of Fertilizer Blending Facilities on NPS Fertilizer Adoption: FD Models 43
B.2 Effects of Fertilizer Blending Facilities on DAP Fertilizer Adoption: FD Models 44
B.3 Effects of Fertilizer Blending Facilities on Overall Fertilizer Adoption: FD Models . . . 44
B.4 Effects of Extension Services on Fertilizer Adoption: FD Models 45
B.5 Distribution of 2013/14 NPS demonstration plots by region and crop. 45

C.1 Goodman-Bacon decomposition of the Effects of Fertilizer Blending Facilities on NPS Adoption 47
C.2 Goodman-Bacon decomposition of the Effects of Fertilizer Blending Facilities on DAP Adoption 48
C.3 Goodman-Bacon decomposition of the Effects of Fertilizer Blending Facilities on Overall Fertilizer Adoption 49

2.1 Comparison of the treatment and control groups at baseline 65
2.2 The effect of treatment on women's perceptions about changes in the level of men's participation in chores 74

Chapter 1

Evaluation of large-scale government effort to improve fertilizer use in Ethiopia

1.1 Introduction

While improved varieties of wheat and rice are the most widely appreciated feature of Asia's Green Revolution, the adoption of chemical fertilizers also played a key role in productivity growth. Despite continuous efforts to replicate Asia's agricultural transformation in Africa, the adoption of chemical fertilizer has not reached the desired level and many farmers apply far below the recommended rates per hectare (Duflo et al., 2011; Sheahan & Barrett, 2017). Some African countries, however, have seen tremendous growth in fertilizer use by smallholder farmers. For example, fertilizer use doubled in Ethiopia between 2013 and 2018. During this time, the Government of Ethiopia (hereafter the government) promoted fertilizer adoption through major policy initiatives designed to improve the supply of fertilizer and improve farmers' demand for it. This paper assesses the impact of these government fertilizer promotion efforts on fertilizer adoption.

On the supply side, Ethiopia invested in improving the availability of blended fertilizers designed to address soil nutrient deficiencies. The Ethiopian Agricultural Transformation Agency (ATA) began to digitally map all of Ethiopia's soil resources in 2012, establishing a national soil information system. The project covered 748 out of 832 woredas,[1] gathering hundreds of thousands of soil samples. The mapping effort uncovered major soil micronutrient deficiencies, with 96% of Ethiopian soils classified as poor in micronutrients such as phosphorus, sulfur, zinc, boron, copper and iron. Policymakers then concluded that the dominant fertilizer types that has been in use in Ethiopia; DAP (which contains Nitrogen and Phosphorus) and urea (which contains Nitrogen), would not address these soil nutrient deficiencies in most locations. The ATA, recommending that farmers instead use blended fertilizer containing the deficient nutrients, established five Fertilizer Blending Facilities (FBFs) in the four major regions of the country

[1] Woreda is the third administrative level in Ethiopia, following regions and zones.

(ATA, 2019). These FBFs, constructed between 2014 and 2016 and owned by Farmers' Cooperative Unions (FCU), blended NPS fertilizer (which contains Sulphur in addition to what DAP fertilizer provides) with additional locally recommended nutrients and supplied these blended fertilizers to farmers [2].

In order to promote demand for improved fertilizers, the government has invested heavily in educating farmers about the new blended fertilizer products, establishing more than 30,000 side by side demonstration plots (DPs) at Farmer Training Centers (FTCs) and on model farms to show how crops respond to blended fertilizers (compared to its conventional alternative, DAP). [3] The government estimated that, on average, 1,000 farmers visited each demonstration plot, thus building awareness among more than 30 million famers across the four regions where fertilizer was promoted.

This study assesses the impacts of these large-scale efforts on fertilizer adoption. To the best of our knowledge, this is the first time these fertilizer promotion efforts were rigorously evaluated. Using Geographic Information Systems (GIS) data, we define a service area for each fertilizer blending facility based on travel time to the nearest blending facility[4]. Then, using detailed household level panel data which collected information at the plot, household, and community levels, we estimate the causal impact of FBF establishment on farm level fertilizer adoption. We use a difference-in-difference (DID) estimator, exploiting temporal and spatial variation in the rollout of the FBFs, to identify the causal impact of establishing FBFs on fertilizer adoption and use. With this approach, we control for time invariant confounding factors, such as differences in soil quality and farmer skills. We also include year and woreda (i.e., district) fixed effects to control for unobserved heterogeneity over time and across space, respectively.

We find that the FBFs facilitated the government's desired shift from DAP to blended fertilizer types, but it did not increase the number of farmers adopting fertilizer. The DPs, on the other hand, increased use of NPS+ blended fertilizer but did not result in dis-adoption of DAP, thus contributing to an increase in overall fertilizer adoption. These results are consistent across three different datasets available to us and robust to checks using the recently proposed Callaway & Sant'Anna (2020) estimator, which uses different control groups for the treatment. We further decomposed our results, into all possible two-groups/two-periods (two-by-two) DID estimates, using a Goodman-Bacon decomposition (Goodman-Bacon, 2018).

By providing an *ex-post* evaluation of large-scale fertilizer promotion policies, this paper contributes to a literature that addresses the importance of market access for technology adoption by smallholder farmers. Second, we contribute to the literature on farmers' market access, which highlights the effect of transaction and transport cost on technology adoption. Within this literature, our study is closely related to Porteous (2020), which uses data from 230 regional markets across sub-Saharan Africa to document increased production and fertilizer use effect from a reduction in trade costs separating producers from input and output markets. We also build on work by Minten et al. (2013), who examine the "last mile(s)" from the input distribution center to the farmer in the chemical fertilizer and improved seed distribution

[2] From now on, we refer to blended fertilizer as NPS fertilizer. It should be noted that there is variation from one location to another in terms of what additional nutrients are blended with NPS depending on the results of the digital soil mapping for the area. But all of them use NPS as the base fertilizer type on top which additional nutrients are added. It is not easy for farmers to distinguish between the different blended fertilizer types and they refer to all of them as NPS fertilizer.

[3] See the distribution of these demonstration plots by region on Appendix B, Table B5.

[4] We compute travel time from each farm to each blending facility using the GPS coordinates of the blending facilities, accounting for road network locations, road types, water bodies and land cover.

system in Ethiopia, finding that remoteness contributes substantially to transaction and transportation costs associated with fertilizer use. Our results are consistent with those of Suri (2011), who find that heterogeneity in the costs of accessing hybrid maize technology provides an explanation for low adoption rates of technologies that could dramatically increase farm profits. Aggarwal et al. (2018) find that, in rural Tanzania, a 50% reduction in travel costs could double technology adoption by smallholders.

Our findings point to important policy implications. First, we show that access to fertilizer markets matter. Investments in transport and communication infrastructure reduced the fixed costs that households incur to access fertilizer markets, thus improving adoption. Second, we find that efforts to promote blended fertilizers (vs the conventional DAP) in DPs led to increased use of blended fertilizer without a corresponding dis-adoption of DAP. This suggests that households choose their own take-away lesson from extension agents' advice. Our results confirm the potential for achieving fertilizer adoption growth by investing in farmer education and agricultural extension services.

The rest of the paper is organized as follows. We first provide institutional background on Ethiopia's fertilizer procurement and distribution system. Then we present a theoretical model, in section three, for fertilizer adoption in which farmers face fixed transaction costs in obtaining fertilizer. We then explain our data in section four, followed by detailed empirical strategy in section five. Section six presents the results, and section seven shows checks to confirm the robustness of our findings. Finally, we conclude in section eight.

1.2 Background

Ethiopia is largely an agrarian economy, agriculture contributing close to 40% and 80% of GDP and employment as of 2017 respectively. Historically, the sector used to have the largest share until the services sector took over in 20010/11. The sector is believed to be the key for economic growth in the country. Within agriculture, the most important subsector was crop production, representing 32 percent of GDP. Cereals represent the largest share within crop production. The share of cereals in total crop output was 63 percent in 2004/05 and remained at the same level up to 2013/14 (Bachewe et al., 2015).

There have been efforts to improve use of modern inputs, like fertilizer, since the 1960s when fertilizer was first introduced to the country. As of now, government controls the decision how much to import, execution of import and distribution in consultation with local farmers' cooperatives. But there have been several policy shifts that have shaped and re-shaped fertilizer supply in the country. Rashid et al., (2013) categorizes these policy shifts into five phases: (i) complete government control (1967-1992), (ii) partial liberalization, with private sector entry and elimination of subsidies (1992-1996), (iii) competition among public, private, and regional holding companies (1997-2000), (iv) exit of private companies (2001-2006), and (v) since 2007, the exit of regional holding companies and the entry of farmers' cooperatives as the distribution channel, with AISE (Agricultural Input Supply Enterprise) as the sole fertilizer importer since 2008 (Rashid et al., 2013; Spielman et al., 2013). Since 2015 AISE merged with other four enterprises to establish a new corporate called Ethiopian Agricultural Businesses Corporation.

The decision how much to import comes from a fertilizer demand assessment which starts at the lowest administrative level (Kebeles). Development Agents (DAs), who are working as agricultural extension agents of the Kebeles, provide expected local fertilizer demand information to the woreda agricultural bureau. These kebele level data are then aggregated to come up with woreda level fertilizer demand and are sent to Zonal offices, which in turn are aggregated and sent to the Regional Bureau of Agriculture. The Regional Bureau of Agriculture replicates the aggregation to estimate regional fertilizer demand for the coming year. Some regions also consider the production targets set in the Growth and Transformation Program (GTP) (AFAP, 2012; Rashid et al., 2013). The federal Ministry of Agriculture will then aggregate the regional estimates to decide national expected fertilizer demand for the coming agricultural year (Rashid et al., 2013).

The imported fertilizer first arrives at Djibouti Port, 950 km from Addis Ababa (AFAP, 2012). It will then be transported to either the regional cooperative unions' warehouses or AISE central warehouses, the large majority (about 70%) being transported to AISE warehouses to be transferred later to the cooperatives(AFAP, 2012). AISE has 33 warehouses strategically located in the country(AFAP, 2012; Rashid et al., 2013). Primary cooperatives then get the fertilizer from cooperative unions or directly from AISE if they don't have cooperative union and distribute it to farmers (AFAP, 2012; Rashid et al., 2013; Wossen et al., 2019).

The price farmers pay for fertilizer is a function of the cost of handling and administration and largely transport cost from the port to the primary cooperatives (Rashid et al., 2013). The fertilizer price that unions pay to AISE is determined by adding transport cost from the port to AISE warehouses and other handling and administrative costs including profit margin. The union and primary cooperative prices are determined by each regions' bureau of agriculture based on AISE hand over prices (Rashid et al., 2013). The bureau determines the price in consultation with the cooperatives considering transportation cost from warehouses to primary cooperatives, profit margins and administrative cost for cooperatives, bank interest, warehouse rent, as well as loading and unloading cost at cooperatives stores(AFAP, 2012; Rashid et al., 2012). In some regions, like Tigray and SNNP, availability of carryover stocks and associated storage costs are considered in the price determination (AFAP, 2012; Rashid et al., 2013).

Despite existence of other costs considered in the fertilizer price determination, transport cost takes the largest share (Rashid et al., 2013). For example, in 2012, the additional cost from Djibouti Port to a cooperative warehouse was $152.3 for DAP and transport cost took about 74% of the increase (AFAP, 2012). Rashid et al., (2013) computed the share of transport cost from the total difference in DAP price between farmgate price and landed costs at Djibouti. It was found to be $138 for Tigray, $113 for Amhara, $110 for Oromia, and $127 for SNNPR (Rashid et al., 2013). Depending on the region, they found transport costs alone to account for 64–80 percent of these price differentials.

Data from Central Statistical Agency's Agricultural Sample Survey shows that amount of fertilizer use increased rapidly in the last decade. Figure A1 shows fertilizer use tripled in eight years since 2009/10 and doubled in five years since 2012/13. DAP and UREA have been the only fertilizer types the country was supplying to farmers for the years preceding 2014/2015 (AFAP, 2015; Bachewe et al., 2015). Since then, the country decreased the amount of DAP it imports and instead started importing and blending fertilizers

of different nutrient types like Nitrogen, Phosphorus and Sulfur. Figures A2 and A3 show the story using two different datasets.

1.3 Theoretical Model

Our aim is to study fertilizer adoption as one form of market participation and model it following Singh et al. (1986), also adopted by Key et al. (2000) and Alene et al. (2008). Starting with the canonical agricultural household model, Key et al. (2000) introduce a fixed cost to input acquisition, which exists alongside proportional (or variable) transaction costs. Our analysis treats access to fertilizer distribution locations as fixed transaction costs, which are household-specific but do not vary with the quantity of fertilizer purchased.

In our simple static model, an agricultural household maximizes utility subject to a production technology constraint:

$$Max\, U = U(p_q Q - I_x[m + \, p_x X]) \qquad \text{(Eq 1)}$$

$$s.t\, Q = Q(X, \boldsymbol{Z}) \qquad \text{(Eq 2)}$$

where p_q is output price, Q is total production, X is the quantity of fertilizer used, p_x is price of fertilizer, m represents the fixed transaction costs (sunk costs) incurred from using fertilizer ($m \geq 0$), I_x is an indicator function equal to one if fertilizer is used (i.e., $X > 0$), zero otherwise, and $\boldsymbol{Z}$ is a vector of household and farm characteristics including other inputs used. The production technology and the utility functions represent well-behaved production and utility functions that $Q^{'} > 0, Q^{''} < 0, U^{'} > 0$ and $U^{''} < 0$.

The fixed transaction costs represented by m includ$e$ costs incurred by the household to buy fertilizer, which are not included in the price of fertilizer. These include search and information costs, bargaining and decision costs, and policing and enforcement costs. But it can also include transportation cost to the fertilizer market (distribution center)(Niehans, 1987) . In the context of this study, cooperative unions who own the blending facilities distribute blended fertilizer largely within their own primary cooperatives. Households around these primary cooperatives, regardless of membership, can purchase fertilizer from the cooperative. Households far from the primary cooperatives that operate blending facilities face higher transaction costs when purchasing fertilizer, due to longer transport distances. Households living far from a cooperative operating a blending facility may also face higher search and information costs in procuring fertilizer from distant cooperatives. Because DAP is marketed differently, through the extension system, farmers' fixed costs in accessing DAP are determined based on their distance to the nearest point of DAP distribution, which is different than the location of the NPS fertilizer facility.

The objective function (Eq 1) can be extended to clearly show the two different fixed transaction costs households face in the DAP and the NPS fertilizer markets;

$$Max\, U = U(p_q Q - p_{NPS} NPS - p_D D - I\,(NPS)\, m_{NPS}(S_1) - I\,(D)\, m_D(S_2)) \quad \text{(Eq 3)}$$

$$s.t\, Q = Q(D,\ NPS,\ \boldsymbol{Z}) \quad \text{(Eq 4)}$$

where p_{NPS} and p_D are prices of NPS and DAP respectively, NPS and D are amount of NPS and DAP, m_{NPS} and m_D are fixed transaction costs of the NPS and DAP fertilizer markets respectively, $I\,(NPS)$ is an indicator function taking value one if the household used NPS fertilizer and zero otherwise and $I\,(D)$ is interpreted similarly for use of DAP fertilizer. S_1 is households' travel time from the nearest NPS fertilizer distributing primary cooperative and S_2 is households' travel time from the nearest DAP fertilizer distributing primary cooperative. We express m_{NPS} as a function of S_1 and m_D as a function of S_2 because these transaction costs increase with distance from the fertilizer distribution centers.

The Lagrangian of the maximization problem of the household is:

$$\mathcal{L} = U\,(p_q Q - p_{NPS} NPS - p_D D - I\,(NPS)\, m_{NPS}\,(S_1) - I\,(D)\, m_D\,(S_2)) + \lambda\, Q(D,\ NPS,\ \boldsymbol{Z})$$

where λ is the Lagrange multiplier associated with the technology constraint. But it can also be written without the Lagrange multiplier as follows:

$$L = U\,(p_q Q(D,\ NPS,\ \boldsymbol{Z}) - p_{NPS} NPS - p_D D - I\,(NPS)\, m_{NPS}\,(S_1) - I\,(D)\, m_D\,(S_2))$$

The fixed transaction costs create discontinuities in the Lagrangian. Therefore, the optimal solution cannot be found by simply solving the first order conditions. The solution is decomposed in to two steps – solving first for the optimal (interior) solution conditional on participation in the specific fertilizer market (adoption), and then deciding whether to participate in the fertilizer market or not (whether to adopt or not) based on comparisons of expected utility between the two discrete alternatives (i.e., adopting at the optimal level vs. not-adopting).

The first order conditions conditional on adoption (market participation) lead to the usual optimality conditions for input use that the value of the marginal product of each type of fertilizer equals the price of fertilizer:

$$p_q \frac{\partial Q}{\partial D} = p_D \quad \text{(Eq 5)}$$

$$p_q \frac{\partial Q}{\partial NPS} = p_{NPS} \quad \text{(Eq 6)}$$

The above equations show that the amount of fertilizer demanded is a function of its own price, output price and the elements in the production technology that determine fertilizer productivity. More importantly, the equations imply that the quantity of fertilizer demanded, conditional on demanding fertilizer, is not affected by the fixed transaction costs. Once the fixed cost of participating in the market is paid, fixed transactions costs do not affect the volume (Alene et al., 2008).

We address the choice of whether to adopt or not by comparing the utility from adopting to the utility from not adopting. Indirect utility is defined as $v(\boldsymbol{P}, y)$, where $\boldsymbol{P}$ is a vector of prices including output price (p_q), fertilizer price (p_x), and other input prices and y is household income. Suppressing other input prices for simplicity, the indirect utility after adopting type x fertilizer at price p_x is represented by $v(p_q, p_x, y)$. For convenience, we define $y_0(p_x)$ as the household income before incurring fixed transaction costs to participate in type x fertilizer market. Household income after incurring the fixed transaction costs is represented by $y_0\left(p_x\right) - m_x$. Therefore, households compare the following indirect utilities to choose whether to adopt or not-adopt type x fertilizer:

$$
\begin{aligned}
V^A &= v(p_q, p_x, y_0\left(p_x\right) - m_x) && \text{If adopting fertilizer type } x \qquad \text{(Eq 7)} \\
V^{NA} &= v(p_q, y) && \text{If not-adopting fertilizer type } x
\end{aligned}
$$

where V^A with the superscript A stands for indirect utility of adopters while V^{NA} with superscript NA stands for non-adopters.

To make the comparison easier, we depict the two indirect utilities as a function of the price of fertilizer x on Figure 1.1 below. The vertical line represents the utility V^{NA} obtainable by a non-adopter household, which is independent of fertilizer price. An adopter household who does not face fixed transaction costs for fertilizer type x (in our context, a farmer this household can be considered to be around NPS fertilizer distributing primary cooperative) and the utility of this household is represented by V_0^A and utility is shown to decrease as price of fertilizer increases (half-line A^0B^0). Let fertilizer price $p_x{=}p_x^0$ solves $V_0^A{=}V^{NA}$, meaning a household facing low fixed transaction costs will be indifferent between adopting and not adopting at $p_x{=}p_x^0$ (also shown as point B^0). At prices lower than p_x^0, such a household with low fixed transaction costs will be better off adopting and enjoy utility level V_0^A on half-line A^0B^0 and a household will be better off not-adopting at any price above p_x^0 and enjoy utility level V^{NA}. Therefore, the optimal choice is represented by path A^0B^0C.

It can be shown using (Eq 7) that an increase in fixed transaction costs lowers household income and therefore decreases the utility level at any given price of fertilizer. This is shown in the figure by a shift of the utility curve to the left (down). An adopter household facing fixed transaction costs (in our context this household can be considered to be far from NPS fertilizer distributing primary cooperatives) achieve utility level V^A(half-line AB) and let fertilizer price $p_x{=}p_x^1$ solves $V^A{=}V^{NA}$, meaning a household facing fixed transaction costs will be indifferent between adopting and not adopting at $p_x{=}p_x^1$ (also shown as point B). Such a household with fixed transaction costs will be better off adopting and enjoy utility level V^A on half-line AB and the household will be better off not-adopting at any price above p_x^1 and enjoy utility level V^{NA}. Therefore, the optimal choice is represented by path ABC.

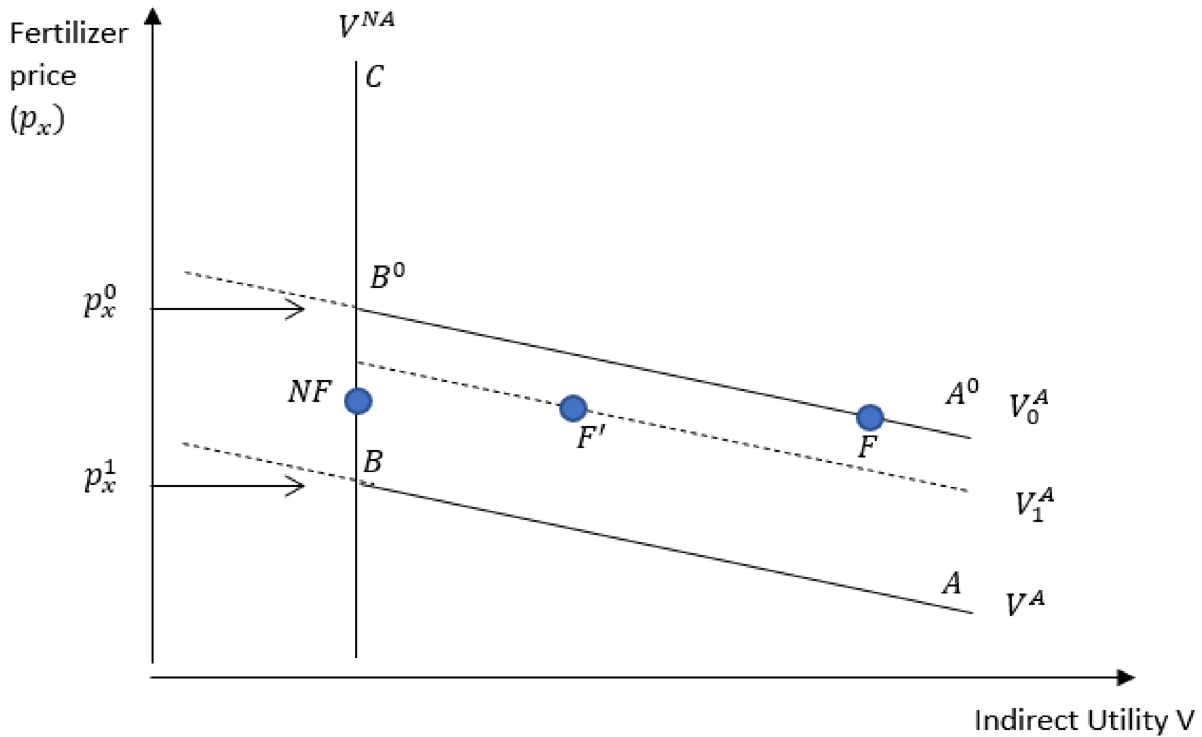


Figure 1.1: Household Indirect Utility with fixed transaction costs

Comparing the cases with and without fixed transaction costs (those close to NPS distributing primary cooperatives and those far from NPS distributing primary cooperatives), we can see that households facing higher fixed transaction costs would need lower price of fertilizer to choose adopting fertilizer compared to households with lower fixed transaction costs. Therefore, although our model predicted that fixed transaction costs do not affect the amount of fertilizer demanded, they do affect the adoption decision. The analysis also shows that the adoption decision is not only a function of fixed transaction costs but also a function of all the determinants of the profitability of fertilizer adoption, i.e. output price, fertilizer price and the elements in the production technology.

In summary, the theoretical model predicts that:

$$I(x) = f\left(p_q, p_x, m_x, \boldsymbol{Z}\right) \quad \text{(Eq 8)}$$

$$X = f\left(p_q, p_x, \boldsymbol{Z}\right) \quad \text{(Eq 9)}$$

1.3.1 Measuring distance to fertilizer distribution center and the critical level of distance $(\overline{S_1})$

With fixed transaction costs (m_{NPS}) assumed to increase in distance to fertilizer distribution centers, and indirect utility increasing in income (y), we expect the household's probability of adoption to decrease in fixed transaction costs. A household facing low fixed transaction cost has an optimal path A^0B^0C, represented by point F (standing for fertilizer used) on Figure A1 above. The household is better for adopting fertilizer as it is on the half-line A^0B^0 portion of the optimal path. If we increase the fixed transaction cost by a smaller amount (smaller than the amount which would shift the indirect utility curve from V_0^A to V^{A}) to shift the curve from V_0^A to V_1^A, the utility level of the household would decrease making the household to move from point F to $F^{'}$. Notice that the increase in the fixed transaction cost that comes from the increase in distance from fertilizer distribution center decreased the utility level, but the household is still better off adopting. But if the increase in S_1 is large enough to shift the indirect utility curve to V^{A}, the household moves to point NF (standing for no fertilizer use) and is no longer better off adopting (will therefore optimally choose not to adopt).

The above analysis shows there is some critical level of distance $\overline{S_1}$, where a household will be indifferent between adopting and not adopting fertilizer. The level of $\overline{S_1}$is determined by the variables in the calculation of net revenue (i.e. price of output, price of input and the elements in the production technology). Therefore, *ceterus paribus*, households beyond this critical distance ($\overline{S_1}$) are better off not to adopt while households living with in $\overline{S_1}$ distance from the distribution center are better off adopting.

Measuring S_1 is not straightforward. Data on distance to the nearest primary cooperative are easier to collect as households are usually aware about the locations of primary cooperatives. They also more likely know whether the primary cooperative distributes fertilizer (e.g., DAP). Given the availability of NPS fertilizer is limited to some primary cooperatives, households face an information barrier about which primary cooperatives distribute NPS fertilizer[5]. It is therefore difficult to collect information about distance to the nearest NPS distribution center from households.

Although we do not directly observe the locations of the primary cooperatives that distribute NPS fertilizer, we do observe the locations of cooperative unions that distribute NPS fertilizer to member primary cooperatives. For each household, we generate a variable representing the distance from that household's geolocation to the nearest NPS distributing cooperative union, which we use as a proxy for S_1. This is justifiable because what matters is the difference in S_1 among households and the difference in S_1 does not change even if we use the distance from NPS distributing cooperative unions to the households for our analysis. To show this, let us define distance between NPS distributing cooperative unions and primary cooperatives as d_1. Therefore, $d_1 + S_1 = T$, where T is the total distance from NPS distributing cooperative unions to households. Consider two households ($x\ and\ y$) where x is living near to a NPS-distributing primary cooperative and y is living far (at S_1 distance) from a NPS distributing primary

[5]This is one of the reasons why households living far away from NPS distributing primary cooperative face higher transaction cost.

cooperative. Therefore:

$$S_1^y - S_1^x = S_1 - 0 = S_1$$
$$T^y - T^x = S_1^y + d_1^y - (S_1^x + d_1^x) = S_1^y - S_1^x = S_1 \text{ since } d_1^x = d_1^y$$

1.4 Data

In the empirical analysis, we use three datasets that are available to us. First is the Ethiopian Living Standards Measurement Survey – Integrated Surveys in Agriculture (LSMS-ISA) dataset. The LSMS-ISA dataset has been collected in three rounds: in 2011/12, 2013/14 and 2015/16. It is collected through systems of multi-topic, nationally representative panel household surveys with a strong focus on agriculture. It collects georeferenced data, information on agricultural inputs used including fertilizer, farm and demographics characteristics of households, among others.

Second is the Agricultural Sample Survey (AgSS) data, a nationally representative annual dataset covering the entire rural parts of the country, collected by the Central Statistical Agency (CSA) of Ethiopia. The data contain information at the field level on whether fertilizer is used and how much, whether it uses improved seed, how much of the land is allocated to the production of the crop, whether the field is irrigated, what crops are planted and what quantity is produced.

Third, the Feed the Future (FtF) impact evaluation dataset, collected by IFPRI's Ethiopian Strategy Support Program (ESSP) in collaboration with CSA, is a household panel dataset representative of the areas where FtF was implemented, also termed as FtF Zone of Influence "ZOI" by the project implementers, which are 149 woredas/districts all in the rural parts of the five major regions in Ethiopia; Oromia, Amhara, Tigray, SNNP and Somali. The FtF data were collected in 2013, 2015 and 2018 from households within the FtF ZOI, in 56 woredas (districts) in Ethiopia. The FtF survey has questionnaires collecting information about the household and the community separately. The household questionnaire collects information on a variety of topics including household demographics, plot level farm characteristics: such as whether fertilizer is used, amount used, size, soil fertility, slops, what crop is planted, how much is harvested ...etc. The community questionnaire collects information, among others, about households' access to health and educational services, access to agricultural and health extension services. It also asks about village level formal and informal groups and access to agricultural inputs like fertilizer.

1.5 Empirical strategy

Our aim is to measure the impact of access to fertilizer blending facilities on fertilizer adoption. Five fertilizer blending facilities were established: one in 2014, and four in 2015 and 2016. Figure A5 shows the location of the blending plants on a map of Ethiopia. Figure 1.2 below depicts the time when the fertilizer blending facilities were established along with the timing of the datasets used on this study. The Becho Weliso blending facility, established in West Shewa zone of Oromia, was inaugurated on June 7, 2014. The remaining four of them were established in 2015 and 2016: Gibe Dedesa blending facility in

Oromia, Enderta blending facility in Tigary, and Merkeb in Amhara regions were established in late 2015, and Melik blending facility SNNP region was established in 2016.

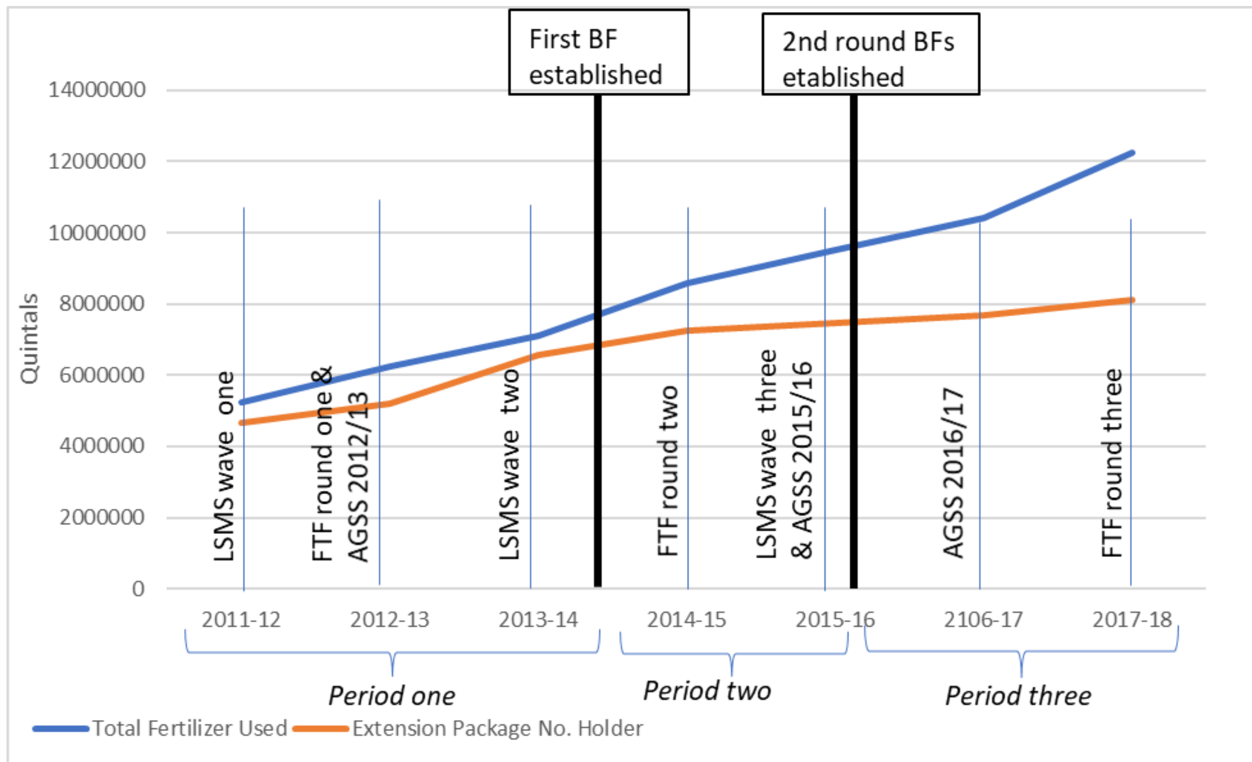


Figure 1.2: Timing of the datasets used and the establishment of fertilizer blending facilities

We rely on temporal and spatial variations in terms of service from these blending facilities for identification. The temporal variation is the difference in the time when the blending facilities were established, and the spatial variation is the difference in proximity of farmers to these blending facilities. In terms of temporal variation, all blending facilities were originally planned to start operating at the same time, but due to delays in procurement of building structures and blending facilities, four of them except the Becho Woliso blending facility were delayed[6]. Since the reason for the delay is unrelated to our outcome of interest, it is an exogenous variation that can be used for identification. Specifically, to estimate the impact of the blending facility established earlier, areas that are getting fertilizer blending facilities later can be used as controls until they get their respective blending facility. In terms of spatial variation, although the blending facilities are strategically located, there is a limit to the geographic area the facilities can serve. The area beyond that limit can serve as a control group area. Obviously, farmers closely surrounding the blending facilities benefit more than others—and are therefore treated, and those who are too far are not served—and are therefore controls.

The key challenge is identifying the effective service area of a blending facility. We use a Geographic Information System (GIS) to define a service area for each fertilizer blending facility. We use GPS coordi-

[6]The Becho Woliso blending facility was funded by the USAID and the other four were funded by the World Bank via the Ethiopian Agricultural Transformation Agency (ATA).

nates of the blending facilities together with GIS datasets on roads, water bodies and land cover, to come up with travel time from any location in the whole country to the nearest blending facility. Following Pozzi & Robinson (2008), we assign different travel speeds to different road types and to distances without road coverage. We then compute the fastest travel time for any location in the country to the nearest blending facility. Inspired by Porteous, (2019), the idea is to assign areas which can be reached at a specific time interval from a blending facility as the service area for that specific blending facility[7]

Figure A9 maps the travel time to each of the blending facilities. In 2014, there was only one blending facility operating: the Becho Woliso blending facility. Therefore, only the area around this blending facility was served. We have shown a different version of Figure A9 when only the Becho Weliso blending facility was operating in Figure A11. We start with using the area which can be reached in 5 hours from the blending facilities as the service area for the blending facilities. This is the area shown by the first two shades around the blending facilities in the figure. Later, we show the sensitivity of results for different cutoff points: 4 hours, 3 hours, 2 hours, including the 5 hours travel time.

Figure A6 shows adoption of blended fertilizer in 2014, for within and outside the service area of the Becho Woliso blending facility. As expected, adoption of blended fertilizer is concentrated within the service area of the blending facility. Adoption of blended fertilizer is higher (10%) within the service area and just 4% outside the service area. Figure A7 shows adoption of blended fertilizer in 2017, for within and outside the service areas of all five blending facilities. Adoption is higher within the service areas (54%), compared to outside the service areas (36%).

These results, obviously, are not showing causality, and are only descriptively presenting differences within and outside the service areas. More importantly, the locations where the blending facilities are located are likely to be where households use more fertilizer. Therefore, there is a reverse causality problem that fertilizer adoption could be causing establishment of fertilizer blending facility, instead of blending facilities causing fertilizer adoption. Next, we explain the estimation techniques we used that address this and other related problems. We use a difference – in – differences approach which requires the trends in fertilizer use over time, prior to treatment, to be parallel between treated and control areas. Specifically, households in treated areas can exhibit higher adoption of fertilizer than control areas but are assumed to have similar change in fertilizer adoption over time prior to treatment. We also test if this assumption holds or not.

Using the different data sets described in the data section, we are able to observe variables of interest in three time periods as depicted on Figure 1.2 above: time period one; before any blending facility was established, time period two; after the first blending facility was established but before the rest of the blending facilities were established and time period three; after all the blending facilities were established. The data before any bending facility was established provides baseline information, the data during the

[7]We first create two GIS road datasets namely: major roads and minor roads by merging many of the classifications we found in the roads data and converted the whole country's landcover raster to a 1km by 1km grid cell size raster. We then built a model using ArcGIS's model builder in which we converted the roads datasets to a cell size the same as the land cover's raster cell size. We then used "raster calculator tool" to assign different speeds to different types of roads and land cover and "cell statistics tool" to create a cost surface (see figure A8). The cost surface is then used as an input for another model in which we used the "cost distance tool" with blending facilities' locations and the 1km by 1km cell size land cover to compute the time it takes from any cell in the whole country to the nearest blending facility (see figure A9).

second time period is post-treatment information for Becho Weliso blending facility service area, and the data after all the blending facilities were established is post-treatment information for all blending facilities. To this end, we can divide households in to three groups based on the time when their nearest blending facility was established and their proximity to the facilities; group one (I) includes households within the service area of the Becho Weliso blending facility which was established first, group two (II) includes households in the service areas of the four blending facilities established later and group three (III) includes households who are outside the service areas of any of the blending facilities.

Following Jensen (2007), let $\overline{Y}_{g,t}$ represent the average value of the outcome variable in group g in time period t. We can then compare the change in $\overline{Y}$ in group one (I) between time period 1 and 2, i.e., before versus after establishment of the Becho Woliso fertilizer blending facility, relative to the change over the same rounds in groups two (II) and three (III)[8], i.e.;

$$\left(\overline{Y}_{I,2} - \overline{Y}_{I,1}\right) - \left(\overline{Y}_{II,2} - \overline{Y}_{II,1}\right) \quad \text{and} \quad \left(\overline{Y}_{I,2} - \overline{Y}_{I,1}\right) - \left(\overline{Y}_{III,2} - \overline{Y}_{III,1}\right) \qquad \text{(Eq 10)}$$

In the same way, to assess impact of the blending facilities established later, we can compare the change in $\overline{Y}$ in group two (II) between time periods 2 and 3, i.e., before versus after establishment of the additional four fertilizer blending factories, relative to the change over the same rounds in groups one (I) and three (III), i.e.;

$$\left(\overline{Y}_{II,3} - \overline{Y}_{II,2}\right) - \left(\overline{Y}_{I,3} - \overline{Y}_{I,2}\right) \quad \text{and} \quad \left(\overline{Y}_{II,3} - \overline{Y}_{II,2}\right) - \left(\overline{Y}_{III,3} - \overline{Y}_{III,2}\right) \qquad \text{(Eq 11)}$$

This comparison can be formally done by estimating either a First Difference (FD) model or a Difference – in – Difference (DID) model, which also allow us to control for other factors that may affect the outcome variable.

The FD model is presented in equation 12, below:

$$\Delta Y_{it} = \beta_1 + \beta_2 Year_{t=3} + \beta_3 \Delta FBF_{it} + \gamma \Delta \boldsymbol{X}_{it} + \mu_w + \Delta\varepsilon_{it} \quad (Eq\ 12) \qquad \text{(Eq 12)}$$

where y_{it} is the outcome variable for individual i at time period t; FBF_{it} is an indicator of being in a Fertilizer Blending Facility (FBF) service area for household i at time t; X_{it} is a vector of covariates including time varying demographic and farming characteristics, and ε_{it} is iid random variable. In this equation, we take first difference of all but the year and woreda fixed effects (μ_w). The woreda fixed effects will wash out if we take their first difference, therefor we use them as they are and will be controlling for

[8] We can either compare the changes in group one with the changes either in group two or group three separately, as in equation (10). But we can also compare the changes in group one with the changes in groups two and three together, by pooling the observations from group two and group three to one comparison group.

any woreda differences of the first differences. For the year fixed effects, (Wooldridge, 2012) shows taking first difference of the year dummies or putting them as they are doesn't affect the coefficients of the other variables and it only affects the intercept. Therefore, we include them without first differencing them. The subscript $t = 3$ on the year variable shows that the year being referenced is within the third time period, based on the three time periods classification explained above.

The first difference model, presented in equation 12, is advantageous in terms of controlling for time invariant heterogeneities that affect the outcome variable. By first differencing, time invariant heterogeneities, e.g., farmer's ability, are washed out. The AgSS dataset is not panel at the household level means that we cannot run first differencing at the household level. We are however able to run the first difference model using the FtF and LSMS datasets. Since it is not our main specification, we put results from the first difference model in the appendix.

The comparisons in equations 1 and 2 can also be estimated using a formal DID approach specified as follows:

$$Y_{it} = \alpha + \beta_1 TREAT_i + \ \beta_2 POST_t + \beta_3^{DD} \left(TREAT_i * POST_t \right) + \gamma \boldsymbol{X}_{it} + \mu_w + \varepsilon_{it} \quad \text{(Eq 13)}$$

where $TREAT_i$ is an indicator variable for the treated unit, and $POST_t$ is an indicator variable for observations post-treatment, $\boldsymbol{X}_{it}$ is a set of covariates that may affect the rate at which fertilizer adoption change over time, and ε_{it} is iid random variable. We have included woreda fixed effects (μ_w) to control for any woreda level differences, and β_3^{DD} is the coefficient of interest measuring the impact of treatment.

In our case, treatment (establishment of the FBF) is happening at different periods of time for group one and group two. Therefore, estimation of equation 13 is not straight forward. In such case, where there is variation in treatment timing, the following two-way fixed effects (TWFE) model has been used to estimate the impact of treatment (Baker et al., 2021; Goodman-Bacon, 2018):

$$Y_{it} = \alpha_i + \ \alpha_t + \beta^{twfe} D_{it} + \varepsilon_{it} \quad \text{(Eq 14)}$$

where α_i and α_t are unit and time period fixed effects, $D_{it} = TREAT_i * POST_t$ is an indicator for a treated unit in treated time periods, and the main effects for $TREAT_i$ and $POST_t$ are subsumed by the unit and time fixed effects (Baker et al., 2021; Goodman-Bacon, 2018). The superscripts on β in equations 13 and 14, indicate from which model they are estimated. DD stands for the standard Difference in Difference model and $twfe$ stands for two-way fixed effects model.

Recent developments in econometrics show that the coefficient estimated by the TWFE DID approach, as in equation 14, is likely biased and can even be of the wrong sign altogether (Athey & Imbens, 2022; Baker et al., 2021; Borusyak & Jaravel, 2017; Callaway & Sant'Anna, 2020; Chaisemartin & Haultfœuille, 2020; Goodman-Bacon, 2018; Sun & Abraham, 2020). The bias arises from two main sources. First, the coefficient is a weighted average of all possible 2x2 DID estimators, and the weights are proportional to the size of the groups in each 2x2 DID and variance of the treatment variable (Goodman-Bacon, 2018). Second, when there is treatment effect heterogeneity across groups and over time, and when already treated groups are used as controls, changes in their treatment effects over time get subtracted from the

changes of later-treated groups. The recent literature on DID calls this negative weighting, and it can lead to opposite sign TWFE model estimate as compared to the true ATT (Baker et al., 2021; Goodman-Bacon, 2018).

Goodman-Bacon (2018) uses a three - groups case, where treatment happens at two different time periods, and graphically illustrates how treatment timing works with the objective of decomposing the single TWFE DID estimate (β^{twfe}) in equation 14 to all possible 2x2 DID estimates within this design. Since we have exactly the same case, three - groups with treatments happening at two different time periods, we have reproduced the graphs used by Goodman-Bacon (2018) on Figure C4 and Figure C5 in the Appendix. The three groups are: those who are never treated (U), treated early at t_k^* (k), and treated later at $t_l^* > t_k^*$ (l), see Figure C4. This creates four possible 2x2 DID designs that compares: early treated units to never treated units ($\widehat{\beta}_{ku}^{DD}$), late treated units to never treated units ($\widehat{\beta}_{lu}^{DD}$), early treated units to late treated units ($\widehat{\beta}_{kl}^{DD,k}$) using the late treated group as control before they get treated, and late treated units to early treated units ($\widehat{\beta}_{lk}^{DD,l}$) using the early treated group as control after they get treated. These four possible designs are presented in Figure C5, and each of these are estimable by equation 13. Goodman-Bacon, (2018) shows $\widehat{\beta}^{twfe}$ is a weighted average of these estimates:

$$\widehat{\beta}^{twfe} = s_{ku}\widehat{\beta}_{ku}^{DD} + s_{lu}\widehat{\beta}_{lu}^{DD} + s_{kl}\widehat{\beta}_{kl}^{DD,k} + s_{lk}\widehat{\beta}_{lk}^{DD,l} \qquad \text{(Eq 15)}$$

where the s_{ku} is the weight for $\widehat{\beta}_{ku}^{DD}$ and so on. Baker et al., (2021) suggests researchers to decompose the TWFE DID estimate using the Goodman-Bacon, (2018) diagnostic (Goodman-Bacon decomposition), as it shows the percentage of the estimate that is driven by different types of treatment timing comparisons, and how the weighted average ATT in each group differs.

Although the literature has not yet settled on a "best" solution to the problems discussed above, there are some suggestions provided by few econometricians (Baker et al., 2021; Borusyak & Jaravel, 2017; Callaway & Sant'Anna, 2020; Chaisemartin & Haultfœuille, 2020; Sun & Abraham, 2020; Wooldridge, 2021). The most notable alternatives suggested so far are: Callaway & Sant'Anna, (2020) estimator, Sun & Abraham, (2020) estimator and Stacked Regression estimator. Baker et al., (2021) provides summary and simulations of each of these estimators. All three of these suggestions aim at creating a "clean" control group to avoid problems associated with comparing late and early treated groups (Baker et al., 2021; Callaway & Sant'Anna, 2020; Sun & Abraham, 2020). Most recently, Wooldridge (2021) shows how the Sun & Abraham(2020) TWFE estimator can be extended to allow covariates to enter flexibly, and argues under a conditional common trends assumption, the extended TWFE approach identifies the average treatment effects for different cohort/time period treatment effects.

The problems highlighted by the literature, and the alternatives being suggested, point at two main take-away messages. First, it suggests not to rely on the single TWFE model estimate, presented in equation 14, and if possible, provide the different 2x2 DID estimates of the different pairs. Second, it suggests avoiding using already treated groups as a control. Therefore, we estimate the simple 2x2 DID model for the first treatment (establishment of the Becho Woliso FBF) and the second treatment (establishment of the remaining FBFs) separately and we do not use the already treated group (households within the service area of the Becho Woliso FBF) as control. Specifically, we estimate the model in equation 13 using

appropriate subset of our datasets to estimate the effect of the first treatment and the second treatment. For the effect of the first treatment, we estimate the following equation:

$$Y_{it} = \alpha + \beta_1 TREAT_i + \ \beta_2 POST_t + \beta_3^{DD} \left(TREAT_i * POST_t \right) + \gamma \boldsymbol{X}_{it} + \mu_w + \varepsilon_{it} \quad \text{(Eq 16)}$$

where the observations come from the first and second time-periods only (i.e. $t = 1,\ 2$); $POST_t = 1$ for $t = 2$ and, zero for $t = 1$; $TREAT_i = 1$ for households within the service area of the Becho Woliso FBF and zero otherwise. Note that we are using both the never treated and later treated households together as controls. But we also present the case when we are using either of them as controls through the Goodman-Bacon decomposition we present later. For the effect of the second treatment, we estimate the following equation after excluding those who are already treated:

$$Y_{it} = \alpha + \beta_1 TREAT_i + \ \beta_2 POST_t + \beta_3^{DD} \left(TREAT_i * POST_t \right) + \gamma \boldsymbol{X}_{it} + \mu_w + \varepsilon_{it} \quad \text{(Eq 17)}$$

where $POST_t = 1$ for $t = 3$ and, zero for $t = 1\ and\ 2$; $TREAT_i = 1$ for households within the service area of the FBFs established later and zero otherwise. Note that we are not using the already treated group, households within the service area of Becho Woliso FBF, as control. Meaning, only never treated households are used as control.

We then try to estimate a single coefficient for the treatment effect, which brings the effects from the first treatment and second treatment together. We first estimate the 'traditional' TWFE model, equation 14, along with its decomposition to the possible 2X2 DID models, equation 15, using Goodman - Bacon decomposition as suggested by Baker et al. (2021). Following estimation of the TWFE model, which is potentially biased, we present results based on Callaway & Sant'Anna (2020) estimator.

Finally, we present some robustness checks. We estimate results based on different cutoff points of travel time for treatment; four hours, three hours, two hours and one hour. We also presented results based on continuous DID estimator.

Parallel trend assumption is a key assumption for DID models. We do not have data for more than one period prior to the first treatment, and therefore we are unable to do the parallel trend assumption test. However, the AgSS data we are accessing is annually collected and runs from just before the first treatment (2012/13) up to a year after the second treatment (20016/17). This means, using the AgSS data and excluding the observations from the areas treated first, we can conduct a parallel trend assumption test for the second treatment. Meaning, we use never treated and later treated observations, and exclude early treated observations, to conduct the parallel trend assumption test for the second treatment.

The figures below present event study dynamic effects estimates, also produced with Callaway and Sant'Anna's estimator and the associated parallel trend assumption test that tests if all pretreatment ATT's are statistically equal to zero. It does not make sense to do the parallel trend assumption test for the NPS fertilizer adoption models because there was no adoption of NPS fertilizer prior to treatment. Figures 3

and 4 below respectively show the pre-trends based on Overall and DAP fertilizer adoption models, with and without controls[9], and the test that all pretreatment ATT's are statistically equal to zero.

Figure 1.3 shows that the parallel trend assumption does not hold for overall fertilizer adoption model without including controls. But it holds when the controls are included. Similarly, the results on Figure 1.4 show that the parallel trend assumption holds, at least at 5 percent level of significance, with controls included for the DAP fertilizer adoption model.

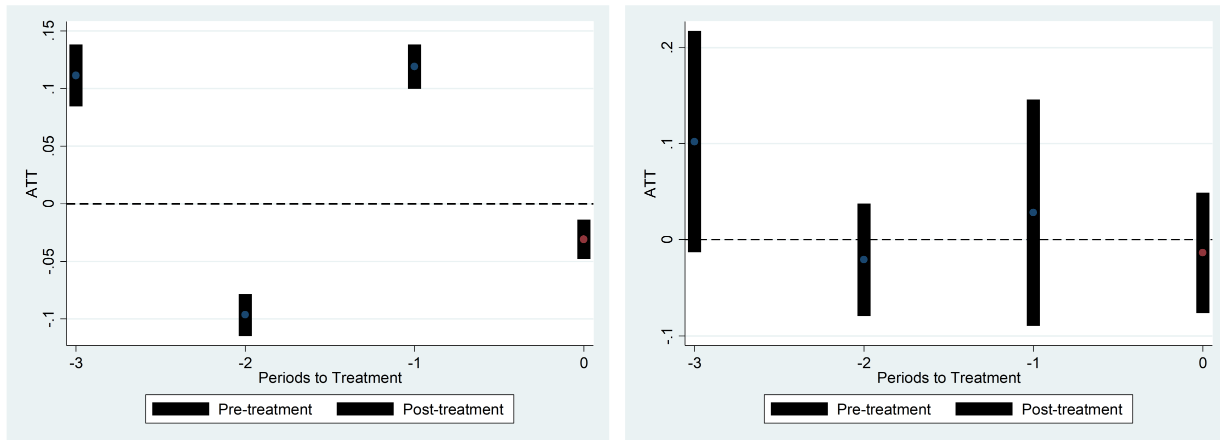


Figure 1.3: Pre-trend test: Overall fertilizer adoption: Without and with controls

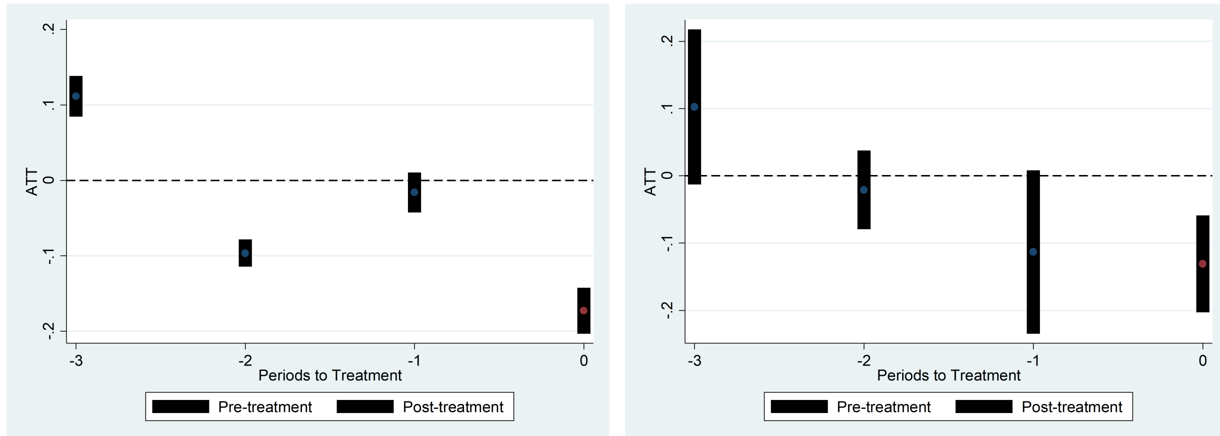


Figure 1.4: Pre-trend test: DAP fertilizer adoption: Without and with controls

1.6 Results

In this section, we present results on the effect of both the supply and the demand side interventions on inorganic fertilizer adoption: adoption of NPS fertilizer, adoption of DAP and adoption of overall fertil-

[9]The controls included are land size, gender of the household head and literacy of the household head. We have included Woreda fixed effects are included in both with and without controls models.

izer. In the first sub section, we assess the effect of establishment of the fertilizer blending facilities (FBFs). We first present results based on coefficients estimated separately for the first treatment and the second treatment. We then present results based on a single coefficient for the treatment effect. The subsequent section will present results on the effects of the demined side intervention, which is establishment of the demonstration plots (DPs). The last sub section will present results on the effects of agricultural extension services in general.

1.6.1 Separate treatment effects of Fertilizer Blending Facilities

Table 1.1 presents the effect of establishing FBFs on adoption of NPS fertilizer by estimating the DID model presented in Equation 13. We present estimates of the first and second treatments separately: the first treatment is establishment of the Becho Woliso FBF and the second treatment is establishment of the rest of the FBFs.

We present estimations using each of the data sets: FtF, LSMS and AgSS. In the first treatment case, the pre-treatment data from each of the datasets comes from 2012/13, except the case for LSMS which is from 2011/12. The post-treatment data from each of the datasets comes from 2015/16, except the case for FtF which is from 2014/15. In the second treatment case, the FtF observations are from 2014/15 and 2017/18 while the AgSS observations are from 2015/16 and 2016/17. We are not using LSMS for the second treatment because the last round of the panel LSMS data was collected in 2015/16 and therefore, it does not have data following the second treatment[10].

In all of the models presented in Table 1.1, we control for woreda-level time invariant heterogeneity with fixed effects, and we cluster standard errors at the EA level. We also control for demographic characteristics like gender and literacy of the household head, farm characteristics like land size. The coefficient of interest in these models, the interaction of the treatment and the post indicator variables, is found to be positive and significant in all of the models. This means households with FBFs around are significantly more likely to adopt NPS fertilizer. This result is found to be consistent across the different datasets: FtF, LSMS and AgSS. We also estimate a first difference model using the panel FtF and LSMS datasets (see Appendix B, Table B1). We are not able to estimate first difference equations using AgSS because the dataset does not comprise a household level panel. In the case of the LSMS data set, we also estimated both the DID and first difference models using wave 2 of the data collected in 2013/14 as pre-treatment data. Our findings are consistent across these alternative specifications.

Unlike the case in some other African countries like Tanzania, the fertilizer market in Ethiopia is highly controlled by the government and farmers access fertilizer mainly from primary cooperatives. Since there is no administrative restriction preventing households who are not around FBFs from buying NPS fertilizer from the primary cooperatives distributing NPS, these results can be interpreted as the effect of changing the fixed costs of accessing NPS. Given that households around FBFs are found on average to be more likely to adopt the NPS fertilizer, as discussed in the theoretical model and assuming households

[10] Please see figure 1.2 for the timing of each round of the different datasets.

are behaving rationally, those who are far from the FBFs must have found it not optimal to incur the fixed costs associated with adopting NPS.

Table 1.1: Effects of Fertilizer Blending Facilities on NPS Fertilizer Adoption: DID Models-LPM

	First Treatment*			Second Treatment**	
	FtF	LSMS	AgSS	FtF	AgSS
Treat_Post	.0353***	.0842**	.116***	.49***	.135***
	(.0133)	(.0404)	(.013)	(.0716)	(.0145)
Treat	-.00674	-.117	-.0199	.0342	-.0447***
	(.0154)	(.0709)	(.0131)	(.0796)	(.00773)
Post	.0291***	.0802***	.0662***	.131***	.0655***
	(.00582)	(.0164)	(.00583)	(.0436)	(.00922)
Land size [in ha]	-.0000307	.000379	16.9***	-.000103*	28.5***
	(.0000251)	(.000336)	(2.37)	(.0000543)	(3.23)
Gender [Female=1]	-.00465	.0169***	-.00709***	-.0426***	-.0114***
	(.00486)	(.00632)	(.00187)	(.00964)	(.00244)
Literacy	.00309	-.00334	.00234	.00483	.00503**
[Can read and write=1]	(.00424)	(.00608)	(.0019)	(.00871)	(.00241)
Constant	.00544	-.0478***	-.0491***	-.172**	-.0483***
	(.00872)	(.00924)	(.00672)	(.0781)	(.00974)
Observations	6512	5844	82507	4227	75437

Note: LPM: Linear Probability Model; Woreda fixed effects are included in all models; Standard errors in parentheses, clustered by EA; * p<0.10, ** p<0.05, *** p<0.01; *FtF observations are from years 2012/13 and 2014/15; LSMS observations are from the years 2011/12 and 2015/16; AgSS observations are from the years 2012/13 and 2015/16. ** FtF observations are from the years 2014/15 and 2017/18; AgSS observations are from the years 2015/16 and 2016/17.

In Table 1.2, we present the results for adoption of DAP fertilizer. The models in Table 1.2 are estimated in the same way as the models in Table 1.1, and all descriptions provided for models in Table 1.1 also apply for the models in Table 1.2, except the change in the outcome variable to adoption of DAP fertilizer. In all of the models, except the last column, the coefficient of interest (i.e., the interaction between "treat" and "post") is found to be negative in sign. The AgSS data on the first treatment and the FtF data on the second treatment show evidence for strongly significant dis-adoption of DAP. This suggests that households close to FBFs are significantly less likely to adopt DAP fertilizer. We also estimated first difference equations using FtF and LSMS datasets (see Appendix B, Table B2). These results, that households near FBFs are less likely to adopt DAP fertilizer, are confirmed when we use a first difference estimator.

Table 1.2: Effects of Fertilizer Blending Facilities on DAP Fertilizer Adoption: DID Models-LM

	First Treatment*			Second Treatment**	
	FtF	LSMS	AgSS	FtF	AgSS
Treat_Post	-.0322	-.0724	-.0914***	-.387***	.0268
	(.0279)	(.0477)	(.0165)	(.058)	(.021)
Treat	.148***	.483***	.0383	.459***	.0728***
	(.0506)	(.0954)	(.0287)	(.0828)	(.024)
Post	.0445**	-.0213	.114***	-.045	-.0415**
	(.0186)	(.029)	(.0114)	(.0551)	(.0162)
Land size [in ha]	.0278**	.00908	84.5***	.00948	90.8***
	(.0118)	(.0115)	(4.66)	(.007)	(5.52)
Gender [Female=1]	-.0678***	.0876***	-.0536***	-.103***	-.0556***
	(.0138)	(.02)	(.00477)	(.0188)	(.0055)
Literacy [Can read and	.038***	.0406***	.0324***	.05***	.0358***
write=1]	(.0119)	(.0157)	(.00404)	(.0131)	(.00438)
Observations	6436	4606	76459	4194	66707

Note: LM: Logit models, and the coefficients are Average Marginal Effects; Woreda fixed effects are included in all models; Standard errors in parentheses, clustered by EA; * p<0.10, ** p<0.05, *** p<0.01; *FtF observations are from years 2012/13 and 2014/15; LSMS observations are from years 2011/12 and 2015/16; AgSS observations are from years 2012/13 and 2015/16; ** FtF observations are from 2014/15 and 2017/18; AgSS observations are from 2015/16 and 2016/17.

The results in Tables 1 and 2 match the general trend in fertilizer use in the country, presented in Figures A2 and A3 in Appendix A, that there is increased use of NPS and decreased use of DAP fertilizer. This is not surprising given the government's policy to shift from DAP to NPS, making NPS more available and gradually taking DAP out of the supply channel. But the results in Table 1.1 and Table 1.2 together point to an interesting finding that the shift from DAP to NPS happened more strongly among households around the FBFs than those far from FBFs. This has important policy implication in terms of the role supply constraints play in the effectiveness of government policy. The FBFs have clearly facilitated the shift from DAP to NPS; households in the areas without the blending facilities are significantly less likely to adopt NPS and dis-adopt DAP.

In Table 1.3, we present estimates of the same models described in Table 1.1 and 1.2 above, except that the outcome variable is changed to overall fertilizer adoption. By overall fertilizer adoption, we mean adoption of fertilizer in general, regardless of which type is adopted. Analyzing the effect on overall fertilizer adoption is important because it shows the net effect of establishing FBFs on fertilizer adoption, given the increase in NPS adoption and the decrease in DAP adoption. The coefficient of interest is statistically insignificant in all the models, except when we use AgSS for the second treatment. We also estimate first difference models using FtF and LSMS datasets and find similar results (see Appendix B3). All in all, we can say that there is no strong evidence to suggest that establishing FBFs improve overall fertilizer adoption. The results show that households around FBFs are equally likely to adopt fertilizer as compared to households far from the FBFs.

Taking the results from Tables 1.1, 1.2 and 1.3 together, we can see that the FBFs facilitated the shift from DAP to NPS fertilizer types but didn't improve overall fertilizer adoption. There could be many reasons why the FBFs did not improve overall fertilizer adoption. It can be related with the quality of NPS compared to DAP, the operation of FBFs and/or farmers awareness towards NPS. The evidence on whether NPS is better than DAP is mixed and detailed information on the operation of FBFs is not easy to access. This suggests the need for further research in this area.

Table 1.3: Effects of Fertilizer Blending Facilities on Overall Fertilizer Adoption: DID Models-LM

	First Treatment*			Second Treatment**	
	FtF	LSMS	AgSS	FtF	AgSS
Treat_Post	-.0184	.00754	.00989	.035	.134***
	(.026)	(.0302)	(.0154)	(.0205)	(.016)
Treat	.151***	.406***	.019	.775	.0398
	(.0427)	(.0768)	(.0261)	(.)	(.0258)
Post	.0763***	.0574***	.228***	.0768	.0739***
	(.0203)	(.0211)	(.00976)	(.)	(.0136)
Land size [in ha]	.03**	.0174	113***	.0257	134***
	(.0123)	(.0259)	(5.5)	(.)	(6.54)
Gender [Female=1]	-.0648***	.122***	-.0559***	-.102	-.06***
	(.0133)	(.0262)	(.00462)	(.)	(.00548)
Literacy [Can read and write=1]	.0455***	.0337**	.0362***	.0696	.0434***
	(.0118)	(.0158)	(.00384)	(.)	(.00447)
Observations	6436	4540	76683	4218	67031

Note: LM: Logit models, and the coefficients are Average Marginal Effects; Woreda fixed effects are included in all models; Standard errors in parentheses, clustered by EA; * p<0.10, ** p<0.05, *** p<0.01; *FtF observations are from years 2012/13 and 2014/15; LSMS observations are from years 2011/12 and 2015/16; AgSS observations are from years 2012/13 and 2015/16; ** FtF observations are from the years 2014/15 and 2017/18; AgSS observations are from the years 2015/16 and 2016/17.

1.6.2 Overall treatment effects of Fertilizer Blending Facilities

So far, we have been discussing the effects of the first treatment and the second treatment separately. From the policy perspective, we would want to know the effect of the treatment in a single coefficient. Given the variation in treatment timing, the traditional approach of getting a single coefficient would be through estimating a TWFE model, Equation 14. As discussed in the empirical strategy section, the TWFE estimate (β^{twfe}) is likely biased due to the mechanism by which weights are attached to all possible two-by-two DID estimates that get into the computation of the single TWFE estimate. As per Baker et al. (2021)'s suggestion, we present the TWFE estimate, all possible two-by-two DID estimates and their associated weights in the Appendix, see Appendix C. Although likely biased, the TWFE estimates reflect the main results we found that FBFs have facilitated the policy of substituting DAP by NPS but did not lead to improvements in overall fertilizer adoption.

We used the Callaway & Sant'Anna (2020) estimator, among the estimators suggested to address the bias on the TWFE model estimate. Table 1.4 below presents the results based on Callaway & Sant'Anna (2020) estimator. We present results from the two datasets: AgSS and FtF. We found positive and significant estimates for NPS fertilizer adoption, negative and significant estimate for DAP fertilizer adoption and conflicting results from the two datasets for overall fertilizer adoption. These results are similar to what we found in Section 6.1, where we presented separate treatment effects for the earlier and later treated observations. Therefore, we conclude that FBFs have facilitated the policy of substituting DAP by NPS but did not lead to improvements in overall fertilizer adoption.

Table 1.4: The Effects of Fertilizer Blending Facilities on Fertilizer Adoption: Callaway and Sant'Anna Estimator

Dataset	NPS		DAP		Overall	
AgSS	.196***	.239***	-.0675***	-.079***	.121***	.176***
	(.0107)	(.0086)	(.0142)	(.0101)	(0130)	(.0084)
FtF	.218***	.218***	-.245***	-.224***	-.038**	-.515***
	(0123)	(.0130)	(.0171)	(.0099)	(.0174)	(.0091)
Woreda FE	No	Yes	No	Yes	No	Yes

Standard errors in parentheses; The data is from 2005, 2008 and 2009; The data is at EA level
* p<0.10, ** p<0.05, *** p<0.01

1.6.3 Effect of Demonstration Plots

In this section, we present the effect of demonstration plot establishment on fertilizer adoption. We follow an approach like the one we used to analyze the effect of FBFs, estimating the DID model presented in equation 13, except that the treatment under consideration is the establishment of demonstration plots within the woreda where the households reside.

Table 1.5 below presents DID model estimates for the effect of DPs on adoption of NPS fertilizer. Woreda fixed effects are included in all models, and we clustered standard errors by Woreda. We also present results from three datasets: FtF, LSMS, and AgSS, as we did in the previous section. The coefficient of interest in these models—interaction of the treatment and the post indicator variables—is found to be positive and insignificant in the first model, where we used 2015 FtF data as a post-treatment observation, while it is found to be positive and strongly significant in the second model, where we used the 2018 FtF data as a post-treatment observation. We also found a positive and significant effect using the AgSS data, while the LSMS data shows an insignificant effect. The difference in results across datasets can be explained by differences in survey timing and geographic coverage. The 2015 FtF data asked about the 2014 meher season in Ethiopia, and this could be too close to the time when the demonstration plots were established (2013/14), to show an impact. Accordingly, we observed a positive effect with the 2018 FtF data, which asked about the 2017 meher season. Although we are not able to see a significant result from the LSMS dataset, the positive sign coupled with the positive and significant effect we found using

the relatively larger AgSS dataset suggests that the demonstration plots have improved adoption of NPS fertilizer.

Table 1.5: Effects of Demonstration Plots on NPS Fertilizer Adoption: DID Models-LPM

	FtF		LSMS	AgSS
	Post=2015	Post=2018		
Treat_Post	.0099	.173***	.0479	.0781**
	(.0161)	(.064)	(.0522)	(.0365)
Treat	-.0126	.325***	.202***	-.0247
	(.0113)	(.0348)	(.026)	(.0184)
Post	.0465***	.393***	.0986***	.101***
	(.00866)	(.0357)	(.019)	(.0121)
Land size [in ha]	-.0000508*	.00535***	.000431	16.2***
	(.0000268)	(.00168)	(.000426)	(2.98)
Gender [Female=1]	-.00483	-.0482***	.0167**	-.00791***
	(.0048)	(.00886)	(.00653)	(.00205)
Literacy [Can read and write=1]	.00272	.00199	-.00322	.00238
	(.00426)	(.00785)	(.00673)	(.00201)
Constant	-.00289	-.128***	-.057***	-.0671***
	(.00858)	(.0235)	(.0106)	(.00795)
Observations	6512	6510	5844	82507

Note: LPM: Linear Probability Model; Woreda fixed effects are included in all models; Standard errors in parentheses, clustered by woreda; * p<0.10, ** p<0.05, *** p<0.01; FtF pre-treatment observation are from year 2012/13; LSMS observations are from the years 2011/12 and 2015/16; AgSS observations are from the years 2012/13 and 2015/16.

Given the major objective of the demonstration plots was to show farmers how better NPS fertilizer performs compared to DAP, it is not surprising to see a positive effect in adoption of NPS fertilizer. In line with this, we expect farmers to dis-adopt DAP fertilizer. Table 1.6 below presents the effect of DPs on DAP fertilizer adoption rates. The coefficient of interest is found to be insignificant in all models, except in the second model, where we find a negative and significant effect when the post-treatment observations come from the 2018 FtF dataset. Therefore, we conclude that demonstration plots did not improve adoption of DAP fertilizer. Although we do not find strong evidence supporting the expectation that DPs will lead to dis-adoption of DAP fertilizer, we have some evidence from the second model that DPs negatively and significantly affected adoption of DAP fertilizer.

Table 1.6: Effects of Demonstration Plots on DAP Fertilizer Adoption: DID Models-LM

	FtF		LSMS	AgSS
	Post=2015	Post=2018		
Treat_Post	.0221	-.171**	.0572	.0629
	(.0274)	(.07)	(.0737)	(.0449)
Treat	.127*	-.0794	-.423***	.344***
	(.0647)	(.0766)	(.000257)	(.0155)
Post	.0211	-.272***	-.0575*	.0615***
	(.0177)	(.0336)	(.0311)	(.0175)
Land size [in ha]	.0278**	.0058*	.00888	84.9***
	(.0117)	(.00306)	(.011)	(6.1)
Gender [Female=1]	-.0685***	-.0444***	.0874***	-.0529***
	(.0138)	(.0115)	(.0217)	(.00558)
Literacy [Can read and write=1]	.0396***	.0291***	.0411***	.0323***
	(.0119)	(.00967)	(.0146)	(.00475)
Observations	6436	6510	4606	76459

Note: LM: Logit models, and the coefficients are Average Marginal Effects; Woreda fixed effects are included in all models; Standard errors in parentheses, clustered by Woreda; * p<0.10, ** p<0.05, *** p<0.01; FtF pre-treatment observations are from year 2012/13; LSMS observations are from the years 2011/12 and 2015/16; AgSS observations are from the years 2012/13 and 2015/16.

Results in Tables 1.4 and 1.5 together show that DPs improved adoption of NPS fertilizer, but did not lead to dis-adoption of DAP. One important qualification of this result emerges from the results based on the FtF datasets. In models where post-treatment observations come from the 2018 FtF dataset, DPs have indeed led to dis-adoption of DAP. The models using post-treatment observations from the 2015 FtF dataset show neither DPs improving adoption of NPS nor leading to dis-adoption of DAP fertilizer. There are two possible explanations for why we find DPs facilitating the intended transition from DAP to NPS with the 2018 FtF dataset, but not with the 2015 FtF dataset. First, the FtF-2015 data record a time period too early for the DPs to have an impact. To clarify, the 2105 FtF survey asked about the 2014 meher season in Ethiopia, and DPs were established in 2013/14. In line with this, when observations from a later year (2018) are used, we observed the positive effect of DPs on NPS and their negative effect on DAP. Second, it could be related with the gradual transition from DAP to NPS that started around 2015 and gets stronger afterwards.

Table 1.7 below presents the effect of DPs on overall fertilizer adoption. The coefficient of interest is found to be positive and strongly significant in all models, except in the first column, where we use FtF-2015 datasets as post-treatment observations. Therefore, we conclude that DPs have significantly improved overall fertilizer adoption. The insignificant result when the post-treatment observations come from the 2015 FtF dataset could be because of the same reason explained for the results in Tables 4 and 5 above that the timing of the survey could be too early for the DPs to create impact.

Table 1.7: Effects of Demonstration Plots on Overall Fertilizer Adoption: DID Models

	FtF		LSMS	AgSS
	Post=2015	Post=2018		
Treat_Post	.0374	.0658*	.116***	.128***
	(.0241)	(.0366)	(.039)	(.0328)
Treat	.11*	.136***	-.47***	.311***
	(.0588)	(.0484)	(.00022)	(.00734)
Post	.0571***	.0815***	.0413**	.212***
	(.0156)	(.0187)	(.0161)	(.0114)
Land size [in ha]	.0299**	.0321*	.0173	113***
	(.0124)	(.0171)	(.0241)	(7.64)
Gender [Female=1]	-.0658***	-.0713***	.122***	-.0564***
	(.0133)	(.0126)	(.0263)	(.00549)
Literacy [Can read and write=1]	.0473***	.0352***	.034*	.0362***
	(.0118)	(.0104)	(.0184)	(.00464)
Observations	6436	6510	4540	76683

Note: LM: Logit models, and the coefficients are Average Marginal Effects; Woreda fixed effects are included in all models; Standard errors in parentheses, clustered by woreda, * p<0.10, ** p<0.05, *** p<0.01; FtF pre-treatment observations are from year 2012/13, LSMS observations are from the years 2011/12 and 2015/16; AgSS observations are from the years 2012/13 and 2015/16.

The findings from Tables 1.5, 1.6 and 1.7 together suggest that DPs have improved adoption of NPS fertilizer, did not lead to immediate dis-adoption of DAP, and have significant increased overall fertilizer adoption. Given our findings that FBFs facilitated the transition from DAP to NPS but did not lead to improvement in overall fertilizer adoption, an interesting story emerges in comparing the effects of the supply-side intervention (establishment of the FBFs) and the demand-side intervention (establishment of the DPs). While the supply-side intervention helped farmers substitute one type of fertilizer (DAP) by the other (NPS), the demand-side intervention improved overall fertilizer adoption. These results suggest the importance of raising awareness among farmers through agricultural extension services, as evidenced by the effect of the DPs. In the next section, we further investigate the effect of broader agricultural extension services on fertilizer adoption in the country.

1.6.4 Effect of Extension Services

In this section, we assess the effect of extension services on fertilizer adoption. The results in the previous section suggested that the demand side intervention, through the establishment of DPs, improved overall fertilizer adoption, which was not achieved by the supply side intervention through the establishment of FBFs. The establishment of DPs is part of a broader extension service program in the country. Ethiopia is one of the few countries with huge public investment in agricultural extension services (Berhane et al., 2018; Blum et al., 2020; Davis et al., 2020). We assess the effect of extension services on fertilizer adoption with the intention of linking our results with the broader extension service program.

Unlike the case in which we analyzed the effect of FBs and DPs using data before and after the specific interventions, we cannot analyze the effect of broader extension services on fertilizer adoption in a similar manner. This is mainly because there is no specific recent year which we can use to categorize data before and after the service: the broader extension services have been in place for many years now. But exploiting the panel nature of our datasets, we estimate fixed effect models in which we control for household or enumeration area fixed effects. Table 1.8 (below) presents the fixed effect models, which are estimated to measure the effects of extension services on NPS, DAP and overall fertilizer adoptions in the first, second and third blocks of the table, respectively. For each dataset, we present household fixed effect and enumeration area fixed effect models, except for the AgSS dataset, for which we present only woreda fixed effect model, as the data is not panel at the household level. We have also estimated first difference models, which are presented in Appendix Table B4 using the FtF and LSMS datasets. We are unable to do so using the AgSS dataset as it is not panel at a household level.

Due to differences in the way each dataset captures extension services provided to households, we are measuring extension service by two different indicators: extension agents' visits and broader extension services. In the FtF models, we use the extension agents' visits variable, which identifies whether the household was visited by an extension agent at least once during the last 12 months. In the LSMS and AgSS models, we use an indicator for a broader extension service that includes services beyond visits by extension agents. The estimated results reflect this difference that the coefficients in the FtF models are relatively smaller than the coefficients in the LSMS and AgSS models. In the case of NPS adoption, the coefficients are even insignificant for the FtF models, while they are strongly significant for the LSMS and AgSS models. All in all, the results show that extension services are positively and significantly associated with fertilizer adoption. These results are consistent with what was found by others: extension services positively and significantly improve adoption of modern technologies including fertilizer (Bachewe et al., 2015; Berhane et al., 2018; Blum et al., 2020; Davis et al., 2020).

Table 1.8: Effects of Extension Services on Fertilizer Adoption: Fixed Effect Models

	FtF		LSMS		AgSS
	HH_FE	EA_FE	HH_FE	EA_FE	W_FE
NPS-LPM					
Visited [Yes=1]	.00179	.00169			
	(.0146)	(.0109)			
Under extension program [yes=1]			.0256**	.033***	1.86***
			(.0118)	(.01)	(.0709)
Observations	9766	9766	8002	8002	80236
DAP-LM					
Visited [Yes=1]	.523***	.619***			
	(.083)	(.0954)			
Under extension program [yes=1]			1.97***	2.69***	1.82***
			(.121)	(.156)	(.0438)
Observations	5655	9685	2858	6300	213897
OVERALL-LM					
Visited [Yes=1]	.628***	.7***			
	(.0958)	(.0999)			
Under extension program [yes=1]			2.87***	3.69***	2.61***
			(.176)	(.182)	(.0439)
Observations	3081	9144	2249	6197	214282
HH fixed effects	Yes	No	Yes	No	No
EA fixed effects	No	Yes	No	Yes	No
Woreda fixed effects	No	No	No	No	Yes

Note: LM: Logit models, and the coefficients on LM are odds-ratios; LPM: Linear Probability Model; HH: Household; EA: Enumeration Area; FE: Fixed Effect; Year FE are included in all models; Standard errors in parentheses, clustered by EA except for HH FE models; * p<0.10, ** p<0.05, *** p<0.01; AgSS observations are from all the years between 2012/13 and 2017/18; FtF observations are all the three rounds in years 2013, 2015 and 2018 and LSMS observations are all the three waves in years 2011/12, 2013/14 and 2015/16.

The evidence that extension services are positively and significantly associated with fertilizer adoption regardless of the type of fertilizer suggests that farmers understand the general message: fertilizers are good to use—instead of focusing on which type to use. This is consistent with the result in Table 1.6: the DPs did not lead to dis-adoption of DAP until later, when DAP was out of the market. The aim of the DPs was to illustrate that the NPS fertilizer performs better than DAP in terms of productivity. As Table 1.6 shows, households around the DPs were more likely to adopt NPS fertilizer but were not more likely to dis-adopt DAP until DAP became less available later in time. This suggests that households also focused on the general message, that any type of fertilizer is good to use, and not only on the specific message, that NPS is better.

1.7 Robustness checks

1.7.1 Sensitivity of results for different cutoff points of treatment

We investigated the sensitivity of results if different cutoff points are used to categorize households to treatment and control groups. Table 1.9 below presents the estimated results for four alternative cutoff points, in which 5 hours, 4 hours, 3 hours and 2 hours of travel time from the blending facilities are considered as alternative radius of service area for the blending facilities. The 5 hours travel time results are the main results we presented and discussed in the previous sections. All models are estimated in the same way as the main results, the only difference being the cutoff point used to categorize households to treatment and control groups.

The results in Table 1.9 show that the main findings from section 1.6 hold when a smaller cutoff point, like 4 hours travel time, is used. The results based on 5 hours and 4 hours cut off points are similar. But when the cutoff point gets smaller and smaller, like 3 hours and 2 hours, coefficients become insignificant. These results make sense because, with smaller cutoff points, households who have been treated may get categorized as controls.

Table 1.9: Effects of Fertilizer Blending Facilities on Fertilizer Adoption: by alternative radius for services areas of blending facilities

		Travel time from blending facilities			
		5 hours	4 hours	3 hours	2 hours
LSMS					
	NPS	.0842**	.118**	.0518	.141
		(.0404)	(.0506)	(.0571)	(.0979)
	DAP	-.458	-.638*	-.375	-.857
		(.311)	(.379)	(.463)	(.663)
	Overall	.0489	.21	.00388	-.285
		(.196)	(.215)	(.276)	(.44)
EtE					
First Treatment					
	NPS	.0353***	.0483***	.0597***	.0158
		(.0133)	(.0152)	(.0216)	(.0236)
	DAP	-.23	-.425**	-.739***	-.435
		(.197)	(.211)	(.215)	(.279)
	Overall	-.145	-.152	-.222	-.32
		(.203)	(.194)	(.209)	(.287)
Second Treatment					
	NPS	.49***	.322***	.243**	.113
		(.0716)	(.0931)	(.0938)	(.105)
	DAP	-2.99***	-1.33	-.734	-.0133
		(.645)	(.838)	(.756)	(.741)
	Overall	.291	.375	.553	.273
		(.328)	(.331)	(.36)	(.457)
AgSS					
First Treatment					
	NPS	.116***	.124***	.113***	.0505*
		(.013)	(.0158)	(.0207)	(.0259)
	DAP	-.568***	-.58***	-.114	.423**
		(.109)	(.12)	(.145)	(.183)
	Overall	.0631	.162	.494***	.44**
		(.0982)	(.107)	(.126)	(.173)
Second Treatment					
	NPS	.135***	.142***	.16***	.135***
		(.0145)	(.0169)	(.0203)	(.0277)
	DAP	.162	.208*	.0507	.0965
		(.126)	(.119)	(.132)	(.176)
	Overall	.77***	.885***	.876***	.917***
		(.0907)	(.0782)	(.0842)	(.107)

Note: Logit Models are used except for the results on NPS, for which we used Linear Probability Models; Woreda fixed effects and other controls are included in all models; Standard errors in parentheses, clustered by EA; * p<0.10, ** p<0.05, *** p<0.01.

1.7.2 Continuous Diff-in-Diff (CDD) estimator

In the previous subsection, we showed results at different cutoff points of treatment. We can also check results using a continuous DID estimator, which doesn't require a cutoff point to be determined. Equation 18 below presents the model to be estimated. Where, α_i and α_t are unit and time period fixed effects, and the interaction of a dummy for the post-treatment period ($Post_t$) with a variable that measures unit $i's$ dose or treatment intensity, Di:

$$Y_{it} = \alpha_i + \alpha_t + \beta^{twfe} D_i Post_t + \varepsilon_{it} \quad \text{(Eq 18)}$$

Callaway et al. (2021) show that this TWFE specification faces similar problems like the case for binary treatment variable, that β^{twfe} will be biased if there are variations in treatment timing and treatment effect heterogeneity across groups and over time as we discussed in the empirical strategy section.

Therefore, we estimate equation 19 below which is the basic DID model, separately for the first and second treatment and excluding already treated households for the second treatment effect estimation, following similar strategy as we do for the binary treatment case.

$$Y_{it} = \alpha + \beta_1 D_i + \beta_2 POST_t + \beta_3^{CDD} D_i Post_t + \gamma \boldsymbol{X}_{it} + \mu_w + \varepsilon_{it} \quad \text{(Eq 19)}$$

where D_i is travel time in hours from the nearest blending facility, and $POST_t$ is an indicator variable for observations post-treatment, $\boldsymbol{X}_{it}$ is a set of the covariates which were also included in the main results, and ε_{it} is iid random variable. We have included woreda fixed effects (μ_w) to control for any woreda level differences, and β_3^{CDD} is the coefficient of interest.

Table 1.10 below presents the results for NPS, DAP and overall fertilizer adoption, all together. The table presents β_3^{CDD} from equation 19. Each column presents separate equations for NPS, DAP and Overall fertilizer adoption and the rows show the coefficient from different datasets: FtF, LSMS and AgSS. The coefficients on NPS fertilizer adoption are negative and significant, meaning households further from the blending facilities are less likely to use NPS fertilizer. The coefficients on DAP fertilizer adoption are all positive, except on the AgSS dataset for the second treatment, meaning households further from the blending facilities are more likely to use DAP. These results confirm our findings that bending facilities play a significant role in the shift from DAP to NPS by increasing adoption of NPS and decreasing adoption of DAP more in the areas around them. The coefficients on overall fertilizer adoption are insignificant, except on the results from the AgSS dataset which shows households further from the blending facilities were more likely to adopt fertilizer in general during the first treatment while this was revered on the second treatment. This also confirms our finding that we do not have strong evidence suggesting FBFs improve overall fertilizer adoption.

Table 1.10: Effects of Fertilizer Blending Facilities on Fertilizer Adoption: Continuous DID model

	Datasets	NPS	DAP	Overall fertilizer
First Treatment				
	FtF	-.00744***	.0569	.0163
		(.00263)	(.0363)	(.0381)
	LSMS	-.0222***	.121***	-.0157
		(.00529)	(.046)	(.0316)
	AgSS	-.0229***	.149***	.0361**
		(.00192)	(.019)	(.0174)
Second Treatment				
	FtF	-.0827***	.459**	-.0553
		(.0195)	(.198)	(.09)
	AgSS	-.0461***	-.0673*	-.334***
		(.00499)	(.0355)	(.0276)

Note: Logit Models are used except for the results on NPS, for which we used Linear Probability Models; Woreda fixed effects and other controls are included in all models; Standard errors in parentheses, clustered by EA; * p<0.10, ** p<0.05, *** p<0.01.

1.8 Conclusions

We assessed large scale Ethiopian government investments in promoting fertilizer use. We relied on the temporal and spatial variation of the interventions to analyze the causal impact of these investments. We found that establishment of FBFs facilitated the government's policy driven shift from DAP to NPS fertilizer types, but it didn't improve overall fertilizer adoption. Further investigations, with detailed data on the operations of the blending facilities, will probably provide justifications to why the FBFs did not lead to improvements in overall fertilizer adoption. We also found that demand side intervention through the establishment of DPs have improved adoption of NPS fertilizer, did not led to dis-adoption of DAP immediately, and resulted a significant increase in overall fertilizer adoption. We further investigated the association between broader agricultural extension services and fertilizer adoption, and we found positive and significant associations.

Our results have important policy implications related to fertilizer market access. Households with better access to the NPS fertilizer distributing primary cooperatives were found to be more likely to switch to NPS from DAP. Transport and communication investments that improve households' access to fertilizer markets reduce fixed and variable costs of accessing fertilizer markets, and thereby facilitate fertilizer adoption. Our results and the policy implications are consistent with what others found (Minten, Koro, et al., 2013; Porteous, 2020a; Suri, 2011).

Our findings that DPs improved adoption of fertilizer without leading to immediate dis-adoption of DAP suggests that households' lesson from the DPs was not specific to NPS fertilizer. Most likely, households understood the general message that fertilizer improves productivity, but they were not focused on the types of fertilizers being demonstrated. This finding has important implication for agricultural

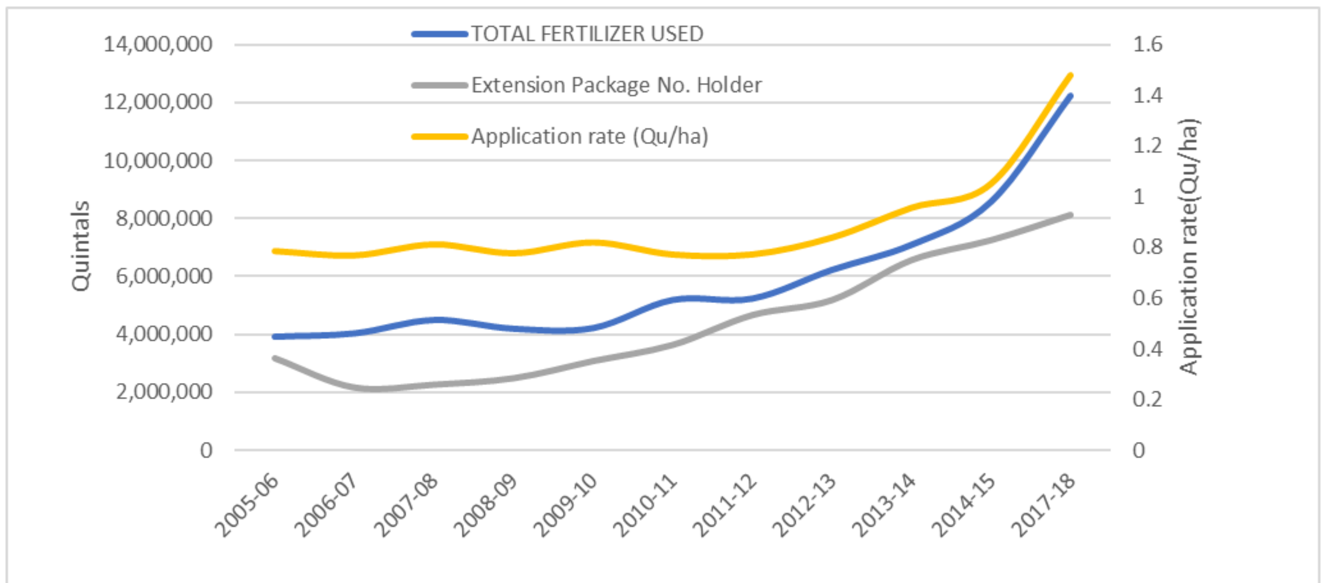
extension services in low-income countries. Rural households are not sophisticated enough to understand complicated messages. Therefore, advises have to be tailored to their capabilities and be communicated in the simplest way possible. Our results also show that households choose their own take-away lesson from extension agents' advice, and a specific advice can go a long way beyond transferring the intended lesson. All in all, our findings suggest the potential of achieving significant fertilizer adoption growth by investing in agricultural extension services.

Finally, the policy implications highlighted above do not stand alone. The importance of investments in transport and communication infrastructure and the potential of achieving huge fertilizer adoption growth through agricultural extension services have interaction effects. The effect of agricultural extension will be multiplied by investments in transport and communication infrastructure, e.g., roads, and vice versa.

In general, our study attempted to answer very complicated, data -demanding research questions. We understand that technology adoption, specifically fertilizer adoption, is a complex household decision that many researchers have analyzed. But the question why farmers in sub–Saharan Africa do not adopt fertilizer and apply at the recommended application rate remains to be largely unexplained fact. We believe we have highlighted some possible explanations relating to the type of the fertilizer itself, fixed costs households incur to access fertilizer, and the role of agricultural extension services. Further research, especially using randomized control trials are required to better understand the issue.

Appendix A

Figures



Source: authors calculations based on AgSS data, CSA.

Figure A.1: Fertilizer use application rate and number of extension package users

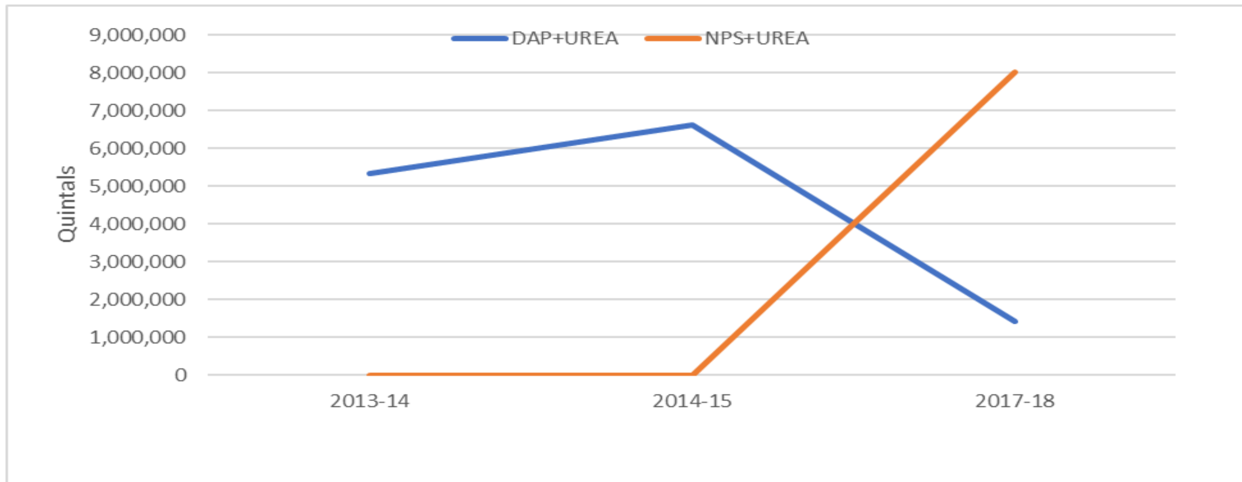


Figure A.2: DAP Vs NPS fertilizer uses (CSA Data)

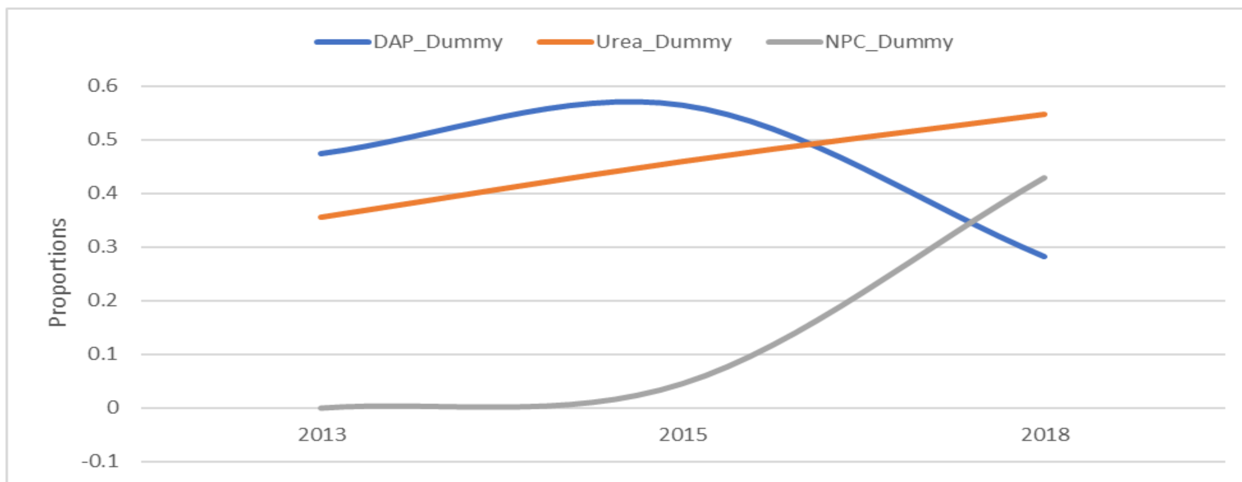


Figure A.3: Adoption rate by fertilizer type (IFPRI-FtF Data)

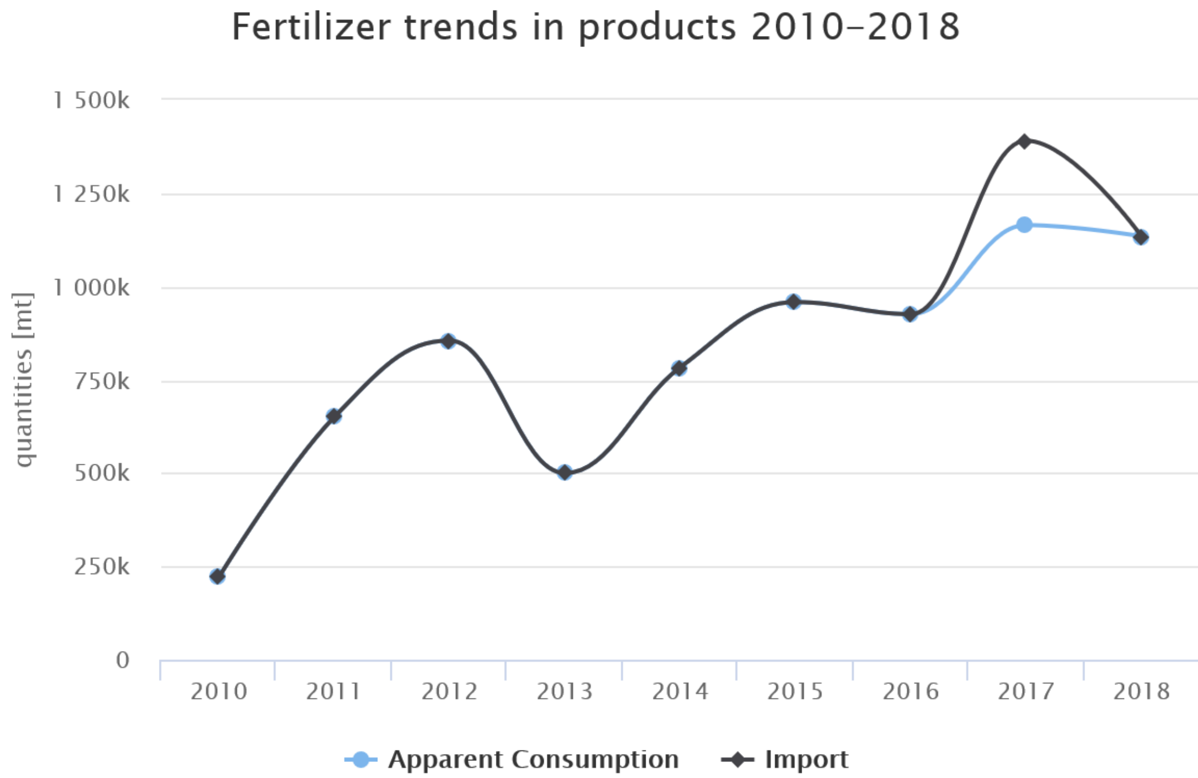


Figure A.4: Fertilizer imports and apparent consumption in Ethiopia from 2010-2018.

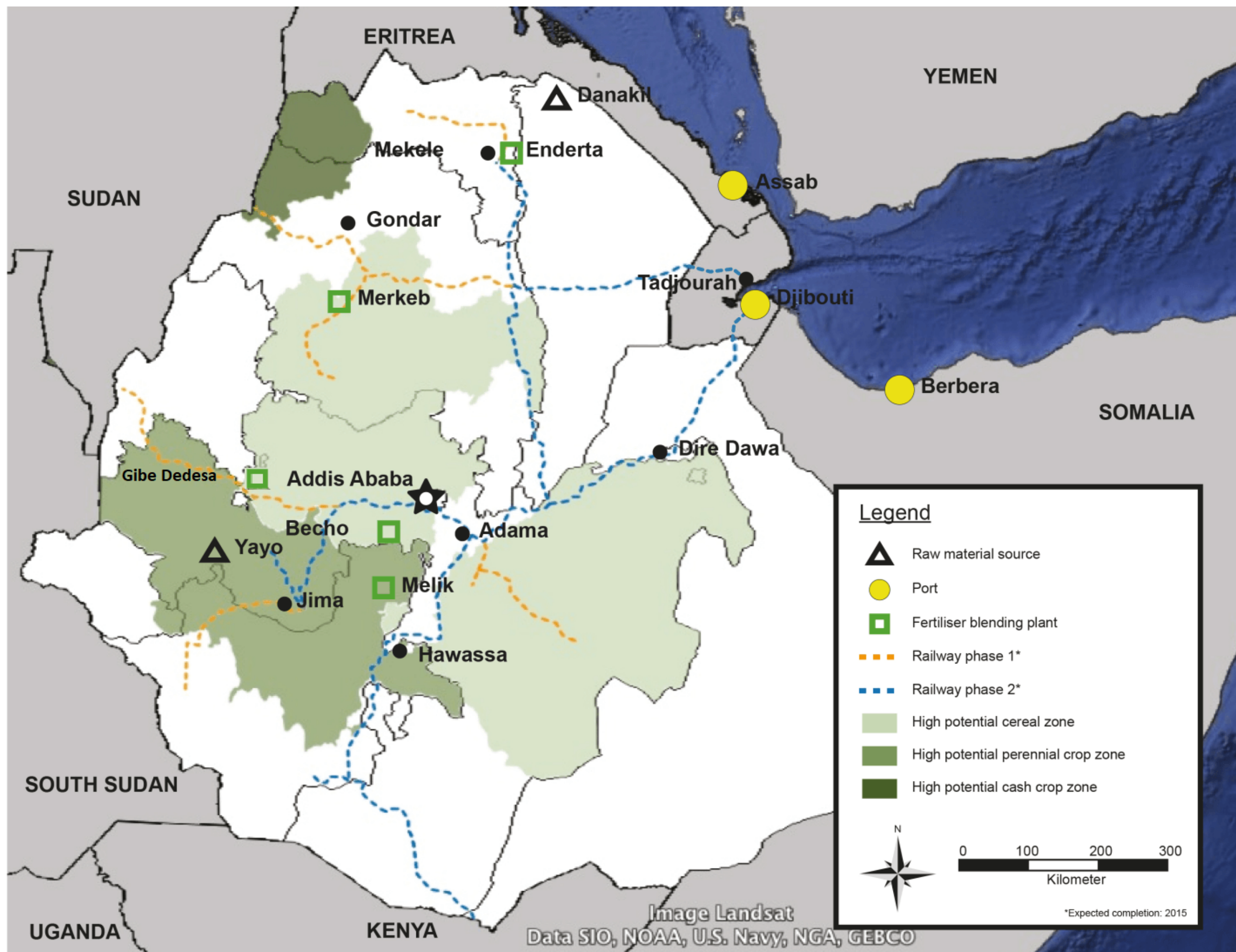


Figure A.5: Locations of five fertilizer blending plants in Ethiopia.

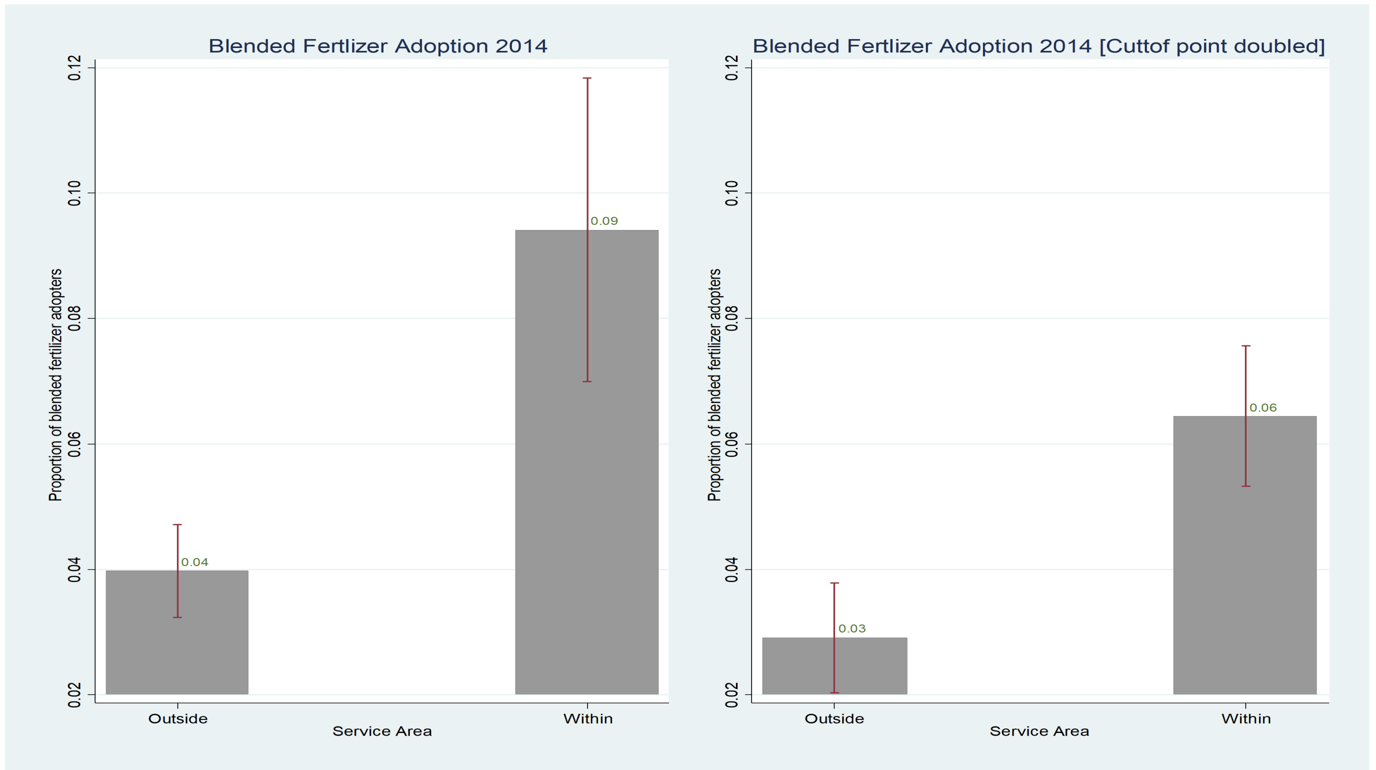


Figure A.6: Blended fertilizer adoption 2014

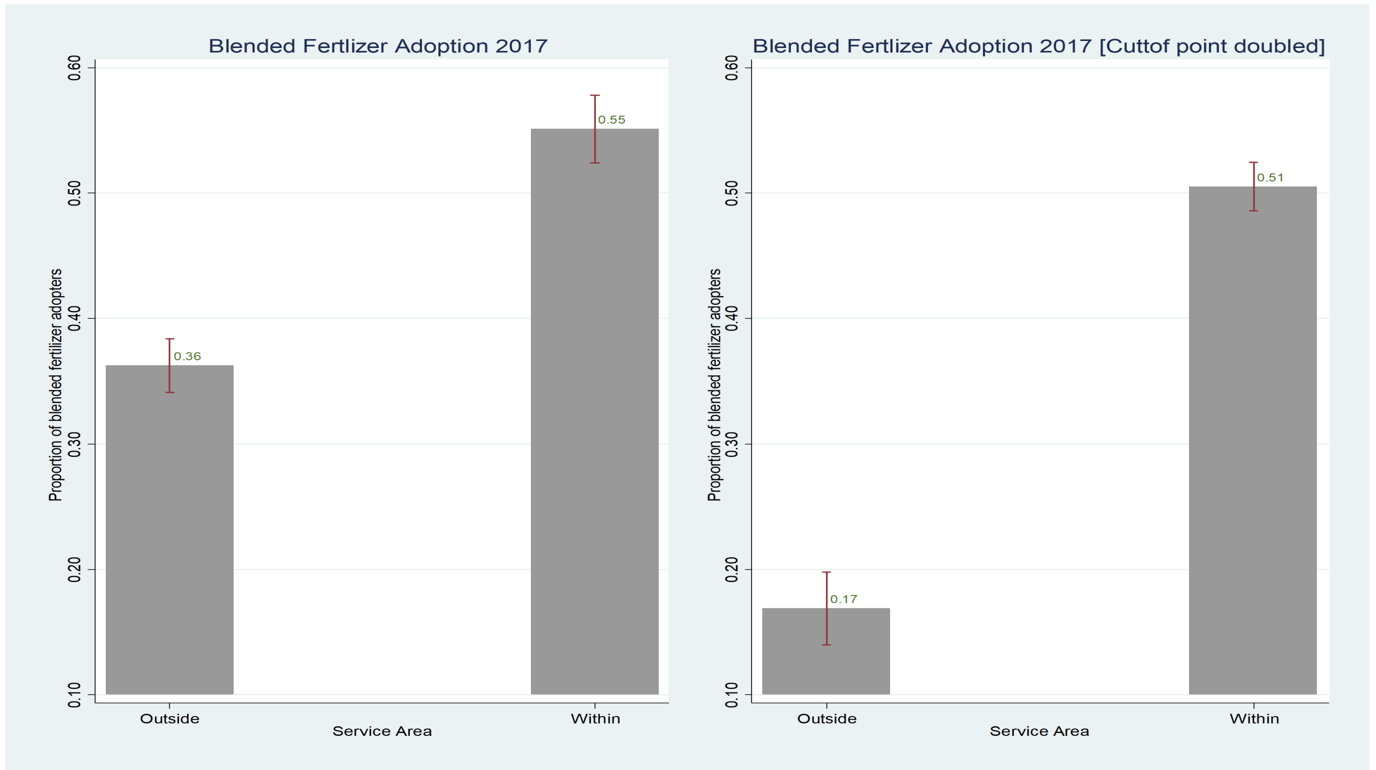
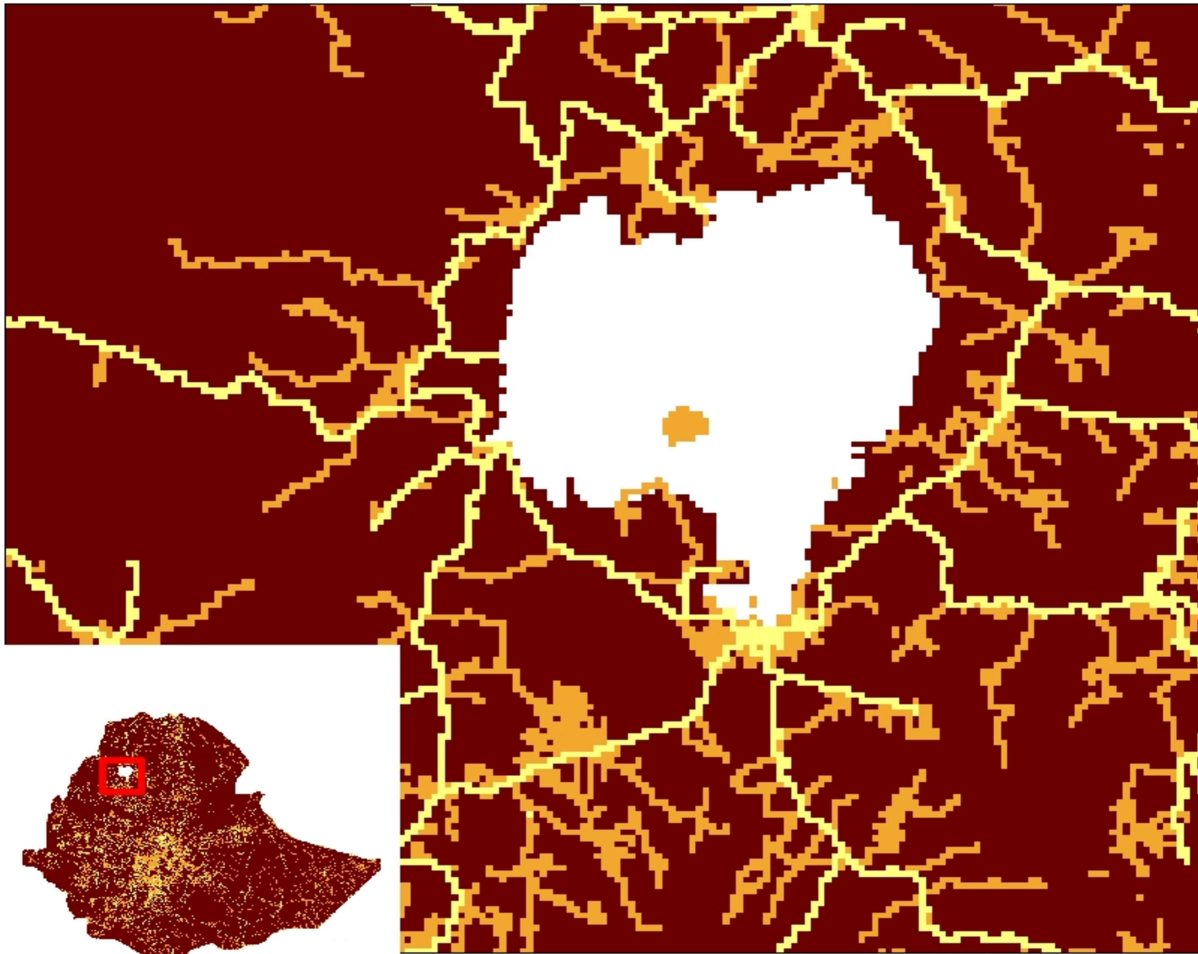
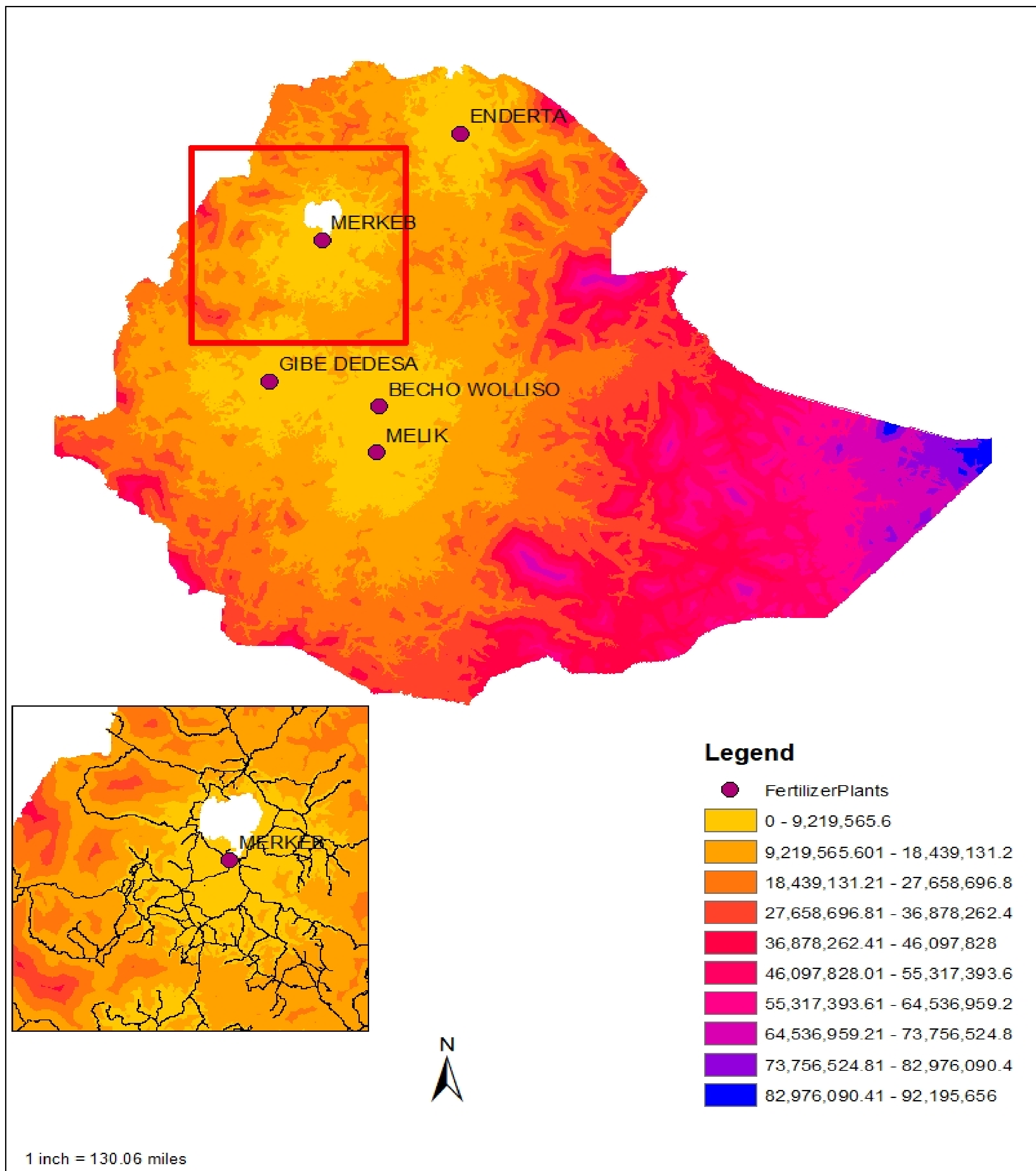


Figure A.7: Blended Fertilizer adoption 2017



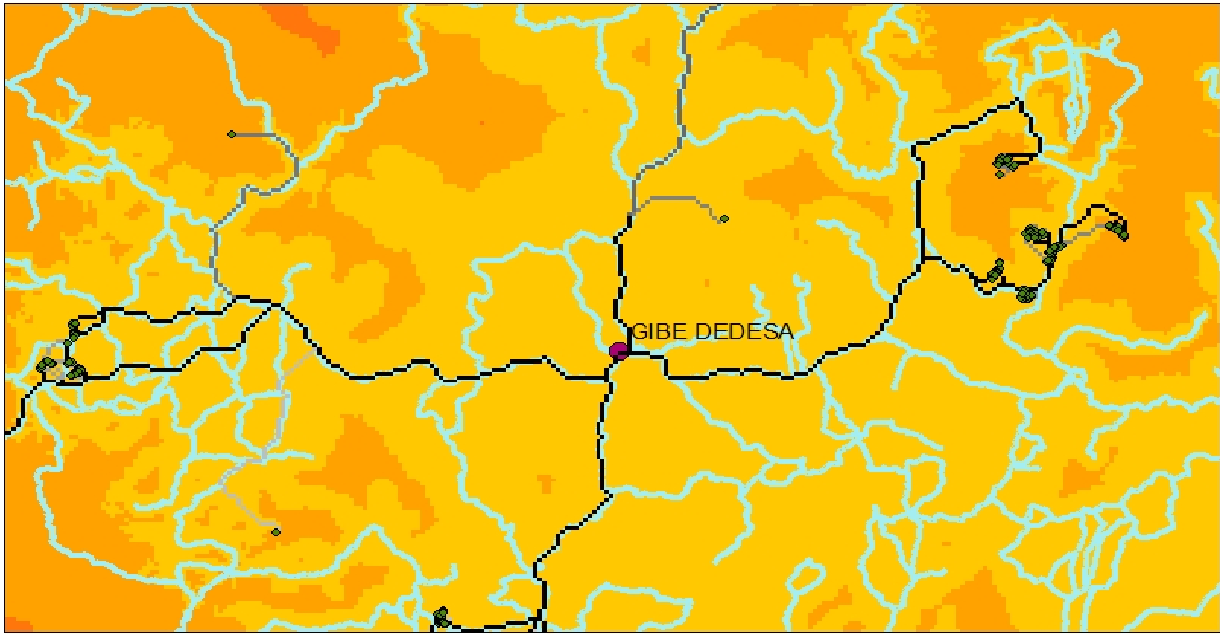
Note: this figure shows the speed assigned for each 1 km by 1km grid cell throughout the country. The darker the cells the slower is the speed. A walking speed of 5km/hour, a driving speed of 35 km/hour on minor roads and 70 km/hour on major roads is assigned. The legend shows the speed it takes to cross a 1 km by 1 km grid cell. Majority of the cells are darker, meaning have only a walking speed assigned, as roads constitute smaller share of the land. The brighter cells are cells with major roads and therefore have faster speed assigned to them.

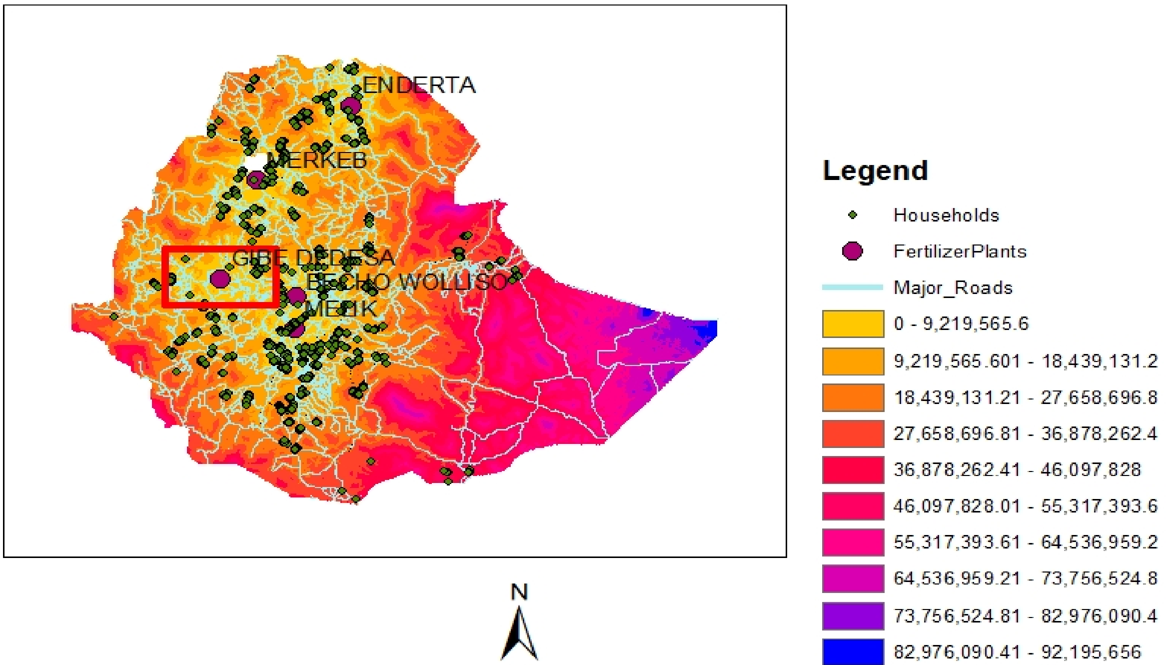
Figure A.8: Cost Surface: the darker the cells the slower the speed



Note: this figure shows the time it takes through the fastest rout from any 1km by 1km grid cell in the country to the nearest blending facility. The shades represent areas which take different ranges of travel time. Since travel time is a function of the speed the shades are aligned with availability of roads. The area around Merkeb blending facility is zoomed in, as a sample, to shows how the shades are aligned with availability of roads (represented by black lines).

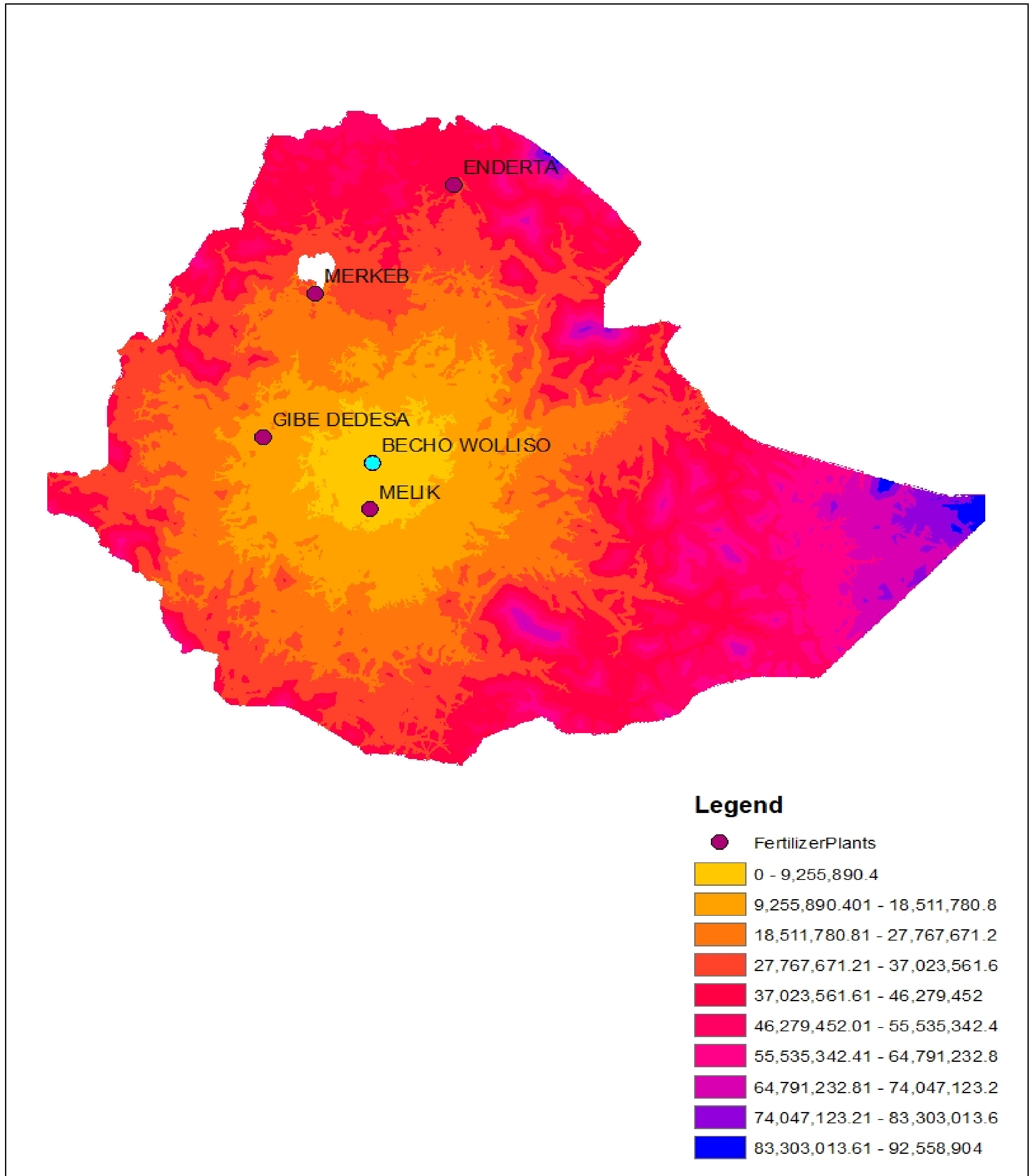
Figure A.9: Cost distance map after all blending facilities are established (2016): the time it takes to the nearest blending facility





Note: this figure shows what is presented in figure 9 with the location of each household surveyed in the FtF data. It also shows the fastest rout assigned from each household to the nearest blending facility. The Gibe Dedesa blending facility surrounding is zoomed in to show how the fastest rout is aligned with availability of roads. Roads are represented with brighter lines while the fastest routs for the sample households are represented by darker lines.

Figure A.10: Cost Path: the least cost path from households to blending facilities.



Note: this figure shows the time it takes through the fastest rout from any 1km by 1km grid cell in the country to the only blending facility in 2014: Becho Weliso blending facility. The shades represent areas which take different ranges of travel time. Since travel time is a function of the speed the shades are aligned with availability of roads.

Figure A.11: Cost distance map in 2014: the time it takes to the nearest blending facility

Appendix B

Tables

Table B.1: Effects of Fertilizer Blending Facilities on NPS Fertilizer Adoption: FD Models

	First Treatment*			Second Treatment**
	FtF	LSMS [Waves 1 & 3]	LSMS [Waves 2 & 3]	FtF
Treat_Post	.036***	.0825**	.063*	.555***
	(.0133)	(.0396)	(.0343)	(.0571)
Land size [in ha]	-.0000482***	.000143	-.0002	-.000298***
	(.0000161)	(.000339)	(.000391)	(.0000449)
Gender [Female=1]	-.0154	-.015	.0232	-.0285
	(.0141)	(.0221)	(.0178)	(.0273)
Literacy [Can read and write=1]	.0184**	-.0235	-.0127	-.000804
	(.00876)	(.0188)	(.0155)	(.0179)
Constant	.028***	.0805***	.0768***	.0508***
	(.00576)	(.0161)	(.0151)	(.0113)
Observations	3256	2910	3339	2818

Note: FD; First Difference, Standard errors in parentheses, clustered by EA, * p<0.10, ** p<0.05, *** p<0.01*FtF observations are from year 2012/13 and 2014/15, Wave 1 LSMS observations are from 2011/12, Wave 2 LSMS observations are from 2013/14, Wave 3 LSMS observations are from 2015/16. ** FtF observations are from 2014/15 and 2017/18.

Table B.2: Effects of Fertilizer Blending Facilities on DAP Fertilizer Adoption: FD Models

	First Treatment*			Second Treatment**
	FtF	LSMS [Waves 1 & 3]	LSMS [Waves 2 & 3]	FtF
Treat_Post	-.0374	-.0726*	-.0381	-.574***
	(.028)	(.0437)	(.0372)	(.0609)
Land size [in ha]	.000157	.00188**	.00186**	.000399***
	(.000191)	(.000812)	(.000788)	(.00012)
Gender [Female=1]	-.0565*	.0132	-.00945	-.00866
	(.0332)	(.0411)	(.0407)	(.0407)
Literacy	.0173	.0225	-.0242	.0427
[Can read and write=1]	(.0212)	(.0251)	(.024)	(.0315)
Constant	.053***	-.0167	-.0715***	.0197
	(.0185)	(.0214)	(.0174)	(.0143)
Observations	3256	2910	3339	2818

Note: FD; First Difference, Standard errors in parentheses, clustered by EA, * p<0.10, ** p<0.05, *** p<0.01*FtF observations are from year 2012/13 and 2014/15, Wave 1 LSMS observations are from 2011/12, Wave 2 LSMS observations are from 2013/14, Wave 3 LSMS observations are from 2015/16. ** FtF observations are from 2014/15 and 2017/18.

Table B.3: Effects of Fertilizer Blending Facilities on Overall Fertilizer Adoption: FD Models

	First Treatment*			Second Treatment**
	FtF	LSMS [Waves 1 & 3]	LSMS [Waves 2 & 3]	FtF
Treat_Post	-.0293	.00825	.0252	-.00742
	(.0273)	(.0237)	(.0201)	(.0282)
Land size [in ha]	.0000783	.00157*	.00156*	.0000785
	(.000181)	(.000894)	(.000847)	(.000103)
Gender [Female=1]	-.0415	.012	.0417	-.0141
	(.0354)	(.0332)	(.036)	(.0366)
Literacy	.0387*	-.0195	-.0375**	.0645***
[Can read and write=1]	(.0208)	(.0212)	(.0176)	(.0242)
Constant	.0877***	.0442***	-.00588	.0719***
	(.0221)	(.0151)	(.0105)	(.0165)
Observations	3256	2910	3339	2818

Note: FD; First Difference, Standard errors in parentheses, clustered by EA, * p<0.10, ** p<0.05, *** p<0.01*FtF observations are from year 2012/13 and 2014/15, Wave 1 LSMS observations are from 2011/12, Wave 2 LSMS observations are from 2013/14, Wave 3 LSMS observations are from 2015/16. ** FtF observations are from 2014/15 and 2017/18.

Table B.4: Effects of Extension Services on Fertilizer Adoption: FD Models

	FtF	LSMS
NPS		
Visited [Yes=1]	.0112	
	(.0137)	
Under extension program [yes=1]		.0217*
		(.0111)
DAP		
Visited [Yes=1]	.0714***	
	(.0172)	
Under extension program [yes=1]		.293***
		(.0267)
OVERALL		
Visited [Yes=1]	.0624***	
	(.0129)	
Under extension program [yes=1]		.344***
		(.0264)
Observations	6510	5175

Note: FD; First Difference, Year FE are included in all models, Standard errors in parentheses, clustered by EA, EA FEs are included, * p<0.10, ** p<0.05, *** p<0.01, FtF observations are all the three rounds in 2013, 2015 and 2018 and LSMS observations are all the three waves in 2011/12, 2013/14 and 2015/16.

Table B.5: Distribution of 2013/14 NPS demonstration plots by region and crop.

		Farmers' plots				FTC* plots			
Region	Woreda/FTC	Wheat	Teff	Maize	Barley	Wheat	Teff	Maize	Barley
Tigray	17/170	3235	3370	2101	948	240	250	190	70
Amhara	91/750	3285	3415	2475	1280	222	225	225	65
Oromia	101/737	1965	2390	1349	849	115	115	120	40
SNNPR	45/390	1015	905	762	485	50	50	50	20
Total	253/2047	9500	10080	6687	3562	627	640	585	195

*FTC stands for Farmer Training Center.

Source; AFAP webpage < https://www.afap-partnership.org/volunteer-programme/success-stories/ethiopia/ >

Appendix C

Goodman – Bacon Decomposition

Figures C1, C2 and C3 present the TWFE model estimate, and the Goodman-Bacon decomposition of the effect of FBFs on NPS, DAP and Overall fertilizer adoption, respectively. The figures also show the weights attached to each estimate. In line with that, we present the statistical significance and standard errors of the coefficients in Tables C1,C2 and C3.

We present results from two of the datasets: AgSS and FtF. The LSMS dataset does not include data beyond 2015/16 and is used to estimate the effect of the first treatment only. Therefore, it does not require Goodman-Bacon decomposition. The TWFE estimates, although likely biased, reflect the main results we found that FBFs have facilitated the policy of substituting DAP by NPS but did not lead to improvements in overall fertilizer adoption.

The decomposed results show the estimates for all possible 2X2 DID estimates. Theoretically, the one which compares later treated groups with earlier treated groups is the most troublesome as it is using already treated households as controls. This is reflected in the estimates across the three outcome variables below. In most of the estimates, it is smaller than the one which compares later treated groups to the never treated households. In Table 8, the estimated coefficient comparing later treated group to already treated group has a negative sign, as opposed to the rest of the estimates. Based on the figures on all of the outcome variables, the one using already treated households as controls was weighted higher than the other estimates, indicating that the TWFE model estimates are likely biased. This is because the effect estimated using already treated groups is smaller than what it should be and is weighted more than the other estimates.

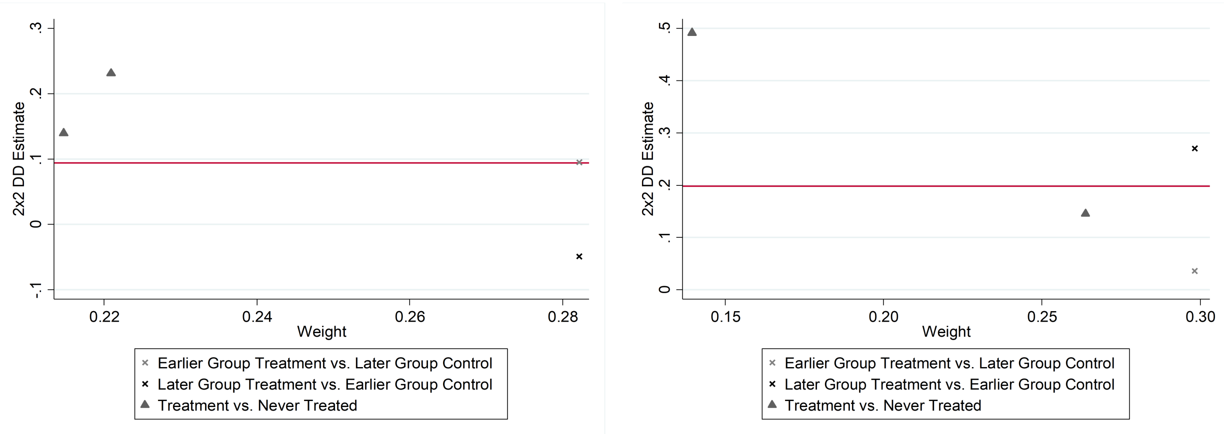


Figure C.1: Goodman-Bacon decomposition of the Effects of Fertilizer Blending Facilities on NPS Fertilizer Adoption: AGSS and FtF

Table C.1: Goodman-Bacon decomposition of the Effects of Fertilizer Blending Facilities on NPS Adoption

	TWFE	Early Treated Vs Late Treated	Early Treated Vs Never Treated	Late Treated Vs Early Treated	Late Treated Vs Never Treated
AgSS	.094***	.095***	.231***	-.0491**	.14***
	(.012)	(.0147)	(.0232)	(.0244)	(.0188)
Observations	5787	3234	3396	3234	3327
FtF	.198***	.0358***	.145***	.27***	.492***
	(.0124)	(.00876)	(.0214)	(.0212)	(.0195)
Observations	9768	8472	6837	8472	4227

Standard errors in parentheses; TWFE: two-way fixed effects; The data is from 2013, 2015 and 2018.
* p<0.10, ** p<0.05, *** p<0.01

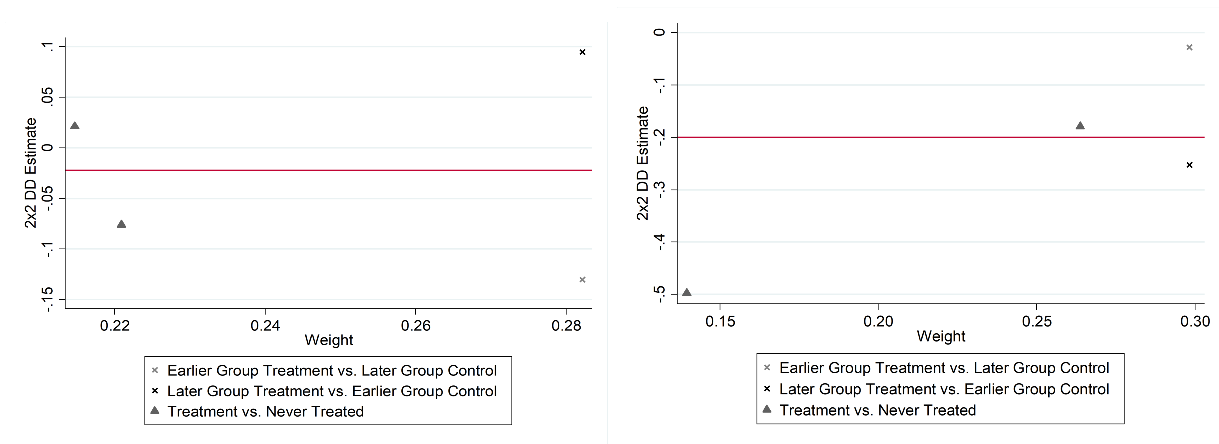


Figure C.2: Goodman-Bacon decomposition of the Effects of Fertilizer Blending Facilities on DAP Fertilizer Adoption: AGSS and FtF

Table C.2: Goodman-Bacon decomposition of the Effects of Fertilizer Blending Facilities on DAP Adoption

	TWFE	Early Treated Vs Late Treated	Early Treated Vs Never Treated	Late Treated Vs Early Treated	Late Treated Vs Never Treated
AgSS	-.0222	-.13***	-.0761***	.0949***	.0211
	(.0145)	(.0244)	(.029)	(.0259)	(.0266)
Observations	5787	3234	3396	3234	3327
FtF	-.2***	-.423***	-.179***	-.0357**	-.498***
	(.0156)	(.0195)	(.0307)	(.0144)	(.0292)
Observations					

Standard errors in parentheses; TWFE: two-way fixed effects; The data is from 2013, 2015 and 2018;
* p<0.10, ** p<0.05, *** p<0.01

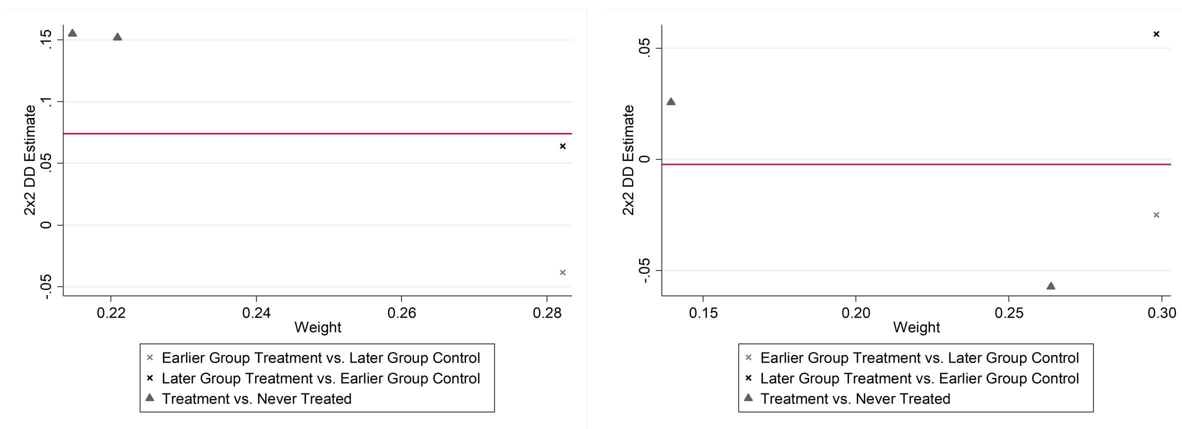


Figure C.3: Goodman-Bacon decomposition of the Effects of Fertilizer Blending Facilities on Overall Fertilizer Adoption: AGSS and FtF

Table C.3: Goodman-Bacon decomposition of the Effects of Fertilizer Blending Facilities on Overall Fertilizer Adoption

	TWFE	Early Treated Vs Late Treated	Early Treated Vs Never Treated	Late Treated Vs Early Treated	Late Treated Vs Never Treated
AgSS	.074***	-.0384*	.152***	.0638***	.155***
	(.0109)	(.0232)	(.0267)	(.024)	(.0283)
Observations	5787	3234	3396	3234	3327
FtF	-.00209	.0322*	-.0572**	.0672***	.0257
	(.0129)	(.0174)	(.0282)	(.0132)	(.0289)
Observations					

Standard errors in parentheses; TWFE: two-way fixed effects; The data is from 2013, 2015 and 2018;
* p<0.10, ** p<0.05, *** p<0.01

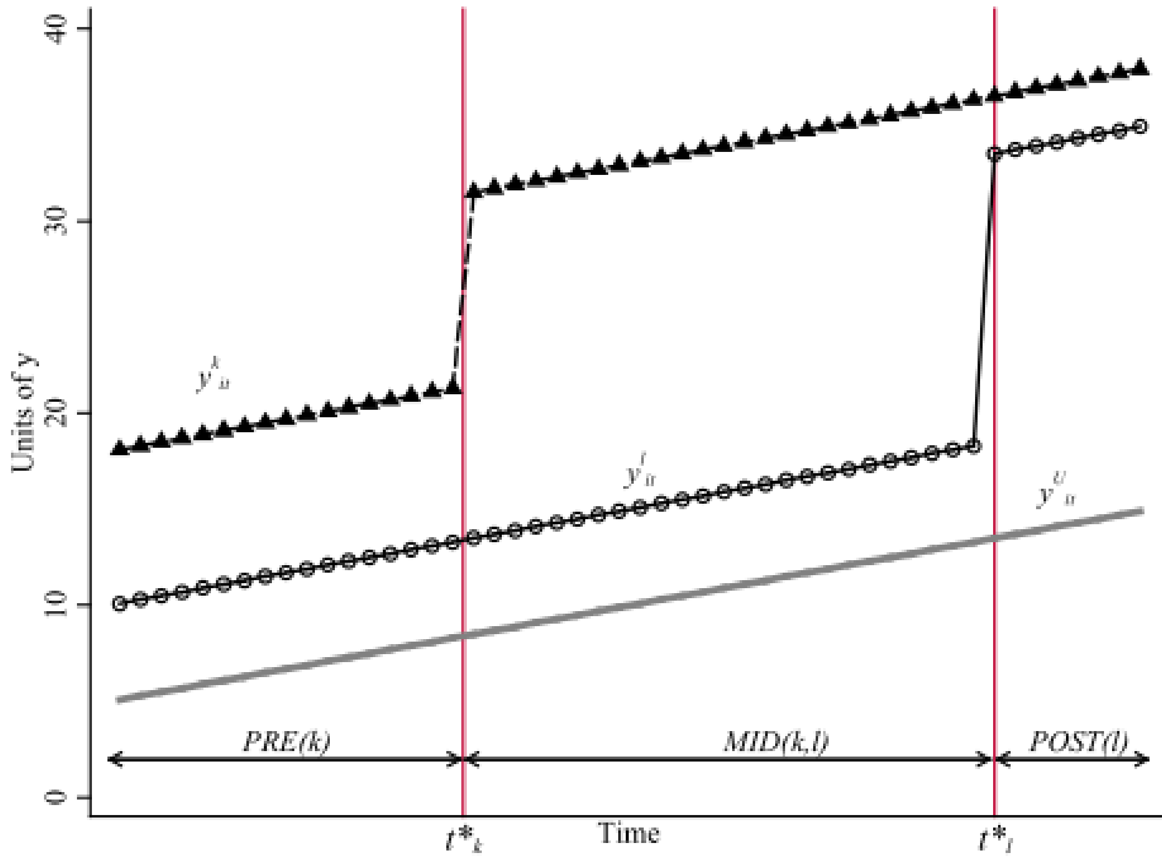


Figure C.4: Difference-in-Differences with Variation in Treatment Timing: Three Groups

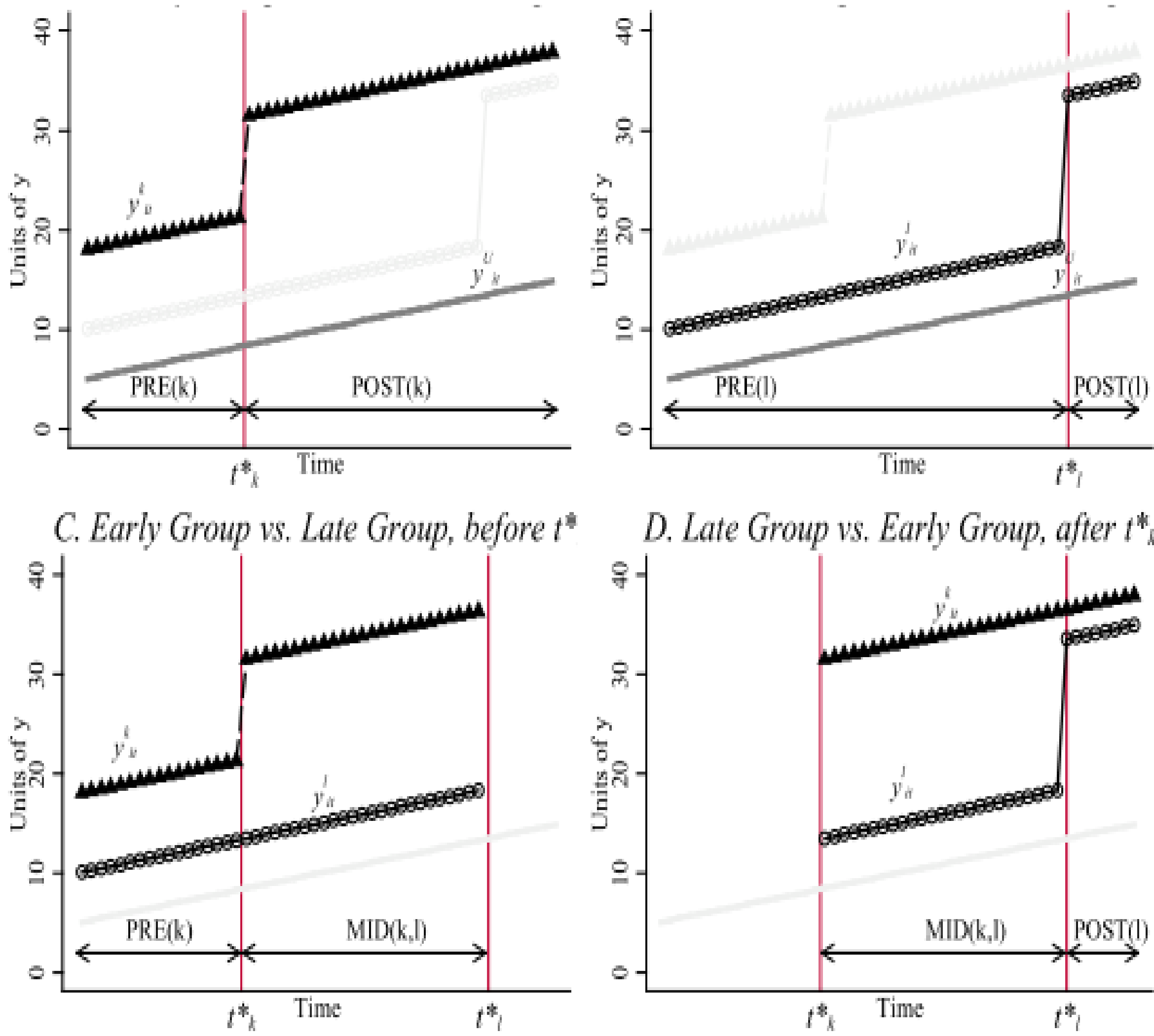


Figure C.5: The Four Simple (2x2) Difference-in-Differences Estimates from the Three Group Case

Bibliography

AFAP. (2012). Ethiopia fertilizer assessment. *African Fertilizer and Agribusiness Partnership*.

AFAP. (2015). 2015 international meeting of association française de commercialisation et de mélanges d'engrais. *African Fertilizer and Agribusiness Partnership*, October 21 –23, Reims, France.

Aggarwal, S., Giera, B., Jeong, D., Robinson, J., & Spearot, A. (2018). Market access, trade costs, and technology adoption: Evidence from northern tanzania. *NBER Working Paper*, (252253).

Alene, A. D., Manyong, V. M., Omanya, G., Mignouna, H. D., Bokanga, M., & Odhiambo, G. (2008). Smallholder market participation under transactions costs: Maize supply and fertilizer demand in kenya. *Food Policy*, *33*(4), 318–328.

Athey, S., & Imbens, G. W. (2022). Design-based analysis in difference-in-differences settings with staggered adoption. *Journal of Econometrics*, *226*(1), 62–79.

Bachewe, F. N., Berhane, G., Minten, B., & Taffesse, A. S. (2015). Agricultural growth in ethiopia (2004-2014): Evidence and drivers. *ESSP Program Working Paper*, (81).

Baker, A. C., Larcker, D. F., & Wang, C. C. (2021). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics*, *119*(1), 249–275.

Berhane, G., Ragasa, C., Abate, G. T., & Assefa, T. W. (2018). The state of agricultural extension services in ethiopia and their contribution to agricultural productivity. *ESSP Working Paper*, (118).

Blum, M., Cofini, F., & Suliaman, V. (2020). *Agricultural extension in transition worldwide: Policies and strategies for reform*. FAO: Rome, Italy.

Borusyak, K., Jaravel, X., & Spiess, J. (2022). Revisiting event study designs: Robust and efficient estimation. *arXiv:2108.12419v2*.

Callaway, B., Goodman-Bacon, A., & Sant'Anna, P. H. (2021). Difference-in-differences with a continuous treatment. *arXiv:2107.02637v2*.

Callaway, B., & Sant'Anna, P. H. C. (2020). Difference-in-differences with multiple time periods. *arXiv:1803.09015v4*.

Chaisemartin, D. B. C., & Haultfœuille, X. D. (2020). Two-way fixed effects estimators with heterogeneous treatment effects †. *American Economic Review*, *110*(17), 2964–2996.

Davis, K., Babu, S. C., & Ragasa, C. (2020). *Agricultural extension: Global status and performance in selected countries.* (K. Davis, S. C. Babu, & C. Ragasa, Eds.; Vol. 38). International Food Policy Research Institute.

Duflo, E., Kremer, M., & Robinson, J. (2011). Nudging farmers to use fertilizer: Theory and experimental evidence from kenya. *American Economic Review*, *101*(6), 2350–90.

Goodman-bacon, A. (2018). Difference-in-differences with variation in treatment timing. *NBER Working Paper*, (25018).

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

Key, N., Sadoulet, E., & DE Janvry, A. (2000). Transactions costs and agricultural household supply response. *American Journal of Agricultural Economics*, *82*(2), 245–259.

Minten, B., Koru, B., & Stifel, D. (2013). The last mile(s) in modern input distribution: Pricing, profitability, and adoption. *Agricultural Economics*, *44*(6), 629–646.

Niehans, J. (1987). Transaction costs. In *The new palgrave dictionary of economics.* Palgrave Macmillan, London.

Porteous, O. (2019). High trade costs and their consequences: An estimated dynamic model of african agricultural storage and trade. *American Economic Journal: Applied Economics*, *11*(4), 327–366.

Porteous, O. (2020a). Trade and agricultural technology adoption : Evidence from africa. *Journal of Development Economics*, *144*(February 2019), 102440.

Porteous, O. (2020b). Trade and agricultural technology adoption: Evidence from africa. *Journal of Development Economics*, *144*(February 2019), 102440.

Pozzi, F., & Robinson, T. (2008). Accessibility mapping in the horn of africa : Applications for livestock policy. *IGAD Livestock Policy Initiative*, (11).

Rashid, S., Tefera, N., Minot, N., & Ayele, G. (2013). Fertilizer in ethiopia: An assessment of policies, value chain, and profitability. *IFPRI Discussion Paper*, (01304).

Sheahan, M., & Barrett, C. B. (2017). Ten striking facts about agricultural input use in sub-saharan africa. *Food Policy*, *67*, 12–25.

Singh, I., Squire, L., & Strauss, J. (1986). *Agricultural household models: Extensions, applications and policy.* (I. Singh, L. Squire, & J. Strauss, Eds.). A World Bank Publication. Baltimore: Johns Hopkins University Press.

Spielman, D. j., Mekonnen, D. K., & Alemu, D. (2013). Seed, fertilizer, and agricultural extension in ethiopia. In P. A. Dorosh & S. Rashid (Eds.), *Food and agriculture in ethiopia. progress and policy challenges* (1st, pp. 84–122). University of Pennsylvania Press (for International Food Policy Research Institute).

Sun, L., & Abraham, S. (2020). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, (40), 1–25.

Suri, T. (2011). Selection and comparative advantage in technology adoption. *Econometrica*, *79*(1), 159–209.

Wooldridge, J. M. (2012). *Introductory econometrics: A modern approach*. Michigan State University.

Wooldridge, J. M. (2021). Two-way fixed effects , the two-way mundlak regression , and difference-in-differences estimators. *Preprint*, (August).

Wossen, T., Gatiso, T. T., & Kassie, M. (2019). Estimating returns to fertilizer adoption with unobserved heterogeneity: Evidence from ethiopia. *Food and Energy Security*, *8*(2), 1–9.

Chapter 2

Engaging men in household chores and child care through mobile phone-delivered Behavior Change Communications (BCC)

2.1 Introduction

Many development programs seek to achieve lasting behavior change among their beneficiaries through program interventions. Development programs typically provide women and their male partners with livelihood and gender training, with behavior change often a core element of the programs' theories of change. As programs in developing countries increasingly target women, the issue of engaging men in the process remains less addressed. Engaging men in behavioral change around gender norms is challenging. Socially constructed behavioral constraints, like gender norms, are not easy to change and require multifaceted interventions. Often, interventions seeking to address gender norms as a means to empower women take the form of a single-gender training event (E. Bulte et al., 2016; E. H. Bulte et al., 2018; Mckenzie et al., 2017). In this study, we experimentally evaluate a behavioral intervention that involves repeated post-training reinforcement of gender norms training contents, delivered to men through phone calls. With increased penetration of mobile phones in the developing world (GSMA, 2018), coupled with decreased costs of ownership and operation (Stork et al., 2017), mobile-based interventions offer a promising, yet understudied, an alternative to interventions relying on costly in-person visits. By delivering behavioral reinforcement messages over the phone, we hope to develop an inexpensive and scalable solution to encourage and sustain behavioral change in developing countries.

To evaluate the impacts of this behavioral intervention, we conduct a randomized control trial (RCT) using respondents from Ethiopia who participated in a rural program seeking to raise participants above the poverty line by enhancing their livelihoods activities. We seek to understand whether cell phones can be used to deepen program impacts by reinforcing behavior change communications. We randomly

assign male household heads either to the treatment group, which receives phone-based reinforcement of training messages designed to encourage men to participate in household chores and child care, or the placebo group, which receives similar phone calls addressing an unrelated placebo topic. The treatment spans three months, with six phone calls placed on a biweekly schedule. During each call, a well-trained enumerator following a carefully designed script discusses gender norms with the male respondent, who is asked to commit to small weekly tasks designed to encourage increased male participation in household chores and child care. For the control group, an enumerator places calls at the same frequency but instead discusses a placebo topic – the man's food consumption in the last 24 hours. We collect endline data from the female spouses of these male household heads two weeks following the final reinforcement phone call. We ask the female respondent detailed information about participation of each household member, including the male spouse, in household chores and child care.

We hypothesize an increase in male's participation in household chores and child care at the end of the intervention period. We measure male's participation in eight household tasks – child care, collecting fuelwood, fetching water, cooking, washing clothes, processing grains, house cleaning, and dish washing. We estimate the treatment effect using analysis of covariance, which controls for the baseline value of the outcome variable.

Our results indicate that men's participation increased, especially in public chores, despite no treatment effect on women's participation due to the treatment. Disaggregating results by types of chores shows men's increased participation in public chores is largely driven by their increased participation in fuelwood collection and to some extent by their increased participation in processing grains. Heterogeneity analysis by age reveals that younger men increased their participation in private chores instead while older men increased their participation in public chores. Younger men's participation in private chores was largely driven by their participation in child care. Treatment effects heterogeneity by participation in chores at baseline shows that the increase in men's participation in fuelwood comes from those who have not been conducting fuelwood at baseline. Our analysis on the woman's level of satisfaction in the man's participation in chores shows that women level of satisfaction significantly increased.

We also found that the treatment led to changes in participation of others, other than the targeted man. Children's participation in chores, specifically daughters' participation in child care and sons' participation in fuelwood collection increased due to the treatment. We observe heterogeneity in these results based on age of the targeted man. Sons of older men increased their participation in chores, especially on fuelwood, while sons of younger men decreased their participation in many types of chores, especially in public chores.

We contribute to the broader Behavior Change Communication (BCC) literature. Psychological interventions such as message framing (Isaac et al., 2016; Madajewicz et al., 2007), making treated individuals to verbally pledge to change behavior or treat them with a commitment to behavior change device (Avdeenko et al., 2019; Dupas, 2011; Luoto et al., 2014) and labeling a person as someone whose behavior is already commendable (Shah et al., 2016) have been found important. Moreover, it has been highlighted that credible information alone brings behavior change (Dupas, 2011; Fitzsimons et al., 2016; Madajewicz

et al., 2007). We add to this literature by providing evidence if repeatedly reinforcing training content, which was already delivered on in-person group training, brings behavior change.

We also contribute to the fast-growing literature on the use of mobile phones in behavior change research. Interventions that use mobile phones usually rely on text messaging. For example, Zurovac et al. (2011) studied the effect of mobile phone text-message reminders on Kenyan health workers' adherence to malaria treatment guidelines, Rodríguez & Saavedra (2019) studied the effect of youth savings reminders, Kerbaj et al. (2017) studied the effect of text message reminders on hand hygiene compliance in health care workers, and Östervall (2017) studied the effect of reminders on antibiotics prescriptions. Text messages limit the content of communication and are well suited for episodic behaviors (such as attendance for immunization, adherence to taking medication...etc.). Complex, socially mediated behaviors rooted in gender norms, however, require deeper and multi-faceted BCC interventions (Mildon et al., 2019). Our intervention administers a two-way phone-based conversation between a well-trained enumerator and respondents. Thus, our study introduces evidence on how mobile phones can be used to alter relatively complex behaviors.

Furthermore, we contribute by providing evidence on Behavior Change Communication (BCC) approach that combines group and individual level interventions. Group level interventions include public campaigns, mass media communications, community mobilization and group trainings (e.g., Gender Group Dialog or GGD) (Chase et al., 2012; Pettifor et al., 2015; Pulerwitz et al., 2015). Individual level interventions target an individual or a household and usually happen through mobile phones, probably due to the costly nature of individual in-person interventions, in the form of short message services (SMS) (Kerbaj et al., 2017; Östervall, 2017; Rodríguez et al., 2019; Zurovac et al., 2011). In this study, we have combined the two approaches by reinforcing gender norms training contents, that have been delivered through in-person group training, to men individually through phone calls.

The rest of the manuscript is organized as follows. We start with a brief discussion of the population from which we draw our sample followed by details on experimental design and randomization in Section 3. Section 4 presents hypotheses and estimation strategy. In Section 6, we present results. We discuss and conclude in Section 7.

2.2 Sample

This study is nested in a research project called "Her time," focused on using cell phones to measure women's time use in a population of poor women in Ethiopia who are participating in multifaceted graduation-from-poverty programs. We focus on the North Wollo zone, located in the northern part of the Amhara region, because this is where both of the NGOs with which we are partnering overlap in their implementation. The two programs are Care Ethiopia's Livelihoods for Resilience (LfR) and World Vision Ethiopia's Strengthen PSNP4 Institutions and Resilience (SPIR). Within North Wollo, both LfR and SPIR programs are implemented in partnership with an Ethiopian NGO, the Organization for Rehabilitation and Development in Amhara (ORDA).

Both LfR and SPIR facilitate the formation of Village Economic and Social Associations (VESA) groups, which then become the main infrastructure through which program interventions are delivered to beneficiaries. In order to be eligible for the programs, members must qualify for and participate in Ethiopia's Productive Safety Net Program (PSNP), a government safety net program that targets poor and vulnerable households in chronically food insecure kebeles (districts) in Ethiopia. Both LfR and SPIR deliver a multifaceted set of interventions designed to permanently raise beneficiaries above the poverty line and thus reduce their future reliance on the PSNP program.

While eligible individuals are broadly encouraged to participate in SPIR and LfR wherever they are working, they ultimately self-select into these livelihoods programs. Our sample is randomly selected from the population of eligible households who choose to participate in either of these safety net programs rather than the larger population of eligible households in the region. Using program enrollment records, we draw a random sample of VESA groups in the Meket and Wadla woredas[1] of the North Wollo zone. To avoid affecting the outcomes observed in a separate program evaluation of the SPIR livelihoods program, we exclude from our sample the kebeles in Meket and Wadla woredas that were randomly selected to be part of that program's concurrent program evaluation.

2.3 Experimental design and randomization

Following VESA level livelihoods and gender training provided to both women and their male partners, we conduct a randomized control trial (RCT) using 231 participants of the LfR and SPIR programs. We randomly assign the men either to the treatment group or the placebo group. Subjects in the treatment group receive phone-based reinforcement of training messages designed to encourage them to participate in household chores and child care and those in the placebo group receive similar phone calls addressing an unrelated placebo topic. The treatment spans three months, with six phone calls placed on a biweekly schedule. During each call, a well-trained enumerator administers a carefully designed 5 minutes script in which he discusses gender norms with the male respondent. For the control group, an enumerator places calls at the same frequency but instead discusses a placebo topic – the man's food consumption in the last 24 hours.

After the treatment is over, two weeks following the final reinforcement phone call, information on the outcome variables (i.e., male's participation in household chores and child care) are collected from the targeted women through phone calls. We collect information about men's participation in household chores from the woman to avoid desirability effect. At the endline, if we were to ask the treated male partners about their participation in chores, they will find it desirable to report a higher participation as they have been told about the importance of men's participation in chores for the duration of the intervention. This could then incorrectly bias the treatment effect upward. To minimize desirability effect, we rely on the women's responses about his participation in chores. We ask the women if the following activities were performed in service of the family yesterday: collecting fuelwood, fetching water, cooking,

[1]Woredas are tertiary administrative units similar to districts, while kebeles comprise the next level of administrative units, similar to blocks or sub-districts.

child care giving, processing grains, house cleaning and washing dishes. We then ask who performed the activities and how many minutes were spent on these activities by the person who performed the activities, including the male spouse. These questions were also asked on the baseline survey. We then evaluate whether men in the treatment group participate in more household chores and child care than men in the control group at the end of the intervention period.

The content of the scripts was largely influenced by a successful intervention conducted by World Vision, in which men conduct in-person group discussions with the objective of making them allies in the process of empowering women. We prepared the scripts in consultation with gender experts at Care Ethiopia and World Vision Ethiopia. Each of the six scripts has clearly stated objectives: the first call aims at showing men the time women spend on housework as compared to men, the second call aims at encouraging men to be involved in childcare, and the third call aims at correcting the man's thinking about some of the myths hindering men's involvement in household chores and child care. The fourth, fifth and sixth calls repeat the content of the first, second and third calls, respectively. The content of the scripts is presented in different formats: open ended questions on which the respondent reflects on his thoughts, scenarios in which he imagines himself in a situation, and statements with which he was be asked to agree or disagree. Well trained enumerators conduct the phone conversations in a very smooth natural way as if they are not using a script but making sure they followed the script content to guide their conversation. By the end of each script, the respondent is asked to commit to small weekly tasks designed to encourage his participation in household chores and child care activity that he has not been performing before. On the next call, enumerators check if he has done the assignment or not.[2].

The "Her Time" research project in which this study is nested collects data from 60 VESAs in 8 waves over the course of a year, randomly assigning VESAs to waves. In each VESA, data collection activities consist of an in-person baseline questionnaire followed by 7-12 days of phone surveys and in-person validation interviews, according to their treatment assignment in the "Her Time" study[3]. The project also conducts gender norms information (GNI) intervention, in which male partners of the women were treated with information about acceptability of men performing chores that are traditionally allocated to women[4]. The gender norms information experiment is conducted just after the baseline, its treatment is assigned orthogonally to the treatment assignment of the time use study and collects an endline data through the 7-12 days of phone surveys and in-person validation interviews. On completion of these activities, households are randomly assigned to a treatment or control group for this reinforcement study, with treatment stratified by treatment assignment of the gender norms information experiment and age of the male partner collected on the baseline. By making sure that no consecutive individuals are assigned to the treatment when respondents are sorted by age, we make the treatment and control groups balanced in terms of age.

[2]The detailed content of the scripts are available upon request.

[3]More information on the time use project interventions can be found in the pre-analysis plan here; AEA RCT Registry (socialscienceregistry.org).

[4]More information about the gender norms information intervention can be found in the pre-analysis plan of the "Her Time" study.

One possible concern in this experimental design is that the treatment (phone calls targeting specific males within a community) could influence the behavior of men in the same community who have been assigned to the control group. Randomly assigning communities to a treatment group (compared to individuals) would minimize this spillover problem, but also requires much larger sample that amplifies treatment and survey cost. After carefully considering both levels of treatment assignment, we decided to assign treatment at the individual level. In our decision, we considered the nature of the treatment, the outcome variables of interest and the direction of spillover bias. First, the treatment is complex enough that is not easily transferable to households in the control group. We administer six calls and discuss complex gender norms behavior that is very difficult for households in the control group to copy with just simply through spillovers. For example, as part of the treatment, we ask the respondent to commit to fulfilling an assignment which requires him to do one of the household chores or child caring activity by the end of each call. Provided the respondent did the assignment, we allow him to slowly grow accustomed to the activities. For spillovers to exist, households in the control group would need to continue keeping up with such activities for three months which is less likely. Second, our outcome variables consist of public and private household chores[5]. For example, collecting fuelwood is public while bathing babies is private in the sense that it can be conducted without others in the neighborhood watching. Since spillovers are expected to be minimal on private household chores and we separately analyze separately differences in the impact on public and private household chores to trace spillovers, if any. Third, spillovers do not make us overinflate our impact estimates because the spillovers bias our treatment effect towards zero.

Two important points require further clarification on the possible spillovers and the justifications for randomizing at the individual level provided above. First, even though the treatment is complex enough to minimize spillovers that may come from control households getting the treatment, we recognize that there could be another channel of spillovers through control households mimicking the behavior of treated households, especially if the behavior change of treatment individuals is visible. Since behavior mimicking is expected to be higher in public household chores than private household chores. We separately analyze the effect on public and private household chores to identify if there is spillover effect. Second, the difference between public and private household chores cannot be entirely attributed to spillovers. To clarify, suppose the treatment effect on the public household chores is smaller than the private household chores. This could be due to spillovers that raise the level of the public chores among the control group, or it could also be due to stigma that prevent men in the treatment group to raise the level of the public chores as much as the private action. Either of these make the treatment effect on public chores smaller than the treatment effect on private chores. Assuming that all behavior is on a flat trend going into treatment, we use before vs. after comparisons to differentiate these two mechanisms. Higher levels of both the private and public chores in both treatment and control suggest the spillover story while no change in the public behavior in the control and a small or no change in public behavior in the treatment group suggests the stigma story.

[5]Public chores include collecting fuelwood, fetching water, washing clothes, processing grains, and Private chores include cooking, child care, house cleaning, and washing dishes.

Another possible concern is related to mobile phone ownership. The experiment requires mobile phone ownership, but respondents may or may not own a mobile phone. Sampling only those who own mobile phones would make our sample non-representative of the population we study. In such a populations with low phone ownership and/or access, Dabalen et al., (2016) suggested that researchers should consider options to improve the representativeness of the sample, for example providing free phones to respondents. Providing free phones has been previously used by the World Bank on its Listening to Africa (L2A) serious of evaluations and by other researchers in similar context (Dabalen et al., 2016; Dillon, 2012; Gaddis et al., 2019). For this reason, the "Her Time" research project distributes free mobile phones to all respondents regardless of treatments status. Although providing phones improve representatives, it threatens external validity of our study in the context where phones are not provided. Furthermore, mobile ownership may have a potential effect on outcome variables. But this effect occurs in both treated and control groups and gets differenced out in the analysis. Accordingly, we interpret our results providing appropriate caveat about their external validity, particularly that the results may not be generalizable to poor households that do not participate in livelihoods groups and do not own mobile phones.

2.4 Outcome variables, Hypotheses, and estimation strategy

2.4.1 Outcome variables

We measure male's participation in eight household tasks – child care, collecting fuelwood, fetching water, cooking, washing clothes, processing grains, house cleaning, and washing dishes. We collected information if the man performed these activities yesterday and how many minutes he spends on each activity. We also collected information on the women's satisfaction with the amount of time that the man contributes towards household chores and child car. We ask this question both in general and specifically towards each of the eight household chores listed above. Furthermore, we collected information about the women's perception on how the male has changed his participation in household chores between baseline and endline.

Using the information collected, we construct two groups of outcome variables: one for men's and women's participation in chores and one for women's perception about men's participation in chores.

Men's and Women's participation in chores

Within this first group of outcome variables, we have the following two indicators: the number of minutes the man/woman spent on chores and whether the man/woman performed the listed household chores, i.e.:

$Y_{1,\ a}$: binary variable indicating whether the man/woman participated *in a given household chore* or not. Subscript a runs from 0 to 10, where subscripts 1 - 8 represent each household chore. For simplicity, we use subscripts $a\ =\ 0,\ \ 9\ and\ 10$ to represent household level aggregate outcome variables: all chores, public chores and private chares respectively. Therefore, $Y_{1,0}$ is an index variable constructed by counting

the number of household chores in which the man/woman participated, $Y_{1,9}$ is an index variable constructed by counting *the number of public household chores* in which the man/woman participated, and $Y_{1,10}$ is an index variable constructed by counting *the number of private household chores* in which the man/woman participated.

$Y_{2,\ a}$: binary variable measuring whether the man/woman participated in at least one household chore. Subscript a stands for $0,\ \ 9\ and\ 10$, repreeting all chores, public chores and private chares respectively. Therefore, $Y_{2,\ 0}$ is a binary variable measuring whether the man/woman performed at least one of the eight tasks yesterday, $Y_{1,\ 9}$ is a binary variable measuring whether the man/woman performed at least one of the *public household chores* yesterday, and $Y_{1,\ 10}$ is a binary variable measuring whether the man/woman performed at least one of the *private household chores* yesterday.

Women's perceptions about men's participation in chores.

Within the second group of outcome variables, we have the following two indicators: the woman's satisfaction level regarding the man's participation in chores and the woman's perceptions regarding the change in the man's participation in chores compared to baseline, i.e.:

$Y_{3,\ a}$: likert scale variable indicating how satisfied the woman is, with the amount of time the man contributes towards household chores. Subscript a runs from 0 to 10, where subscripts 1 - 8 represent *each household chore*. For simplicity, we use subscripts $a\ =\ 0,\ \ 9\ and\ 10$ to represent household level aggregate outcome variables: all chores, public chores and private chares respectively. Therefore, $Y_{3,0}$ is likert scale variable indicating how satisfied the woman is, with the amount of time the man contributes towards *household chores in general*, $Y_{3,9}$ is likert scale variable indicating how satisfied the woman is, with the amount of time the man contributes towards *public household chores*, and $Y_{3,10}$ is likert scale variable indicating how satisfied the woman is, with the amount of time the man contributes towards *private household chores.*

Y_4 : a variable indicating the women's perception of the change in the man's participation in household chores in general since baseline. This is a variable collected by asking the change in the man's participation in chores in general, without specifying chores, from baseline to endline. The responses were five levels of change options: *increased a lot, increased a little, no change, decreased a little, decreased a lot.*

2.4.2 Hypotheses and estimation strategy

Assuming randomization succeeded in providing comparable treatment and control groups, any differences between the groups is not correlated with the treatment assignment. Taking advantage of the baseline data collected, we estimate the treatment effect using. Analysis of Covariance (ANCOVA) which controls for the baseline value of the outcome variable[6]. The main specification is the following ANCOVA model:

$$Y_{iv,t=1} = \alpha + T_{iv} + \gamma Y_{iv,t=0} + \delta_v + \varepsilon_{iv} \quad \text{(Eq 1)}$$

[6]The 2nd group of outcome variables was only recorded on the endline. Therefore, we are not able to control for the baseline values for them.

where $Y_{iv,t=1}$ is an outcome variable for individual i from VESA v measured at endline and $Y_{iv,t=0}$ is its value measured at baseline. T_{iv} is the treatment indicator which equals one if individual i in VESA v is assigned to treatment. The coefficient of interest is β, which shows the intent-to-treat effect, δ_v is a VESA fixed effect. ε_{iv} is iid across individuals.

We estimate equation Eq(1) for all outcome variables listed in section 4.1. In line with the two groups of outcome variables, we have two groups of hypotheses:

Hypotheses group A: men's/women's participation in chores

We hypothesize an increase in men's participation and a decrease women's participation in chores due to the reinforcement calls. Using the indicators for this group of outcome variables listed above, we estimate the following equations:

$$Y_{a,iv,t=1}=\alpha+\beta_a T_{iv}+\gamma Y_{a,iv,t=0}+\delta_v + \varepsilon_{iv} \quad \text{(Eq 2)}$$

where all definitions in equation (1) apply and outcome variable Y_a represents each outcome variable (Y_{1a} and Y_{2a}). Subscript a runs from 0 to 10, where subscripts 1 - 8 represent *each household chore*, and subscripts $a = 0,\ 9\ and\ 10$ represent household level aggregate outcome variables: all chores, public chores and private chares respectively. Therefore, equation (2) represents eleven equations within each outcome variable. These regressions generate the following core hypotheses:

Null Hypothesis A: $\beta_a = 0\ for\ each\ a$

For the chores level outcome variables ($a = 1 - 8$), the coefficient of interest β_a measures the effect of treatment on men's/women's participation in household chore a. When the outcome is measured as a dummy indicating participation in specific chore (Y_{2a}), it is interpreted as the difference in the likelihood of men's/women's participating in chore a between treated and control households. The set of *'null hypotheses A'* states that the treatment effect is zero for each household chore. Rejection of any of these null hypotheses implies $\beta_a \neq 0\ for\ each\ a$

For the household level aggregate outcome variables ($a = 0,\ 9\ and\ 10$), the coefficient β_a measures the effect of treatment on males' participation in all household chores, in public chores and in private chores, respectively. These coefficients measure the aggregated effect on males' participation in household chores as opposed to the effect of treatment on participation in specific chores. When the outcome variable is measured by the number of minutes men/women spent performing chores (Y_{1a}), it is interpreted as the average difference in the number of minutes men/women spent on the aggregated chores (all chores, public chores and private chores) between treated and control households. When the outcome is measured as an index constructed by counting the number of household chores in which the man/woman participated (Y_{2a}), it is interpreted as the difference in the number of household chores in which the man/woman participated out of the eight household chores, the four public chores and the other four private chores listed.

Hypotheses group B: women's perceptions about men's participation in chores

We hypothesize an imrpovment in women's perception about men's participation in chores due to the reinforcement calls. We identified two indicators within this group of outcome variables. First, the woman's satisfaction level on the man's participation in chores. Second, the woman's perception on the change in the man's participation in chores from baseline to endline. Using these indicators, we estimate the following equations:

$$Y_{a,iv,t=1}=\alpha+\beta_a T_{iv}+\gamma Y_{a,iv,t=0}+\delta_v + \varepsilon_{iv} \quad \text{(Eq 3)}$$

where all definitions in equation (2) apply and outcome variable Y_a represents each outcome variable (Y_{3a} and Y_4). Subscript a runs from 0 to 10, where subscripts 1 - 8 represent *each household chore*, and subscripts $a = 0,\ 9\ and\ 10$ represent household level aggregate outcome variables: all chores, public chores and private chares respectively. Therefore, equation (3) represents eleven equations for outcome variable Y_{3a} and one equation for outcome variable Y_4. These regressions generate the following core hypotheses:

Null Hypothesis B: $\beta_a = 0\ for\ each\ a$

For the woman's satisfaction level on the man's participation in each chore outcome variable ($a = 1-8$), β_a measures the treatment effect on women's level of satisfaction about the man's participation in chore a. For the household level aggregate outcome variables ($a = 0,\ 9\ and\ 10$), the coefficient β_a measures the effect of treatment on women's level of satisfaction about the man's participation in all household chores, in public chores and in private chores, respectively. When using Y_4 as an outcome variable, β_a measures the effect of treatment on the woman's perception about the change in the man's participation in chores compared to baseline. The coefficient measures the aggregated effect on women's level of satisfaction as opposed to the treatment effect on the women's level of satisfaction about men's participation in specific chores. The set of *'null hypotheses B'* states that the treatment effect is zero. Rejection of any of these null hypotheses implies $\beta_a \neq 0\ for\ each\ a$.

2.4.3 Heterogeneity analysis

Heterogeneity by the age of the targeted men

Younger men may show a different treatment impact compared to older men. We sorted the list of individuals by age within GNI treatment assignment before assigning individuals to treatment for this experiment. To see if there is heterogeneity of treatment impact by age, we estimate the following equation.

$$Y_{iv,t=1}=\alpha+\beta_1 T_{iv}+\beta_2 Young_{iv} * T_{iv}+\gamma Y_{iv,\ t=0}+\delta_v + \varepsilon_{iv} \quad \text{(Eq 4)}$$

where $Young_{iv}$ refers to a binary variable indicating that individual i from VESA v is young. The men in our data are on average 44 years of age, with a median value of 42. Based on this information, and also for simplicity, we decided to use 45 years of age as a cutoff point for young and old treated men.

Heterogeneity by participation in chores at baseline

To see if treatment effect varies by participation in household chores at baseline, we estimate a similar equation like equation 4 where a variable indicating the man participated in chores ($Y_{iv,\ t=0}$) at baseline interacted with treatment is included. i.e.

$$Y_{iv,t=1}=\alpha+\beta_1 T_{iv}+\beta_2 Y_{iv,\ t=0} * T_{iv}+\gamma Y_{iv,\ t=0}+\delta_v + \varepsilon_{iv} \quad \text{(Eq 5)}$$

where $Y_{iv,\ t=0}$ is a binary variable which indicates whether the individual i participated in the household chore under consideration at baseline. The coefficient of interest is β_2, which measures the heterogeneous treatment effect that individuals who participated in chores at baseline may exhibit.

2.5 Results

2.5.1 Balance test and descriptive results

We start with a test of balance between the treatment and control groups across some covariates and values of outcome variables. Table 2.1 presents the results, and it suggests that randomization succeeded in providing comparable treatment and control groups. Table 2.1 shows that none of the differences between the treatment and control group households are statistically significant. That is, any differences between the groups based on the baseline values of covariates and outcome variables are not correlated with treatment assignment. Table 2.1 also presents descriptive statistics on the differences between the woman and her partner in their participation in chores. Not surprisingly, women perform a greater number of household chores and spend more time on household chores than their male partners.

Table 2.1: Comparison of the treatment and control groups at baseline

	(1)		(2)		(1-2)#
	Treated [N=125]		Control [N=106]		T - C
Variables	Mean	[SE]	Mean	[SE]	Difference
Did the man ever attend school [1=Yes]	0.333	[0.043]	0.286	[0.044]	0.048
HH participated in non-farm business [1=Yes]	0.282	[0.041]	0.292	[0.044]	-0.01
HH owns at least one phone [1=Yes]	0.363	[0.043]	0.443	[0.048]	-0.08
The Man's age in years	42.756	[1.096]	44.924	[1.275]	-2.168
The Man's participation in chores at baseline					
Number of tasks performed: All	0.232	[0.072]	0.208	[0.051]	0.024
Private	0.176	[0.046]	0.151	[0.042]	0.025
Public	0.056	[0.035]	0.057	[0.023]	-0.001
Number of minutes spent: All	8	[2.051]	8.396	[2.913]	-0.396
Private	7.28	[2.008]	7.311	[2.833]	-0.031
Public	0.72	[0.510]	1.085	[0.614]	-0.365
Performed at least one task [1=Yes]: All	0.169	[0.034]	0.179	[0.037]	-0.01
Private	0.145	[0.032]	0.132	[0.033]	0.013
Public	0.038	[0.019]	0.067	[0.026]	-0.029
The Woman's participation in chores at baseline					
Number of tasks performed: All	3.296	[0.162]	3.34	[0.169]	-0.044
Private	2.432	[0.110]	2.368	[0.122]	0.064
Public	0.864	[0.080]	0.972	[0.083]	-0.108
Number of minutes spent: All	233.368	[16.632]	222.274	[18.290]	11.094
Private	192.824	[15.760]	181.528	[16.617]	11.296
Public	40.544	[5.347]	40.745	[5.638]	-0.201
Performed at least one task [1=Yes]: All	0.951	[0.020]	0.943	[0.023]	0.008
Private	0.935	[0.022]	0.925	[0.026]	0.01
Public	0.714	[0.044]	0.789	[0.043]	-0.075

We conducted a t-test on the differences of the means, and *s represent significance levels with * p<0.10, ** p<0.05, *** p<0.01

2.5.2 Effects of treatment on participation in chores

Men's and Women's participation in chores

We have two main outcome variables that measure men's and women's participation in the eight household chores: the number of tasks performed and whether at least one task was performed. For both outcome variables, we disaggregate results by whether the chores are private or public. Figure 2.1 presents the treatment coefficients based on equation 2 estimated using the 'number of tasks performed' outcome variable, also disaggregated by public vs private tasks. The results show that men's participation increased in public chores, although the increased contribution from men did not lead to a decrease in the woman's participation in chores. Figure 2.2 presents the results on the probability of performing at least one of the eight tasks. We observe that the men's probability of performing at least one public task increased significantly. Although not significant, the coefficients on performing at least one task (Figure 2.2) are positive and larger for men and others compared to the woman.

The findings above suggest that an increase in men's participation in household chores doesn't necessarily lead to a decrease in women's participation in chores, at least in the short run. It might be the case that women's participation in chores will decrease if such increases in men's participation are sustained over a long period. Therefore, given the short period between the intervention and the endline data collection, it is not surprising not to see men's added contribution crowding out women's participation in chores. Another possible explanation could be that the household was not getting enough labor to fully undertake household chores. In such cases, additional labor contribution by men would not lead to a decrease in labor contribution by others like the woman.

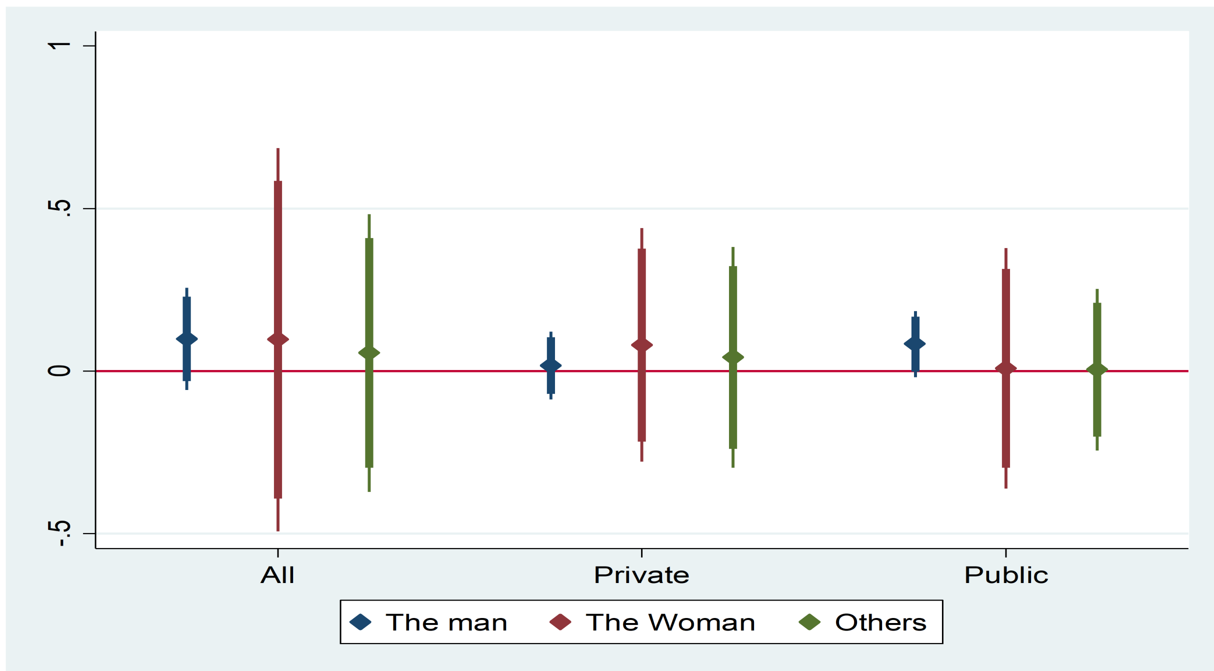


Figure 2.1: Treatment effects on the number of tasks performed by the woman, the man, and others

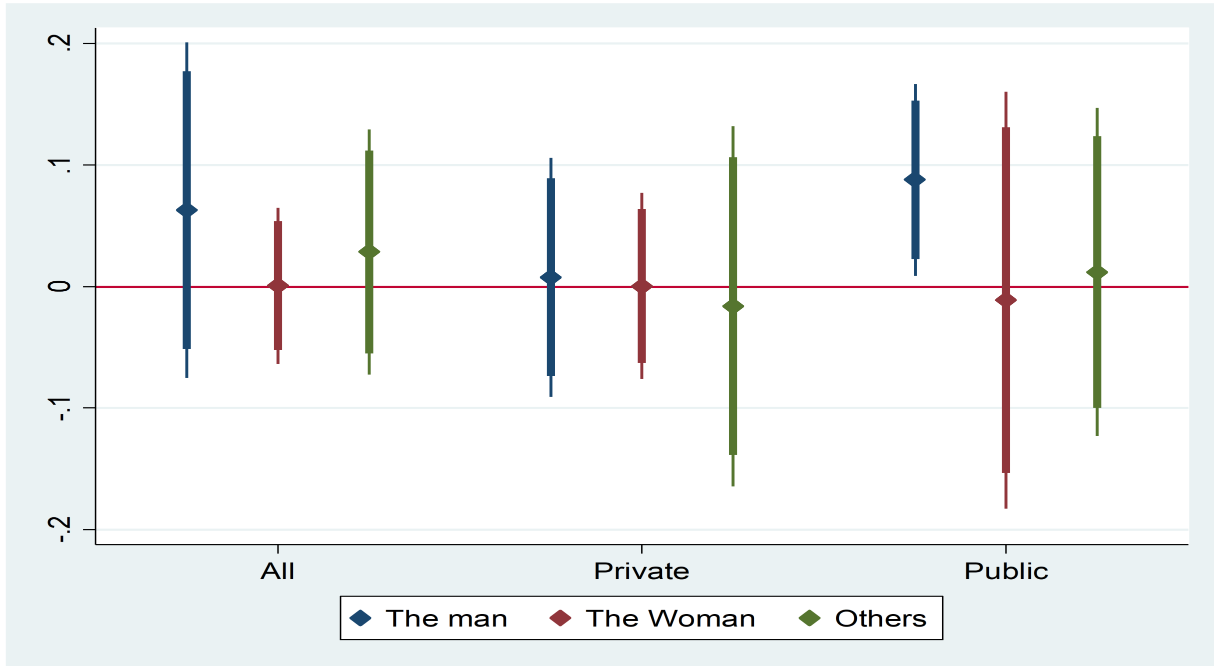


Figure 2.2: Treatment effects on 'performing at least one task' by the woman, the man, and others

Other household members and non-members' participation in chores

The results in the previous section show participation by others (household members and non-members) has not changed due to the treatment. But further disaggregating results by types of household members reveal new findings. Although insignificant, children's participation in chores shows a modest increase. Daughters' participation in chores increased, largely due to increased participation in private chores. Daughters' participation in terms of performing at least one task (Figure 2.3) and in terms of the number of tasks they perform (Figure 2.4) increased more compared to others in private household chores. Sons' participation in terms of performing at least one task (Figure 2.3) increased more compared to others in public chores. Figures 3 and 4 also show that participation in chores from non-members decreased more compared to others, especially so in private household chores where non-members' participation is the only negative coefficient.

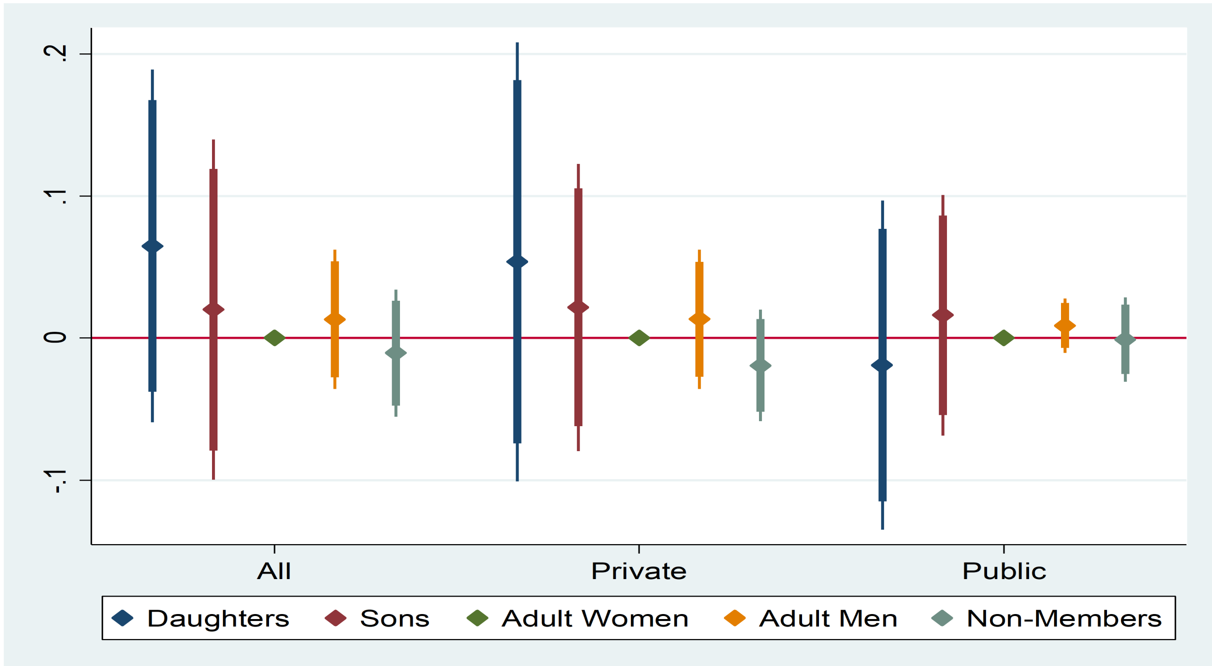


Figure 2.3: Treatment effects on 'performing at least one task' by types of members

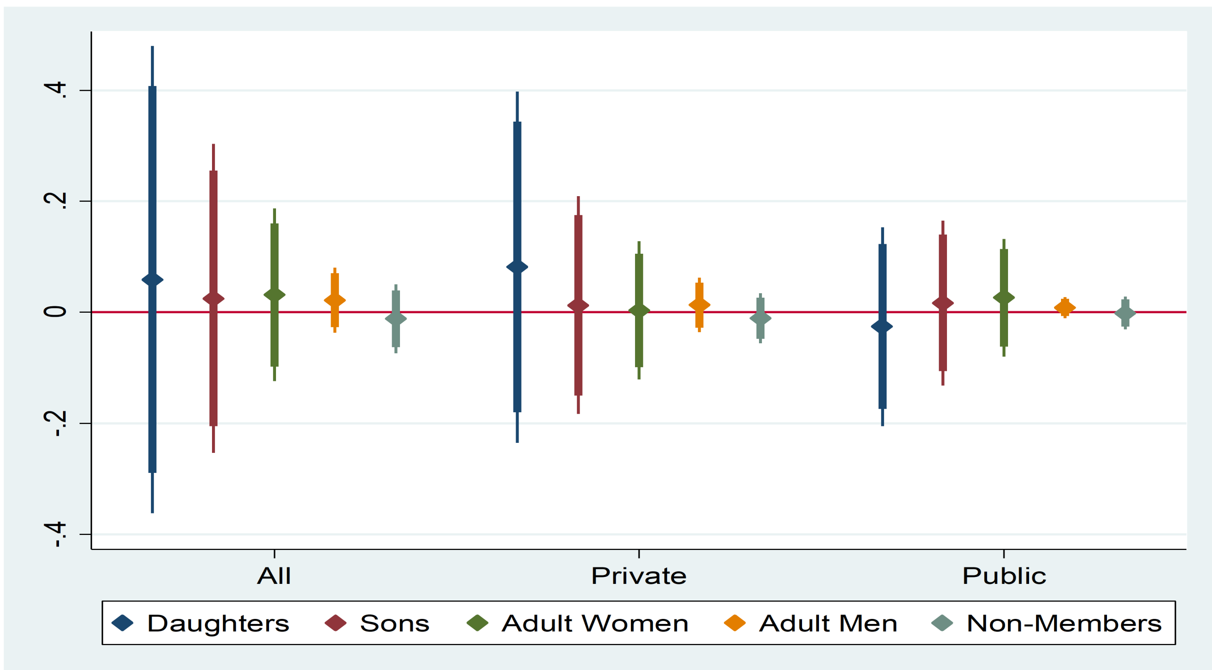


Figure 2.4: Treatment effects on the number of tasks performed by types of members

Men's and Women's participation in different types of household chores

We further disaggregated results by participation in each type of household chores under consideration. Doing so allows us to understand what drives the results discussed above. Figure 2.5 below presents the man's and the woman's participation in chores side by side. The results show positive coefficients on most of the tasks for the man, with a considerable effect on processing grains, and even a statistically significant coefficient on firewood. The results are mixed for women, with a significant increase in the probability of participating in washing clothes and a relatively higher coefficient on her participation in washing dishes. Although not significant, we also see a decrease in the woman's probability of performing fetching water and processing grains with a relatively larger decrease in her participation in child care.

The findings in Figure 2.5 are consistent with what we found in the previous section that men's participation in public household chores has significantly increased. The coefficients on the man's participation in public chores (firewood, fetching water, washing clothes, and processing grains) are relatively higher than the coefficients on the man's participation in private chores (cooking, child care, cleaning, and washing dishes). The significant increase in the man's participation in public chores could be largely driven by his increased participation in firewood collection, although the increase in his participation in the other public tasks, e.g., processing grains, is also not negligible.

Given public chores are easily observable, it is surprising to see a significant increase in the man's participation in firewood collection compared to private chores. There can be a couple of reasons why we are seeing these results. First, firewood is usually collected using cart animals which make the task easier. Therefore, it could just be easier than the private chores. Second, the gender norm on the division of labor could be weaker on firewood collection compared to the private chores. Therefore, firewood may have relatively better acceptability to be performed by men compared to the private chores. Generally, the difference between public and private chores may not only be observability and other factors like difficulty of the task and the level of gender norm attached to the task determine whether the man will perform it or not.

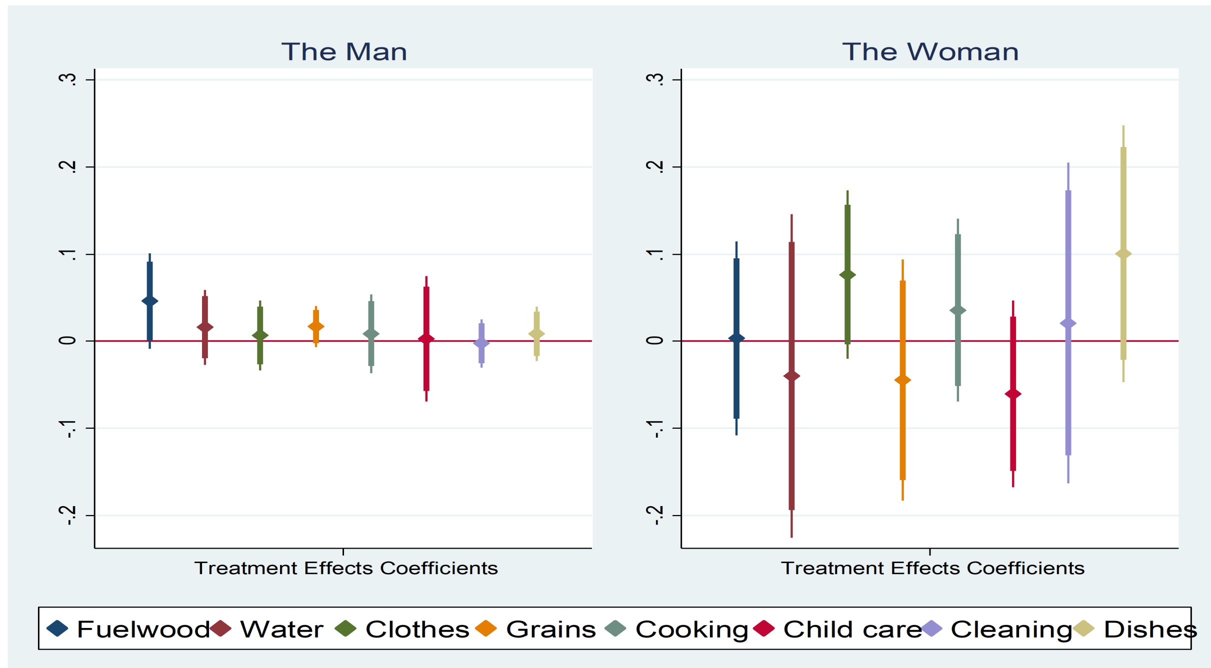


Figure 2.5: Treatment effects on the probability of participating in different types of household chores

Other household members' and non-members' participation in different types of household chores

In previous sections, we found that children's participation in chores increased: daughters increased their participation in private chores while sons increased their participation in public chores. In Figure 2.6 below, we disaggregated the results by different types of household chores. The results show that daughters' increased participation in private chores is largely driven by their significantly increased participation in child care while sons' increased participation in public chores is largely driven by their significantly increased participation in firewood.

These results indicate that increased participation from children is observed, although the intervention targeted to increase men's participation in chores. It is important to note that such increased participation of children in chores may have a long-term negative effect on households' wellbeing if children's contribution leads to decreased time spent on schooling.

Furthermore, the results provide some insight into the gender norms in the division of labor that is getting passed to the next generation. We observe that sons increased their participation in collecting firewood, exactly what the man is doing. Daughters increased their participation in child care, and it is a norm in rural parts of Ethiopia that child caring is "the women's task". Although the intervention increased men's and children's participation in chores, the types of household chores they picked follow the existing gender norms to some extent. It is important not to undermine the increased men's and sons' participation in collecting firewood, which is also assumed to be a "women's task", but the norm is weaker on collecting firewood compared to those tasks with strong norms, e.g., cooking or child caring. Even

after men are convinced that they should increase their participation in household chores, they tend to pick tasks with weaker gender norms. These results show the need for multifaceted interventions to tackle deep rooted gender norms.

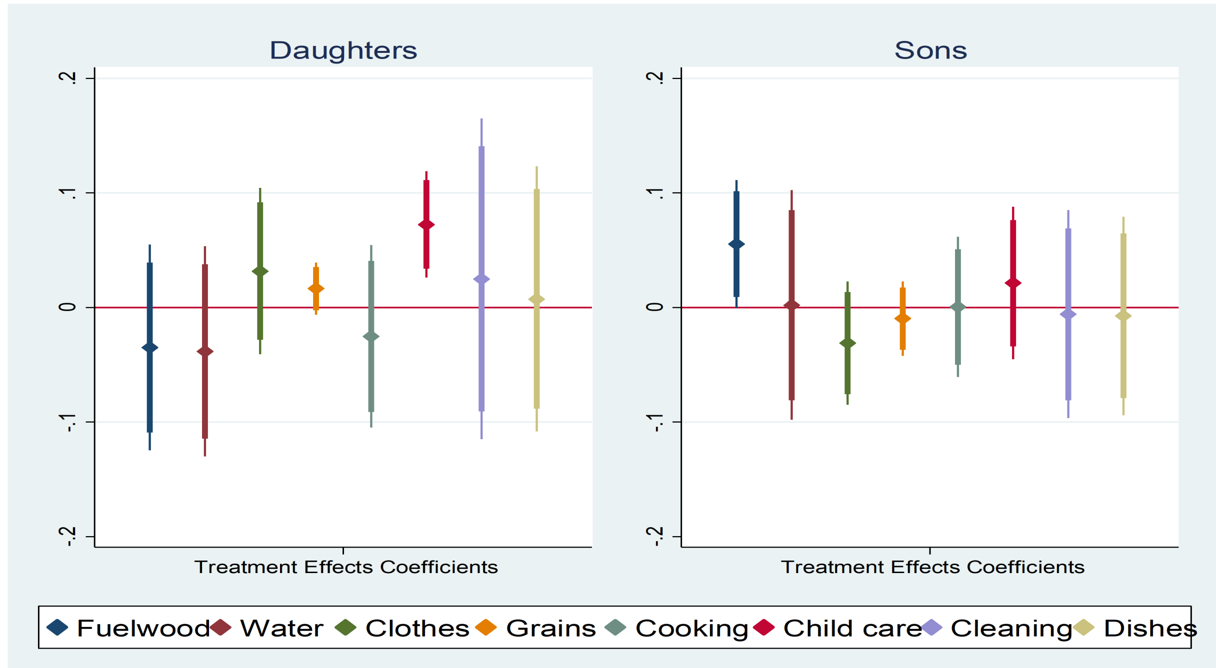


Figure 2.6: Treatment effects on children's probability of participating in different types of household chores

We were unable to do similar estimations disaggregated by tasks for adult women due to few observations who performed the chores. Figure 2.7 below shows the treatment effects on the participation of adult men and non-members in household chores. There is a significant decrease in the participation of non-members in child care. This could be because the man encouraged daughters to participate more in child care (see Figure 2.6) after he realized the increased burden on the woman. Therefore child care that used to be performed by non-members is now being performed by daughters.

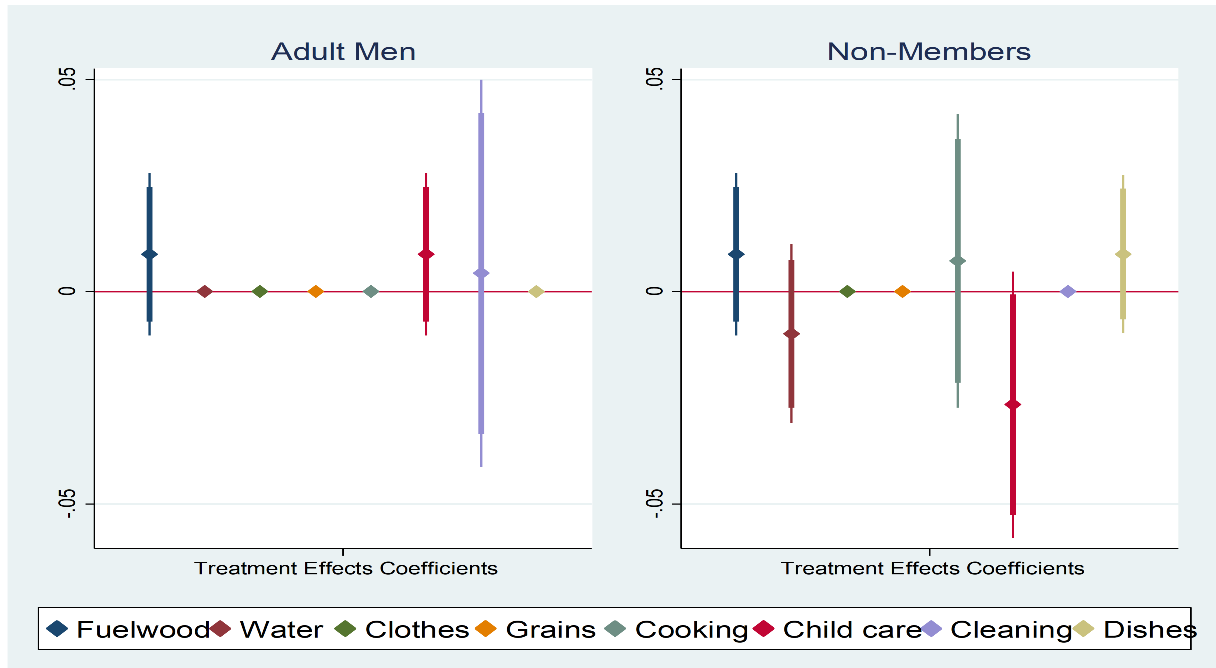


Figure 2.7: Treatment effects on adult men's and non-members' probability of participating in different types of household chores

2.5.3 Effects of Treatment on women's perceptions of men's participation in chores

We use two indicators to measure the woman's perception in household chores: the woman's level of satisfaction in the amount of time the man contributes towards chores, and the woman's perception of the change in the level of the man's participation in chores. We asked the woman's level of satisfaction for household chores in general and each activity. Therefore, we disaggregate results to private and public chores (Figure 2.8) and each type of household chore (Figure 2.9). Responses were recorded on a Likert scale ranging from "very unsatisfied" to "very satisfied."

The results in Figure 2.8, an estimation of equation 3, show that the women's level of satisfaction in men's participation in private household chores improved significantly. This finding makes sense given the strong stigma attached to public chores as compared to private chores. Although insignificant, we also observe positive coefficients for public chores and all chores.

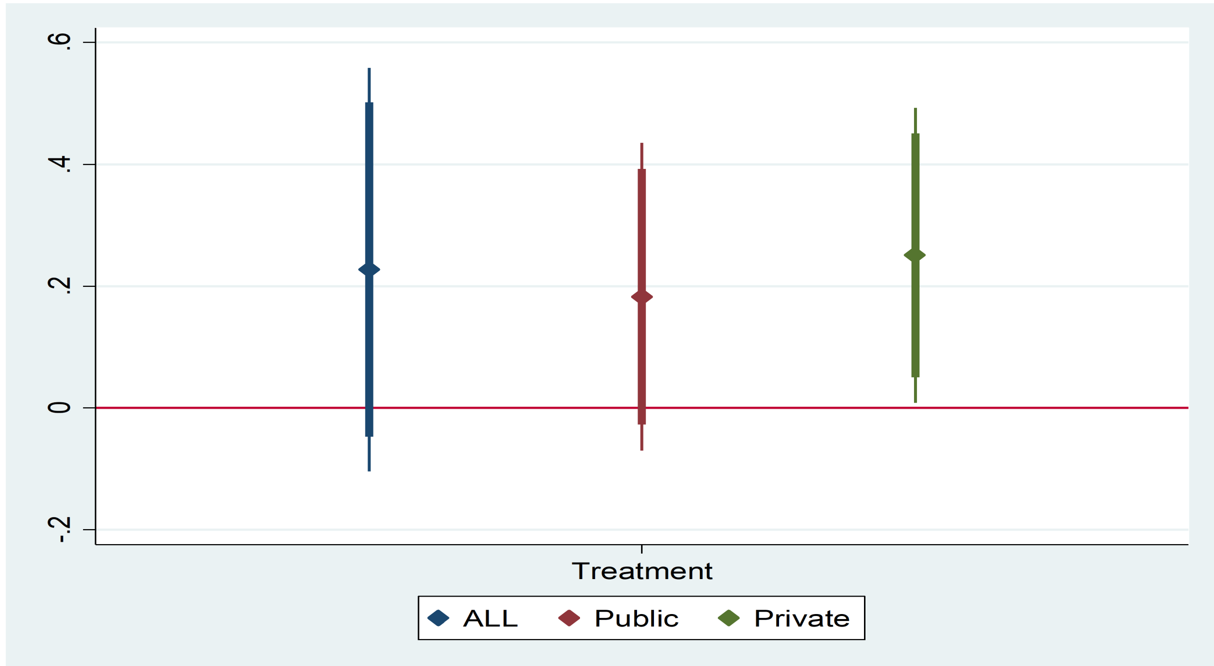


Figure 2.8: The effect of treatment on the Women's Level of satisfaction in the amount of time that Men contribute towards household chores

Further disaggregating results to different types of household chores, presented in Figure 2.9 below, shows that we observe positive and significant treatment effects for collecting fuelwood and cooking. The effect on the women's level of satisfaction in the amount of time the man is spending on fuelwood is consistent with our findings in the previous sections that the man significantly increased his participation in firewood collection. The increase in the woman's satisfaction from the man's participation in cooking reveals an interesting story about the level of gender norms attached to different types of chores. Although we do not observe a significant increase in the man's participation in cooking, the women's level of satisfaction in his participation in cooking significantly increased. This could be because the gender norm attached to cooking is so high that even a small increase in men's participation in cooking is rewarded by a significantly higher level of satisfaction from the woman.

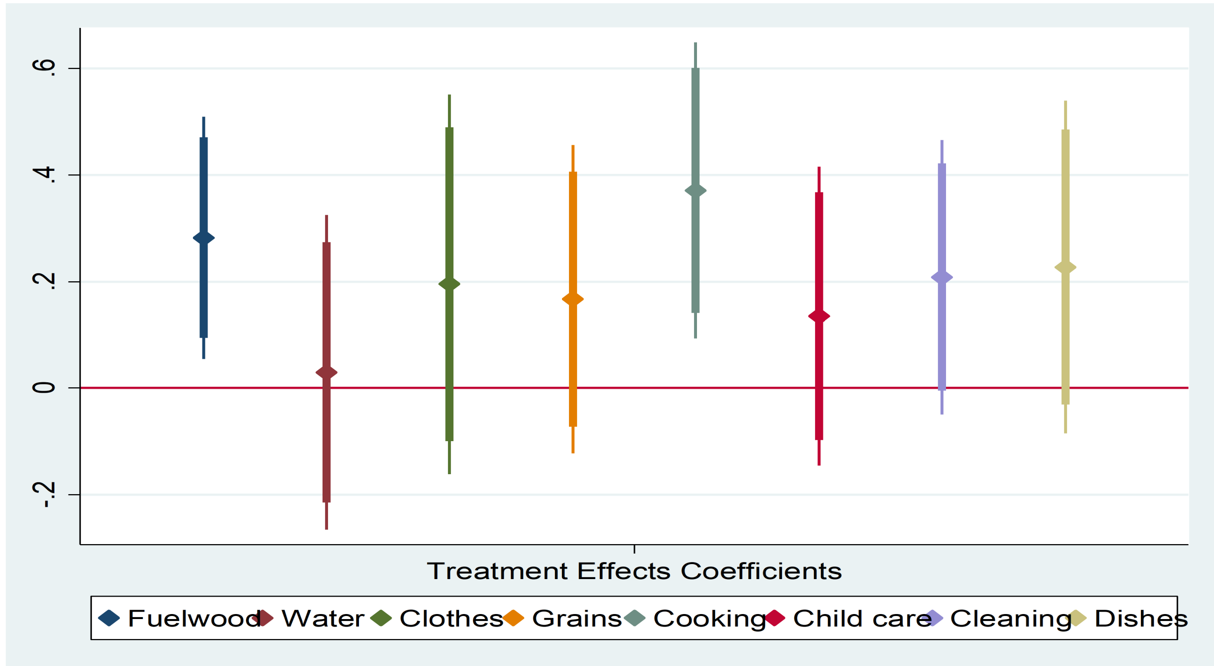


Figure 2.9: The effect of treatment on women's level of satisfaction in the amount of time that men contribute towards household chores, by activity

Finally, we wanted to see how the women perceive the change in the man's participation in chores from baseline to end line. Her responses were recorded on a Likert scale ranging from "decreased a lot" to "increased a lot.". Table 2.2 below presents the results based on equation 3 and it shows a positive and significant increase in women's perceptions about the change in men's participation in household chores from baseline to end line as a result of the treatment.

Table 2.2: The effect of treatment on women's perceptions about changes in the level of men's participation in chores

	The Women's perception on the change to Men's participation
Treated	.314**
	(.135)
Constant	3.51***
	(.0677)
N	229

Standard errors in parentheses; * p<0.10, ** p<0.05, *** p<0.01

2.6 Heterogeneity analysis

2.6.1 Heterogeneity by age

The treatment effects vary by age. The effect of treating young men is different from treating older men. Figure 2.10 presents the treatment effects on the number of tasks performed based on equation 4 for the man and the sons in the household. The results show that the treatment effects are larger and significant for young men, especially for private household chores. There is no heterogeneity by age in the treatment effects on the number of public tasks performed by the man. Previously in section 5.1, we found that treatment increased men's participation in public chores and not in private chores. Our results here show that this is not the case for young men, for whom we found a significant increase in their participation in private chores. This result makes sense given the less observable nature of private chores. The result also suggests that young men are more interested in less observable household chores than old men, which may indicate that young men are more restricted by gender norms on the division of labor, especially if it is observable by others. This is also reflected in the difference between young and old men in terms of letting their sons participate in public household chores. Figure 2.10 shows that participation in public household chores by sons of old men increased due to the treatment, while it decreased participation by sons of young men. Old men may have encouraged their sons either through advice or by serving as a model to participate in public chores.

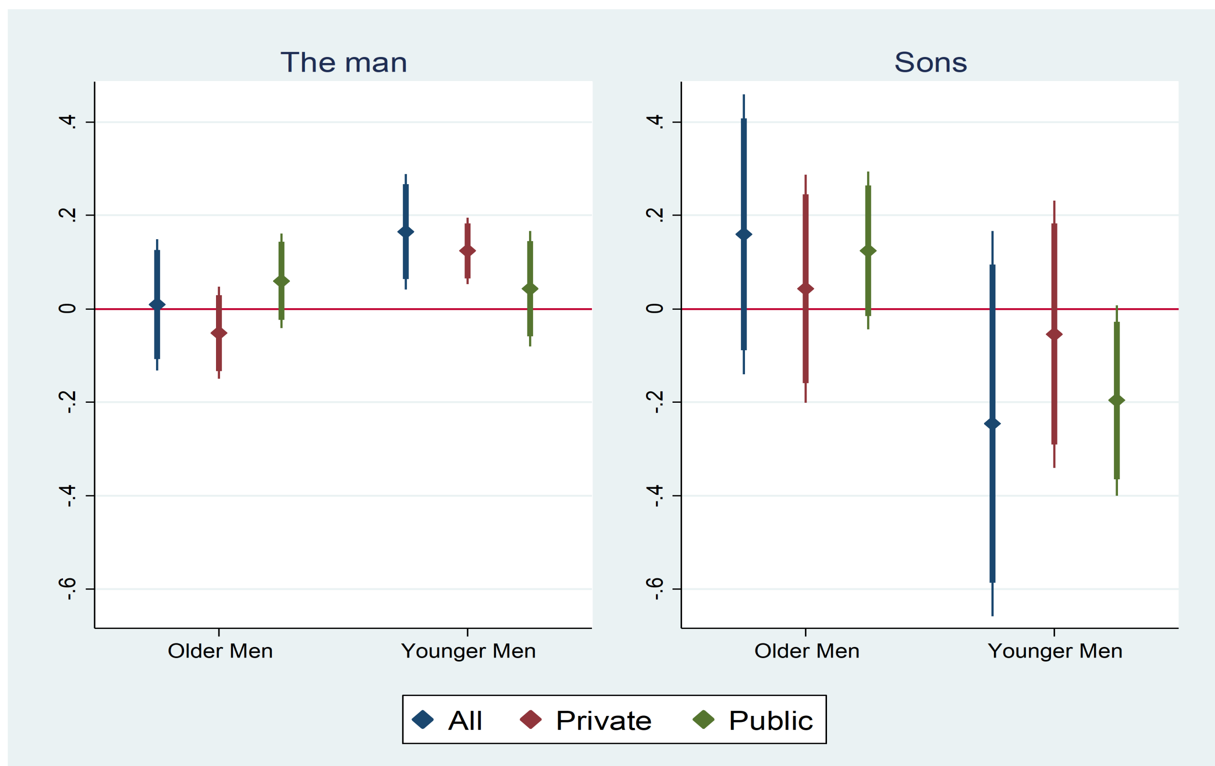


Figure 2.10: Treatment effect heterogeneity by age: Number of tasks performed by the Man and Sons

Figure 2.11 below presents the treatment effect heterogeneity by age on performing at least one task based on equation 4. We observe very similar results as presented in Figure 2.11, with an even stronger and significant treatment effect on older men's increased participation in public chores and younger men's increased participation in private chores.

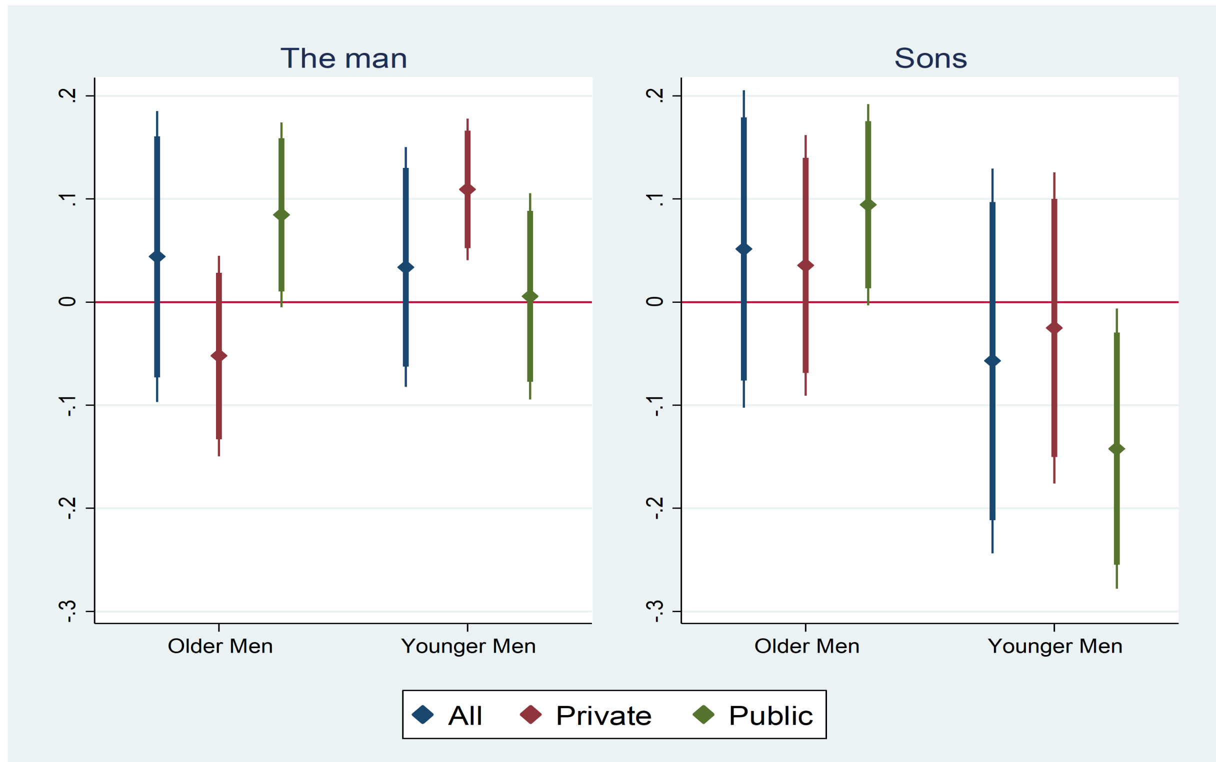


Figure 2.11: Treatment effect heterogeneity by age: Performing at least one task by the Man and Sons

Further disaggregation by types of household chores strengthens the above story and provides more information on the types of public and private chores being performed by old and young men. Figure 2.12 presents the treatment effects on the probability of performing selected types of household chores. Previously (Figure 2.5), we found that men's probability of performing fuelwood significantly increased due to the treatment. But heterogeneity of results by age, presented in Figure 2.12 below, reveals that young men respond to the treatment differently by increasing their participation in child care instead. The woman's participation in processing grains shows a significant decrease in the young men category. This decrease could be a reflection of the modest increase in the young men's participation in processing grains.

Figure 2.12 also shows that older men and their spouses decreased their participation in child care due to the treatment. The results in Figure 2.13 below show that daughters of older men have increased their participation in child care due to the treatment. Therefore, the treatment may have shifted child caring responsibility from parents to daughters. Figure 2.13 also shows treatment effects on sons' probability of

collecting fuelwood with a positive and significant effect on sons older men while we observe the opposite on sons of younger men.

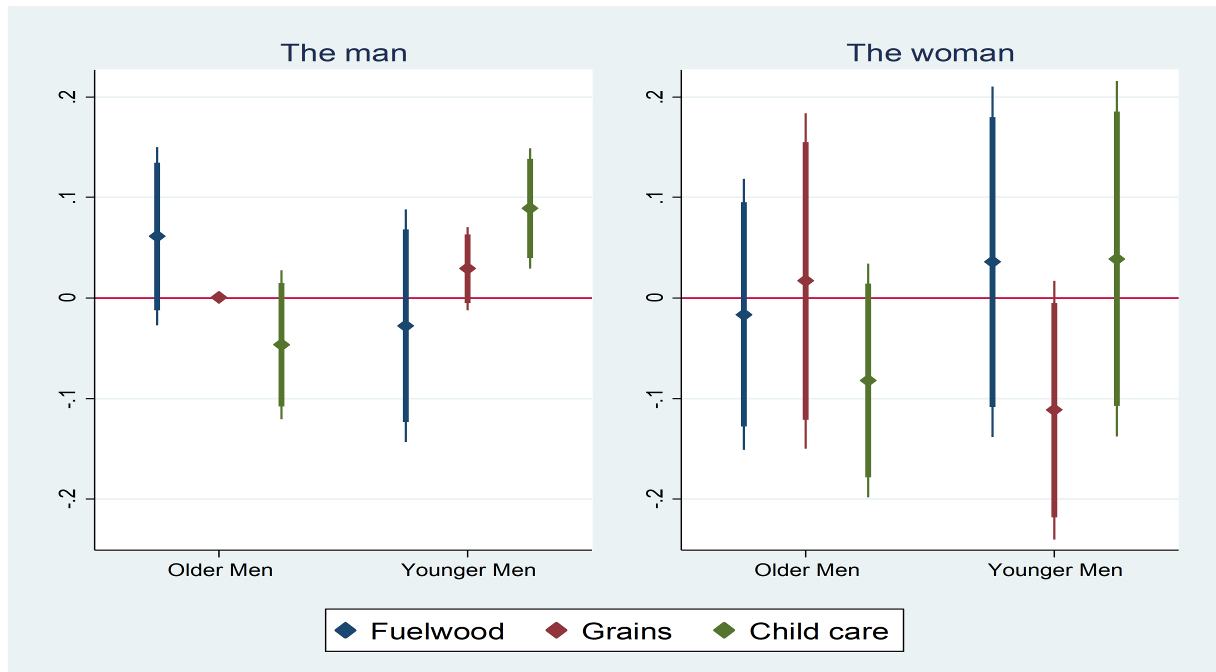


Figure 2.12: Treatment effect heterogeneity by age: Probability of Performing tasks: the man and the woman

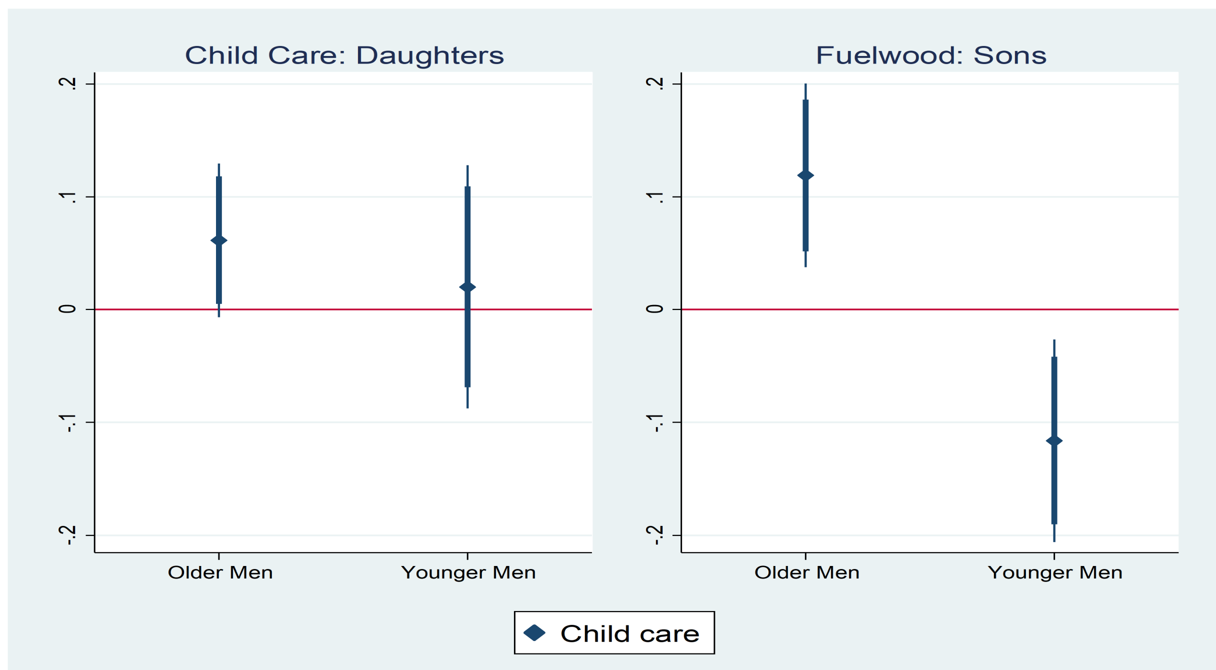


Figure 2.13: Treatment effect heterogeneity by age: Probability of Performing tasks: Daughters and Sons

2.6.2 Heterogeneity by participation at baseline

Figure 2.14 below presents the estimation results based on equation 5. The results show that the probability of men participating in fuelwood collection increased significantly among those who have not been conducting fuelwood collection and decreased significantly among those who have been conducting fuelwood collection at baseline. This suggests that the treatment effect on men's participation in fuelwood discussed in previous sections is largely driven by men who have not been conducting fuelwood collection. The negative treatment effect on those who have conducting fuelwood warrants further analysis. One possibility is that these men switched to performing other household chores. We cannot check this from Figure 2.14 because coefficients for participation in other types of chores tell us treatment effects among those who conducted those other types of chores at baseline and not necessarily fuelwood. We want to see treatment effects on all other types of chores among those who conducted fuelwood collection at baseline. Therefore, we estimate slightly modified version of equation 5 for all household chores, but interacting treatment with participation in fuelwood collection at baseline. [7] The results, also presented in Figure 2.14, show that although men who conducted fuelwood at baseline decreased their participation on fuelwood, they increased their participation in child care due to the treatment.

Figure 2.14: Treatment effect heterogeneity by the man's participation at baseline

[7]The specific equation we estimated is the following,

$$Y_{aiv,t=1}=\alpha+\beta_1 T_{iv}+\beta_2 Y_{a=fuelwood,\ iv,\ t=0} * T_{iv}+\gamma Y_{aiv,\ t=0}+\delta_v+\varepsilon_{iv}$$

where all the definitions for equation (5) apply and a stands for types of chores.

Figure 2.15 below presents the results for the woman, daughters, and sons. We see that spouses and daughters of men who have been conducting fuelwood at baseline have increased their participation in fuelwood collection due to the treatment. This matches the man's decreased participation in fuelwood collection. We also see a modest decrease in the woman's participation in child care, matching men's increased participation in child care from those who have been conducting fuelwood at baseline. Furthermore, the results also show that spouses of men who have been conducting the chores at baseline significantly increased their participation in washing cloths and cooking, and significantly decreased their participation in processing grains and house cleaning due to the treatment. It is important to note that only few men are reported to do washing clothes and house cleaning both at baseline and endline. For that reason, we were not able to estimate the regressions for those who have been conducting these chores at baseline. Given this information, it is hard to rely on the significant increase in the woman's participation in washing clothes and the significant decrease in the Waman's participation in house cleaning.

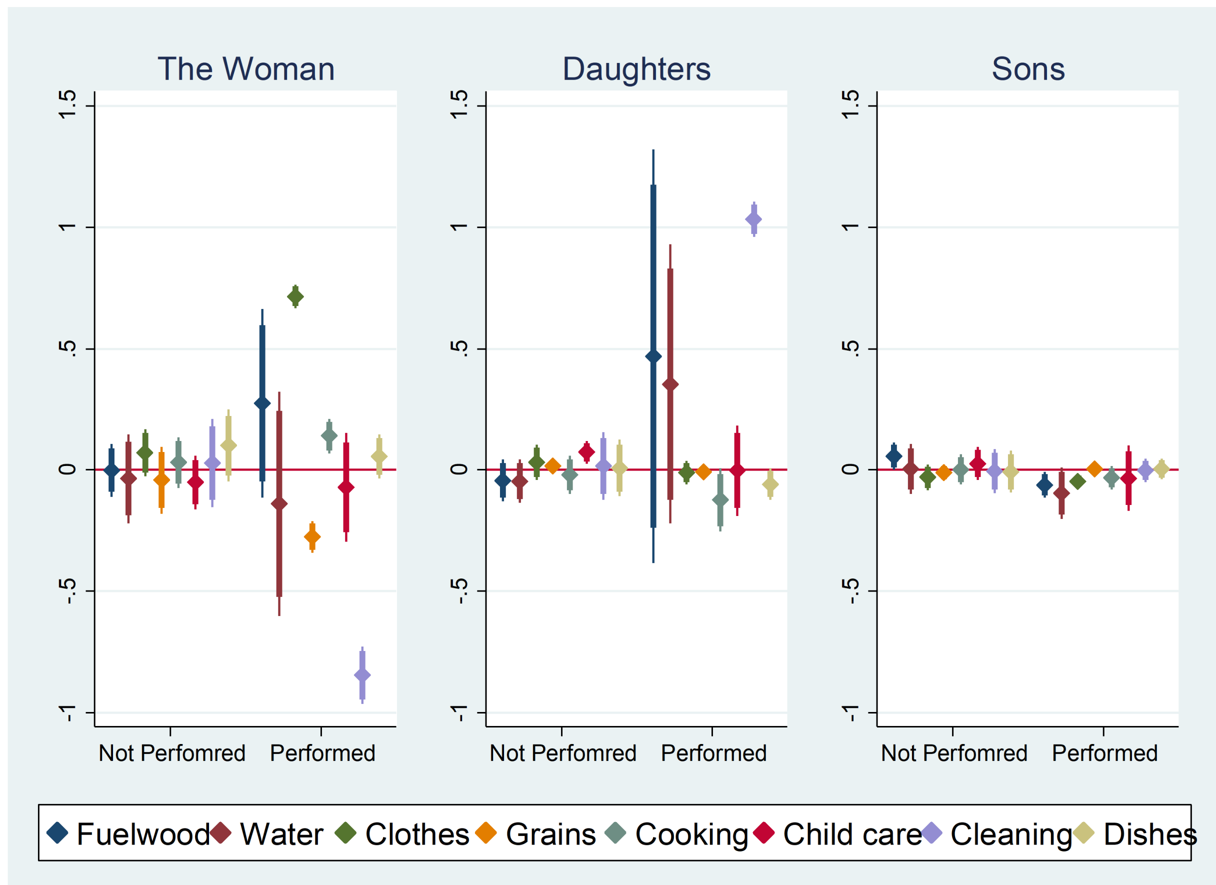


Figure 2.15: Treatment effect heterogeneity by the man's participation at baseline

2.7 Discussions and conclusion

We used a randomized control trial to test if mobile phones can be used to encourage and sustain behavior change in developing countries. We tested if men's participation in household chores can be affected by delivering repeated behavioral reinforcement messages over the phone. Our results indicate that men's participation, specifically in public chores, increased due to the treatment without a corresponding decrease in women's participation. Lack of decrease in the women's participation in chores despite an increase in men's participation may indicate that there was a labor shortage to fully conduct all household chores and/or the increase in men's participation needs to be sustained over a relatively long period to realize a decrease in women's participation. Further disaggregation by types of household chores shows that men's participation in public chores was largely driven by their increased participation in fuelwood collection, and this could be either because it is easier, or the gender norm for men to collect fuelwood is relatively weaker.

Women's level of satisfaction was found to increase significantly due to men's increased amount of time spent on fuelwood collection. Surprisingly, women were found more satisfied in men's participation in cooking specifically, for which we did not find a significant increase in men's participation. Overall, we observed a significant increase in men's participation in public chores, but women were more satisfied with men's increased participation in private chores. It could be because women do not expect men to do chores with probably strong gender norms like cooking, and therefore get easily satisfied with even an insignificant increase in men's participation in private chores.

Although the treatment targeted increased participation in chores by men, we also observed significant changes in other members' level of participation in chores due to the treatment. This happens because treatment raised awareness among men about women's disproportionate burden regarding chores and that men respond to this treatment by encouraging others in the household to perform more chores. Specifically, daughters' and sons' participation in chores increased significantly. This needs to be further investigated for policy implications. If any increase in children's participation in chores is crowding out their time on schooling, it will have a long-term effect on the children's wellbeing. Furthermore, the difference in the types of chores that daughters and sons take up seems to indicate that gender norms are being passed over to the next generation. Daughters are found doing more private chores largely child caring while sons do more public chores largely fuelwood collection. It looks like sons are mimicking their father or even working together as we also observed men doing more fuelwood collection due to the treatment. Furthermore, daughters seem to be taking the larger burden in terms of undertaking household chores; in addition to a significant increase in child care, also due to a significant decrease in child care from non-household members, daughters were found doing more of processing grains due to the treatment. Therefore, interventions aiming to empower women must take this into account and work towards empowering women and daughters to break inter-generational gender norms. This finding highlights the complicated nature of addressing gender norms in developing countries.

Our analysis shows heterogeneity in treatment effects by age of the man being treated. Young men are found to take up private chores (specifically child-caring) due to the treatment while older men are

found taking up public chores (specifically fuelwood). In line with this, we do not see a significant increase in daughters of young men's participation in child caring while we see that daughters of older men significantly increased their participation in child caring. But daughters of young men are found doing significantly more of washing clothes instead. By doing more private chores, it looks like young men are avoiding being observed by others in the community. This is also reflected in whether they let their sons do household chores compared to sons of older men. Sons of young men have not increased participation in any household chores while sons of older men increased their participation in fuelwood collection. All in all, the intervention created not only an increase in the man's participation in chores but also a shift in who does which type of household chores by letting some members increase their participation in one type of household chore while decreasing their participation in other types of chores.

Generally, our intervention shows mobile phones can be used to change behavior in developing countries. Our analysis shows that reinforcing messages through phones have created changes in the division of labor in household chores. Similar phone-based interventions can potentially be used to address other behavioral constraints in low-income countries.

Bibliography

Avdeenko, A., Bohne, A., & Frölich, M. (2019). Linking savings behavior , confidence and individual feedback : A field experiment in ethiopia. *Journal of Economic Behavior and Organization*, *167*(2019), 122–151.

Bulte, E., Lensink, R., & Vu, N. (2016). Gender training and female empowerment: Experimental evidence from vietnam. *Economics Letters*, *145*, 117–119.

Bulte, E. H., Lensink, R., & Winkel, A. B. (2018). The impact of a gender and business training on income hiding: An experimental study in vietnam. *Journal of Economic Behavior and Organization*, *148*, 241–259.

Chase, C., & Do, Q.-T. (2012). Handwashing behavior change at scale evidence from a randomized evaluation in vietnam. *Global Scaling Up Handwashing Project*, (September).

Dabalen, A., Etang, A., Hoogeveen, J., Mushi, E., Schipper, Y., & Engelhardt, v. J. (2016). *Mobile phone panel surveys in developing countries: A practical guide for microdata collection*. The World Bank, Washington D.C.

Dillon, B. (2012). Field report using mobile phones to collect panel data in developing countries. *Journal of International Development*, *24*(February), 518–527.

Dupas, P. (2011). Health behavior in developing countries. *Annual Review of Economics*, *3*(2011).

Fitzsimons, E., Malde, B., Mesnard, A., & Vera-hernández, M. (2016). Nutrition , information and household behavior : Experimental evidence from malawi. *Journal of Development Economics*, *122*(September), 113–126.

Gaddis, I., Oseni, G., Palacios-Lopez, A., & Pieters, J. (2019). Measuring farm labor: Survey experimental evidence from ghana. *World Bank, Policy Research Working Paper*, (8717).

GSMA. (2018). The mobile economy. *GSMA Intelligence*, (35).

Isaac, O., & Delmas, M. A. (2016). The dynamics of behavior change : Evidence from energy conservation. *Journal of Economic Behavior and Organization*, *126*, 196–212.

Kerbaj, J., Toure, Y., Soto, A., Ms, A., Boudjema, S., Giorgi, R., Dufour, J. C., & Brouqui, P. (2017). Smartphone text message service to foster hand hygiene compliance in health care workers. *American Journal of Infection Control*, *45*, 234–239.

Luoto, J., Levine, D., Albert, J., & Luby, S. (2014). Nudging to use : Achieving safe water behaviors in kenya and bangladesh. *Journal of Development Economics*, *110*(2014), 13–21.

Madajewicz, M., Pfaff, A., Geen, V. A., Graziano, J., Hussein, I., Momotaj, H., Sylvi, R., & Ahsan, H. (2007). Can information alone change behavior ? response to arsenic contamination of groundwater in bangladesh. *Journal of Development Economics*, *84*, 731–754.

Mckenzie, D., & Puerto, S. (2017). Growing markets through business training for female entrepreneurs : A market-level randomized experiment in kenya. *World Bank Policy Research Working Paper*, (7993).

Mildon, A., & Sellen, D. (2019). Use of mobile phones for behavior change communication to improve maternal, newborn and child health: A scoping review. *Journal of Global Health*, *9*(2).

Östervall, L. W. (2017). Nudging to prudence ? the effect of reminders on antibiotics prescriptions. *Journal of Economic Behavior and Organization*, *135*, 39–52.

Pettifor, A., Lippman, S. A., Selin, A. M., Peacock, D., Gottert, A., Maman, S., Rebombo, D., Suchindran, C. M., Twine, R., Lancaster, K., Daniel, T., Gómez-Olivé, F. X., Kahn, K., & MacPhail, C. (2015). A cluster randomized-controlled trial of a community mobilization intervention to change gender norms and reduce hiv risk in rural south africa: Study design and intervention. *BMC Public Health*, *15*(1), 1–7.

Pulerwitz, J., Hughes, L., Mehta, M., Kidanu, A., Verani, F., & Tewolde, S. (2015). Changing gender norms and reducing intimate partner violence: Results from a quasi-experimental intervention study with young men in ethiopia. *American Journal of Public Health*, *105*(1), 132–137.

Rodríguez, C., & Saavedra, J. E. (2019). The persistent effects of youth savings reminders : Experimental evidence from text-message campaigns in colombia. *Journal of Development Economics*, *139*(June), 135–156.

Shah, B. A. K., & Ludwig, J. (2016). Option awareness : The psychology of what we consider. *American Economic Review*, *106*(5), 425–429.

Stork, C., Esselaar, S., & Chair, C. (2017). Ott - threat or opportunity for african telcos? *Telecommunications Policy*, *41*(7-8), 600–616.

Zurovac, D., Sudoi, R. K., Akhwale, W. S., Ndiritu, M., Hamer, D. H., Rowe, A. K., & Snow, R. W. (2011). The eff ect of mobile phone text-message reminders on kenyan health workers ' adherence to malaria treatment guidelines : A cluster randomised trial. *The Lancet*, *378*(9793), 795–803.