UC Santa Cruz
UC Santa Cruz Electronic Theses and Dissertations

Title
Essays on Technology Adoption in Developing Countries

Permalink
https://escholarship.org/uc/item/68x778w8

Author
Jeong, Dahyeon

Publication Date
2020

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

ESSAYS ON TECHNOLOGY ADOPTION IN DEVELOPING COUNTRIES

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

ECONOMICS

by

Dahyeon Jeong

June 2020

The Dissertation of Dahyeon Jeong
is approved:

Professor Jonathan Robinson, Co-Chair

Professor Ajay Shenoy, Co-Chair

Professor Alan Spearot

Quentin Williams
Acting Vice Provost and Dean of Graduate Studies

Copyright © by

Dahyeon Jeong

2020

Table of Contents

List of Figures

List of Tables

Essays on Technology Adoption in Developing Countries

Dahyeon Jeong

Abstract

This dissertation contains three essays broadly related to technology adoption in developing countries. The dissertation discusses a job search technology, an agricultural technology, and a monitoring technology in the context of agricultural input and labor markets in developing countries. Chapter 1 studies whether the adoption of a new job search technology reduces search frictions in rural labor markets in Tanzania. I develop an SMS-based technology that connects agricultural workers and employers instantly. In a clustered RCT experiment, I find that treatment villages experience a 16-40 percent reduction in within-village wage dispersion. Consistent with reduced wage dispersion, I find evidence that labor is reallocated within villages. Dispersion in per-acre labor input across farms decreases. Workers divert job applications from lower-paying to higher-paying employers.

In joint work with Shilpa Aggarwal, Brian Giera, Jonathan Robinson, and Alan Spearot, Chapter 2 studies the effect of remoteness on the adoption of chemical fertilizer, one of the most important agricultural technologies to enhance productivities. We quantify and estimate the effect through the lens of a structural spatial model, facilitated by an enormous data collection exercise in Tanzania. We find that villages in remote places face very different input and output prices than villages near a city hub. In our reduced form analysis, we find that one standard deviation increase in travel time is associated with 20-25 percent lower input adoption and output sales. Our simulated structural model predicts that reducing transportation costs to reach input-retailers by 50 percent doubles input adoption.

Chapter 3 (joint with Ajay Shenoy and Laura Zimmermann) studies whether corruption "greases the wheels" of government in the context where a digital monitoring platform is used for implementing government programs. We test this corruption-as-compensation hypothesis in Indian villages whose presidents, in charge of running a massive public works scheme, claim such behavior is widespread. We link millions of administrative job records from the digital platform to election outcomes, enabling us to measure both the presidents' performance and their self-dealing. We do not find evidence in favor of corruption-as-compensation hypothesis. Self-dealing declines over time in villages with plausibly better monitoring capacity (e.g. the presence of internet cafes), even though there is no decline in performance. In villages without such capacity, self-dealing persists.

Acknowledgements

I am incredibly fortunate to have Jonathan Robinson as my primary advisor who guided me from the very start of the doctoral program. He showed me first-hand how to implement a large scale field research. I am deeply grateful to him for showing me how this kind of fieldwork is possible in rural Africa and guiding me tirelessly throughout the entire cycle of the project. Without his caring support, my dissertation would not have materialized. I am heavily indebted to Alan Spearot for his warm encouragement and support. He was always incredibly generous with his time in providing guidance. Ajay Shenoy was exemplary for his self-discipline, organization, and critical thinking that are crucial for research. I would like to express my heartfelt gratitude to him for his persistent support and guidance.

I would also like to thank other economics faculty members at UCSC, who attended my seminars numerous times and offered me new and constructive feedback every time. I thank George Bulman, Jesse Cunha, Carlos Dobkin, Robert Fairlie, Laura Giuliano, Gueyon Kim, Kenneth Kletzer, Natalia Lazzati, Jessie Li, Justin Marion, Brenda Samaniego de la Parra, and Jeremy West. I also thank UCSC staff members, Susan Leach and Sandra Reebie for making this Ph.D. process seamless for me. I thank James Berry, Bryant Hyuncheol Kim, Tapan Mitra, and David Sahn from Cornell University, who supported my decision to pursue a Ph.D. economics program.

I would also like to thank IPA Tanzania staff members and enumerators who worked extremely hard, collecting data in remote areas of rural Tanzania. I will not forget that you travelled 8 hours in a car to collect one data point and slept in a place with no electricity and couldn't even charge your phone. In particular, I would like to thank Joseph Kissiri, my field manager, who worked tirelessly and showed great leadership. And I would like to thank Lusekelo Andrew, who always worked with perfection.

I want to thank people who supported me throughout the program, many of whom are previous and current students in the UCSC department of economics, Shilpa Aggarwal, Valentina Brailovskaya, Jaehoon Choi, Brian Giera, David Sungho Park, Bryan Pratt, and Liam Rose.

Last but not least, I would like to thank my family members. First and foremost, I would like to thank my wife, Jennifer Rhee, for agreeing to come to the US from Korea to build a life with me. Thank you for having a tremendous trust in me. Thank you for engaging in boring discussions with me about my economics research ideas for six years, and even suggesting creative ideas for me. I thank my parents, Intai Jeong and Hyosoon Kim, who put educating me as their utmost priority. Thank you for your love and sacrifices. I am also immensely indebted to my parents-in-law, Jungkeun Rhee and Mikyung Cho, who provided tremendous support for Jennifer and me in every possible way. Finally, I thank my sister and brother-in-law, Dahye Jeong and Sungju Cho, for their support for my education. Thank you for keeping our families united.

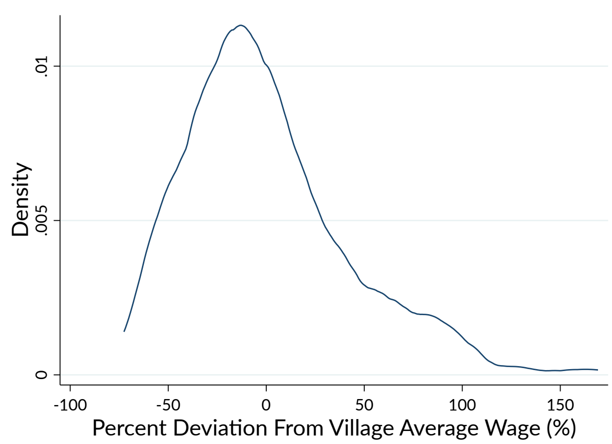
Part 1

Creating Digital Labor Markets in Rural Tanzania

1.1 Introduction

Job search costs are high in rural labor markets in developing countries. Most farmers do not have access to the internet and there are no online job markets for agricultural daily laborers. There is rarely a central place where workers are gathered and employers tend to rehire the same workers whom they had hired in previous seasons. As a consequence, the flow of information on jobs and wages is imperfect, and the law of one price may fail to hold. In my study area of rural Tanzania, I find evidence that there exists substantial within-village wage dispersion. Figure 1.1 presents the percent deviation of an individual daily wage relative to the village average. Only about half the reported wages are within a 25 percent deviation from the village average wage, which is surprising given how jobs are relatively homogeneous for manual farm work.[1]

Figure 1.1: Wage Dispersion In Rural Labor Markets



I study the effect of reducing search frictions on wage dispersion through a field experiment in Tanzania. To reduce job search costs, I develop an SMS-based messaging

[1] This data is collected by the author across 66 villages in rural Tanzania in 2019 and the distribution is at the employer-hiring event level. In my study area, a daily wage ranges from \$1.2 to \$6.5 with a mean of \$3.1.

app that connects agricultural workers and employers. The app was designed to mimic an online job portal like *monster.com*, except that ads are announced over feature phones without internet data. Employers post a job ad to a gateway phone and the ad is sent to all registered workers in the village. Once a worker replies to the job ad, the worker's information is instantly forwarded back to the employer who initiated the request. This service effectively connects all employers and workers at near-zero user cost, increasing the size of labor markets.[2] Treatment villages were offered to use the app service throughout the 2019 agricultural season, while control villages received nothing.

I find that the app has a sizeable effect. It reduces within-village wage dispersion by 16-40 percent (depending on how dispersion is measured).[3] The results are robust to controlling for wage seasonality as well as job characteristics. Labor market search theory predicts that lower search cost may raise wage by creating more jobs (Pissarides 2000; Van den Berg and Van Vuuren 2010). However, while there is a clear reduction in the wage dispersion, I find no effect on job creation or the average wage.

To reconcile the wage compression result along with the null effect on the average wage, I look at heterogeneity in the initial level of wage paid by employers. I first identify the employers who paid a higher wage relative to other employers in the same village before the intervention. I find that the treatment induces initially high-paying employers to reduce the wage. On the other hand, lower-paying employers increase the wage. These competing effects cancel each other out, resulting in little change to the average wage.

A primary effect of the messaging app is that it increases the size of labor markets. By sending a job ad to all registered workers in the village, it is easier for employers

[2] SMS costs are borne by farmers. An SMS voucher costs as little as 22 cents for 1000 messages, while an average daily wage is \$3.1. Additionally, most talk-time vouchers already come with free text messages. From the project side, the annual cost of keeping the messaging app is \$950 (= A subscription fee of a 3rd party platform (\$610) + SMS and mobile data plans (\$72) + operation cost (\$268)).

[3] The measures include standard deviation, coefficient of variation (i.e., the standard deviation divided by the mean), p50-p10 percentile wage ratio, and mean-min wage ratio. The magnitude is in the range of estimates found in the price dispersion literature. For comparison, Aker (2010) finds a 10-16 percent reduction in grain price dispersion and Jensen (2007) finds a 75 percent reduction in fish price dispersion after the introduction of mobile phone service.

to consider a new set of workers whom they had not hired in the past. In addition to integrating previously disconnected employers and workers, the app can also change other aspects of labor market conduct. One feature of the app is that it asks employers (workers) to specify the wage they would like to pay (get paid) to facilitate the transactions. This bidding feature could potentially affect the way participants bargain over wages, for example by encouraging participants to bid more aggressively and effectively. Second, the wage signal in the job ads and job applications might help market participants to update their beliefs on prevailing market wages. If information frictions are prevalent, the size of the update can be large, influencing market wages.

I isolate the channel of search frictions from the change in bargaining behavior and wage signaling by randomizing the disclosure of wage information. In a random subset of treatment villages, I remove the wage information from job ads and/or job applications before sending out a message. In those villages, the app does not carry any explicit wage signal and therefore there is no bargaining effect through the information channel. I find no evidence that the wage disclosure feature has any impact on wage compression. Moreover, I find modest evidence that the overall treatment induces initially high paying employers to increase the probability of hiring a new worker after the intervention. Those employers also face an increased number of applicants per vacancy. Taken together, these findings suggest that the app improves the competitiveness of labor markets, and sharing job availability alone is sufficient to improve the functioning of labor markets (without explicit wage information).

One consequence of wage dispersion is that employers face different prices for labor due to market imperfections, which contributes to the dispersion in labor demand. By reducing wage dispersion, my intervention is predicted to reduce dispersion in labor input for farmers in the same labor market. Consistent with this prediction, dispersion in per-acre labor input across farms is lower in treatment villages by 20-30 percent, suggesting that labor allocation has improved.

My paper is the first study to examine the role of search frictions on wage dispersion in rural labor markets. While there exist many studies on the effect of information communication technologies (ICTs) on price dispersion, they mostly focus on commodity prices (Aker 2010; Aker and Fafchamps 2015; Allen 2014; Goyal 2010; Jensen 2007).[4] More recent studies on labor markets in developing countries use experiments to reduce frictions in urban labor markets by organizing job fairs (Abebe et al. 2017 ; Beam 2016), offering monetary incentives to travel to job sites (Abebe et al. 2019; Franklin 2017; Bryan et al. 2014), and providing information on skills of job-seekers (Abel et al. 2016; Bassi and Nansamba 2018; Groh et al. 2015).[5] A few related studies on rural labor markets show that labor market outcomes improve when villages are connected to outside markets through the construction of roads or footbridges (Aggarwal 2018; Brooks et al. 2019; Shamdasani 2016).

Another contribution of this paper is to document the extent of wage dispersion attributed to search frictions through a well-identified experiment. Early search-theoretic literature shows that wage dispersion is the equilibrium outcome of imperfect wage competition in a market with search frictions (Stigler 1961, Butters 1977, Burdett and Judd 1983, Mortensen 1988, and Burdett and Mortensen 1998). Notably, these papers show that wage dispersion may arise even in settings where workers and employers are identical.[6] However, search models do not agree empirically on the contribution of search frictions to observed wage dispersion. For example, depending on whether on-the-job search and/or sorting are incorporated in the models, some studies show that search frictions explain a large fraction of observed wage dispersion (Postel-Vinay and Robin 2002; Ortego-Marti 2016), while other studies find that search frictions explain little (Hornstein et al. 2011; Bagger and Lentz 2018). While these studies attempt to explain frictional dispersion through models, my paper takes a different approach and experimentally reduces frictions with the messaging app. My findings support the idea that search friction accounts for a sizable variation in wage dispersion.

[4] The relationship between search cost and price dispersion depends on assumptions on search methods and market environments. Counterintuitively, some theories predict that lower search cost increases price dispersion (MacMinn, 1980). See Baye et al. (2006) for an excellent review of search-theoretic models and price dispersion.

[5] See McKenzie (2017) for excellent review of papers on this topic.

[6] On the other hand, Autor (2001) discusses that lower search cost may *increase* wage dispersion within skill groups if worker talents are heterogeneous and the levels of demand for talent are different across markets. The intuition is that heterogeneity in demand for talent across markets could hide the price difference in talent, which is revealed after market integration.

This paper is also related to a recent development literature that focuses on imperfections in rural labor markets (LaFave and Thomas 2016; Foster and Rosenzweig 2017; Dillon et al. 2019). Recent papers explore possible mechanisms of labor market failures. Fink et al. (2018) find that seasonal liquidity constraints distort farm labor allocation because farmers choose to work on other people's farms to cope with food shortage even though returns are lower. Kaur (2019) and Breza et al. (2019) show that downward wage rigidity based on social norms prevents wages from fully adjusting in response to shocks. My paper adds to this literature by documenting another source of rural market inefficiency, i.e., search costs to find workers and jobs, which has not been studied extensively in a rural setting in developing countries.

Section 1.2 explains the messaging app and the intervention protocols. Section 1.3 describes sampling, data, and context of the study. Section 1.4 presents the experimental results. Section 1.5 discusses allocative consequences. Section 1.6 concludes.

1.2 Experimental Design

A key motivation of the intervention in this paper is to make hiring and job search much easier than traditional methods allow. While smart phones are still rare, feature phones shown in Appendix Figure A.1 are almost universal. In the study regions of Tanzania, mobile phone ownership rate is 93 percent. Furthermore, the literacy rate is 84 percent,[7] making an SMS-based messaging app a feasible solution for digital labor markets.

I develop an app which works autonomously on a mobile messaging platform called *Telerivet*, a third party API that was integrated with JavaScript.[8] See Appendix Figure A.2 for an example of the backend development. When farmers send a message to a gateway phone, the system first identifies whether the farmer is a registered user and whether the person is an employer or a worker. To register, an individual responds to

[7] The mobile phone ownership rate is from the author's own farmer survey data collected in 2019. The literacy statistics is computed from National Panel Survey (Wave 4) from National Bureau of Statistics, Tanzania, after restricting the regions to the study regions, Kilimanjaro and Manyara

[8] The JavaScript codes are available publicly on GitHub, https://github.com/regulusweb/ucsc-tz-labor.

six messages one by one to answer basic questions (e.g. name, location, age, gender, and whether the person intends to be an employer or worker).

A registered employer can post a job by answering a few questions about the job, e.g., the type of task, crop, starting day, and the wage (See Appendix Figure A.4 for the full message interactions). Once posted, the ad is sent to registered workers located nearby. The messaging app allows employers to reach a large number of workers instantly, reducing search costs dramatically. A unique job code is attached to a given job ad, and workers can then text back with the job code to apply for the job. Workers' applications are forwarded to the original employer in real time. Both parties are given the phone numbers and names of each other, which they can use to negotiate details over the phone. A week after posting, a feedback survey is sent automatically via SMS to ask about the hiring result, the final wage paid, and the worker ratings.

Another useful feature of the app is the ability to disclose the wage information in a job ad and/or in a job application to facilitate transactions more efficiently. The app asks all users to specify the wage they would like to pay (for employers) and the wage they would like to get paid (for workers). While the bidding feature is intended to reduce transaction costs, it could affect the way people bargain over wages. For example, it might encourage users to bid more aggressively, thereby changing the wage. It also helps users to update the distribution of wage offers in the market. I isolate these two channels from the reduction in search cost, by randomizing the disclosure of wage information. The wage information could be displayed either in the job ad or in the job application. I cross-randomized the non-disclosure of wage information as shown in Table 1.1 and the example messages are in Appendix Table A.1. Overall balance between control and treatment group at baseline is shown in Appendix Table A.2 at a farmer level.[9] Another balance table using the recall data from the phone survey and endline survey for the pre-period is shown in Appendix Table A.3 at a village level. At

[9] While the randomization selected 30 control villages and 40 treatment villages, some villages are excluded from the analysis because not enough farmers reported hiring. In particular, I require at least three reported wage observations within a village-production stage when calculating dispersion measures. Therefore, the analysis involved 30 control villages and 36 treatment villages. To be consistent with the analysis sample, I report the balance table for the same set of villages.

both farmer and village level, almost all of the differences between the treatment and control group are insignificant.

Table 1.1: Treatment Design

	Show Employer's Wage in the Job Posting	Show Worker's Wage in the Application	Number of Study Villages
Treatment	No	No	10
	Yes	No	10
	No	Yes	10
	Yes	Yes	10
Control			30
Total			70

The intervention was rolled out in February and March 2019, as shown in the timeline in Appendix Figure A.6. The village meeting was pre-announced before our visit and everyone interested in hiring and working was invited. They were also told to bring mobile phones. During the intervention period, field enumerators visited each treatment village to conduct two meetings. The first meeting was in the center of the village and the second was in a more remote part of the village. During the meetings, field enumerators conducted a hands-on training session to demonstrate how the messaging app works.

At the meeting, potential employers and workers were given instructions on how to register, which they did by texting *SAJILI* ("register"). Some had no prior experience using SMS messages; thus, enumerators walked them through basic functions of sending and replying to messages on their feature phones. Once everyone was registered, employers were instructed to send a text *WAFANYAKAZI* ("workers") to post a job. The job ads were automatically sent to the workers sitting in the crowd, and field officers guided workers to apply to those jobs by helping them to send a text *KAZI* ("job") During the training session, one randomly selected person per village was given a mobile phone as a gift to incentivize the practice of the messaging app. See Appendix Figure A.7 for an example of a village meeting and Appendix Figures A.8 and A.11 for the flyers distributed which also contain the step-by-step message flows.[10]

The users were sent reminder messages to encourage the use of the app, approximately five times throughout the agricultural season. Furthermore, to incentivize farmers to continue using the messaging app, one farmer for each village was randomly selected to win a $10 or the equivalent. The raffle was done three times throughout the 2019 agricultural season.[11]

1.3 Sampling, Context, and Empirical Specification

1.3.1 Sampling and Data

The study was conducted in two northern regions of Tanzania, Kilimanjaro and Manyara. I draw on the sample of study farmers from a related project (Aggarwal et al., 2019) where we had obtained a census of households from village offices, and had randomly sampled 18 farmers for each village. In the original study, 147 villages were randomly selected after stratifying by market in the two regions.

To study rural labor markets, I excluded villages located in Moshi town, the major hub in Kilimanjaro Region. I also excluded villages with mobile ownership less than 80 percent because the intervention relies on mobile phone technology. After removing pilot villages, I randomly sampled 70 villages out of 86 villages to be included in the study. Treatment was randomized at the village level, resulting in 30 control villages and 40 treatment villages. Within each village, those farmers who did not participate in rural labor markets at the time of the baseline survey were excluded from the study. The final study sample comprises 650 farmers from 70 villages. However, when constructing wage dispersion measures, I require at least three reported wages within a village-production stage. This drops four villages because few employers report hiring in those villages. The map of the study villages is shown in Figure A.5. The average distance from a village centroid to any other nearest study village centroid is 15km, and the closest control-treatment pair is 3km apart by geodetic distance. Given that rural labor markets are formed closely within the village boundary, spillover effects between control and

[10] The original flyers were distributed in the local language, Swahili.

[11] Farmers were eligible to enter the raffle conditional on using the app.

treatment villages are extremely unlikely. Furthermore, the registration for the app was declined if the user is not from a treatment village.

Note that the farmers who were treated by the intervention and the farmers who were part of the survey data collection are not entirely the same. *Everyone* in the treatment villages was invited to the village meetings and was eligible to use the messaging app. However, to ensure the comparability between the treatment and the control group, the random sample of 650 farmers was independently selected as explained above. Only those farmers were surveyed by phone and in-person interviews. The village meetings were followed by three rounds of phone surveys in April-July 2019 and endline surveys in September 2019. The compliance rates for each round of phone survey and endline survey are presented in Appendix Table A.4 and A.5. The phone survey compliance rates are slightly lower in treatment villages (59 percent vs. 62 percent in control), but the difference is statistically insignificant. The compliance rates for in-person endline interviews are quite similar (91 percent vs 90 percent).

The universe of hiring history from 2018 is constructed by merging the phone survey and the endline interviews. If a farmer was successfully surveyed on the phone, the hiring events that occurred after the phone survey are supplemented by the endline interview data. On the other hand, if a farmer was not reachable by phone, all hiring events data solely come from the endline interview data. The merged hiring events from the three rounds of phone survey and the endline survey form the basis of wage dispersion analysis. I also use reported wages by employers only. The data cover all hiring events, 1,867 events by 448 employers from 2018 to 2019 September.[12]

1.3.2 Context

In Tanzania, the North-Eastern regions including Kilimanjaro and a small part of Manyara region have two farming seasons annually: a longer, more productive "long rains" season, which runs from March to August, and a less productive "short rains" season

[12] While the messaging app also collected wage information through the system-generated follow-up texts, this data is not used for analysis. There is no comparison data in the control group because the access to the app was limited to treatment villages.

from October to February.[13] The intervention was conducted right before the planting season of 2019 long rains. Panel A of Table 1.2 shows that rural villages are not small geographically. According to the 2012 Population and Housing Census of Tanzania, a typical village has 532 households. In the farmer surveys collected in 2019, farmers report that it takes on average 4 hours to walk from one end to the other end of the village. They also estimated that it will take 42 hours if they were to visit every single household in the village.

Table 1.2: Village Size and Dispersion In Labor Input

	Mean	SD
A. Village Size (Village-Level)^a		
Number of Households in The Village (2012 Census)	532.48	373.03
Estimated Hours to Pass Through The Village	3.95	1.58
Estimated Hours to Visit Every Household	41.69	33.93
B. Labor Input (Household Level)^b		
On-Farm Labor Days	84.09	68.86
On-Farm Family Days	49.55	48.23
On-Farm Hired Labor Days	24.36	33.67
On-Farm Exchange Labor Days	3.26	8.00
Labor Per Acre	34.88	29.35

Notes: *a*.Village-Level statistics are from 66 villages, and are computed using the median value across farmers within village. *b*.Household-level statistics are based on 566 farmers in 66 villages who participated in the endline survey and cultivated in 2018 long rains. Labor input statistics are conditional on cultivating.

Farming in Tanzania is small-scale and labor intensive and most production is for subsistence. A median plot size is two acres, and most people plant a combination of maize and beans. The average value of production was only \$246 in 2018 and \$141 in 2019.[14] The low productivity is in part due to low adoption of input technologies. In the study sample, only 20 percent of farmers used fertilizer and 50 percent used hybrid seeds in 2018 long rains. Lacking access to credit and farm machineries, the most important input to agricultural production for most farmers is manual labor. Panel B of Table 1.2

13 See the agricultural cycle in Appendix Figure A.10 during 2018 and 2019

14 Low production in 2019 is potentially due to low rainfall. Ninety percent of farmers said the rain in 2019 was lower than the typical rainfall, and 23 percent of them said it was the worst rain they had seen in their life.

shows that the typical labor input amount in 2018 long rains is 84 labor days. Much of this is own family labor - about 60 percent of the total labor usage. However, casual workers also account for a large part of the labor force. About 30 percent of total labor input is provided by hired casual workers, while only 4 percent is covered by exchange labor scheme between fellow farmers. Note that there is a large variation in labor input amount across households. For example, the average labor days per acre is 35, while the standard deviation is 30.

Table 1.3 presents more detailed statistics on rural labor markets. As shown in Panel A, a large fraction of households participate in labor markets. Roughly 50 percent of farmers hired casual workers in the 2017 long rains, while 35 percent of households reported working as a casual worker. A small proportion simultaneously bought and sold labor (6 percent).

The remaining statistics show responses from employers in the farmer surveys. Conditional on hiring, farmers typically hire two times during the season. As expected, job durations are short, typically lasting 2-3 days. In a given hiring event, seven workers are typically hired at a time. Another feature of rural labor markets is that employers rarely hire workers outside the village. Only 8 percent of workers are hired outside the employer's own village.[15] Eighty-one percent of workers are those whom the employer had already hired in previous seasons.

Employers in the village rely on traditional search methods. Panel B of the table shows that 62 percent of employers reported contacting workers whom they already know. However, an equally large fraction of employers reported being visited by workers, while 44 percent of them reported that workers called them first asking for a job. There is rarely a central place in rural areas where workers are gathered, a fact reported by only 10 percent of farmers. Furthermore, word of mouth is not a popular way to find workers, either. Only 9 percent of employers reported asking village leaders, friends, and families. This suggests that employers and workers try to find each other directly. Panel C of the same table shows that a daily wage is 3 USD. To measure the total wage

[15] An unreported survey response indicates that it is largely due to transportation cost rather than preferential treatment over own villagers.

Table 1.3: Rural Labor Markets

	Mean	SD
A. Labor Market Participation and Characteristics		
=1 if Employer[a]	0.49	0.50
=1 if Worker[a]	0.35	0.48
=1 if Both Employer And Worker[a]	0.06	0.24
Worker-Employer Ratio[b]	0.87	0.94
Number of hiring events	1.80	0.97
Job duration in days	2.73	2.39
Number of workers hired	6.81	4.58
Fraction of workers the employer hired outside the village	0.08	0.22
Fraction of workers the employer had hired previously	0.81	0.27
B. Search Methods By Employers		
=1 if I called or/and visited workers I know	0.62	0.49
=1 if Workers visited me asking for a job	0.66	0.48
=1 if Workers called me asking for a job	0.44	0.50
=1 if I went to a gathering place	0.10	0.30
=1 if I asked leaders/friends/families	0.09	0.28
C. Wage Compensation		
Daily Raw Wage Per Person (USD)	3.12	1.18
=1 if Paid Workers for Food	0.26	0.44
=1 if Paid Workers for Transportation	0.03	0.17
Daily Raw Wage And Benefits Per Person (USD)	3.24	1.20

Notes: *a*. Summary statistics are from the baseline survey in 2017 long rains from 566 farmers. *b*. The number of employers-workers ratio is at the village level. The remaining statistics are conditional on hiring in 2018 long rains, from 352 farmers in 66 villages.

compensation precisely, I also asked employers if they paid other benefits. 26 percent of employers reported paying for food on top of the wage and 3 percent of them paid for transportation. In the analysis section below, I show most of results for both raw wage as well as the total wage compensation which include food and transportation payments.

1.3.3 Empirical Specification

The main outcome of the study is wage dispersion. Wage dispersion is measured within village because the intervention treats everyone in the labor market at the village level. I use commonly used measures of dispersion in the literature: standard deviation, coefficient of variation, mean-minimum wage ratio, and p50-p10 percentile ratio. Because of seasonality in the agricultural production (See Panel B of Appendix Figure A.10), the wage in lean season and the wage in peak season reflect different labor market conditions. Therefore, I divide agricultural production stages as follows: 2018 planting, weeding, harvesting, dry season, 2019 planting, weeding, and harvesting. The stages are chosen by the most popular job tasks reported in the farmer surveys in a given time period. I construct dispersion measures within village and agricultural production stage using wages reported by employers only. The main regression uses a simple difference-in-differences specification where the source of exogenous variation is the randomization of the treatment:

$$Dispersion_{vs} = \beta_0 + \beta_1 TREAT_v + \beta_2 TREAT_v \times Post_s + \delta_s + \varepsilon_{vs}. \tag{1}$$

$Post_s$ is a dummy variable to indicate whether a production stage s is after the intervention,[16] δ_s is a stage fixed effect, and $TREAT_v$ is a dummy variable indicating that village v is in the treatment group. β_2 estimates the effect of the intervention on wage dispersion, by differencing out pre-post difference as well as control-treatment difference.

While the main specification relies on the aggregated village level data, the raw wage

[16] Village meetings took place for a month. I use a uniform meeting date across treatment and control villages to define $Post_s$. The results are robust to using the median village meeting date as well as the last day of all village meetings.

data is collected at the individual-hiring event level. As a robustness check, I estimate the wage dispersion at a farmer-event level as well. I construct the absolute percent deviation of the individual wage from the village-stage average wage as follows:

$$abs\left(\frac{w_{ivsh} - \bar{w}_{vs}}{\bar{w}_{vs}}\right) = \beta_0 + \beta_1 TREAT_v + \beta_2 TREAT_v \times Post_s + \phi_{C(i,s,h)} + \eta_{J(i,s,h)} + \delta_s + \varepsilon_{ivsh}, \tag{2}$$

where w_{ivsh} is a wage paid by an employer i in village v in the production stage s for a hiring event index h, and $\bar{w}_{vs}$ is the average wage paid within village v and stage s. The $\{\phi_c\}_{c=1}^{C}$ are crop-specific effects on wage and C(i,s,h) is a function indicating the crop of the hiring event index h of an employer i in stage s. Similarly, the $\{\eta_j\}_{j=1}^{J}$ are task-specific effects of the hiring event, and δ_s is a stage fixed effect.

As noted in Table 1.1, the disclosure of wage information was randomized within treatment villages. A regression specification that tests whether wage dispersion is influenced by the additional bidding feature is:

$$\begin{aligned} Dispersion_{vs} = \beta_0 &+ Post_s \times (\gamma_1 TREAT_v + \gamma_2 TREAT\_BID_v) \\ &+ \gamma_3 TREAT_v + \gamma_4 TREAT\_BID_v + \delta_s + \varepsilon_{vs}, \end{aligned} \tag{3}$$

where

$$TREAT\_BID_v = \begin{cases} 0 & \text{if village } v \text{ is in the control group or the wage is not disclosed} \\ 1 & \text{if village } v \text{ is in the treatment group and the wage is disclosed} \end{cases}$$

In this equation, the coefficient γ_2 captures the additional effect of the bidding feature on wage dispersion.

One caveat of the wage dispersion analysis is that it relies on the wages reported by employers. The results would suffer from endogenous selection of farmers if the treatment causes some farmers to become employers. Therefore, I explore the treatment effect on hiring and other labor input measures using farmer-crop season level, which include all study farmers regardless of whether they participate in the labor market:

$$Y_{ivr} = \beta_0 + \beta_1 TREAT_v + \beta_2 TREAT_v \times Post_r + \rho_r + \varepsilon_{ivr}, \tag{4}$$

where Y_{ivr} measures various outcomes at the crop season level including whether a farmer

i in village v hired workers during agricultural season r. ρ_r is a season fixed effect. Farmers were asked questions for three rainy seasons, 2018 long rains, 2018 short rains, and 2019 long rains. Some farmers cultivated in all three seasons, while other farmers cultivated in only one or two seasons.

1.4 Results

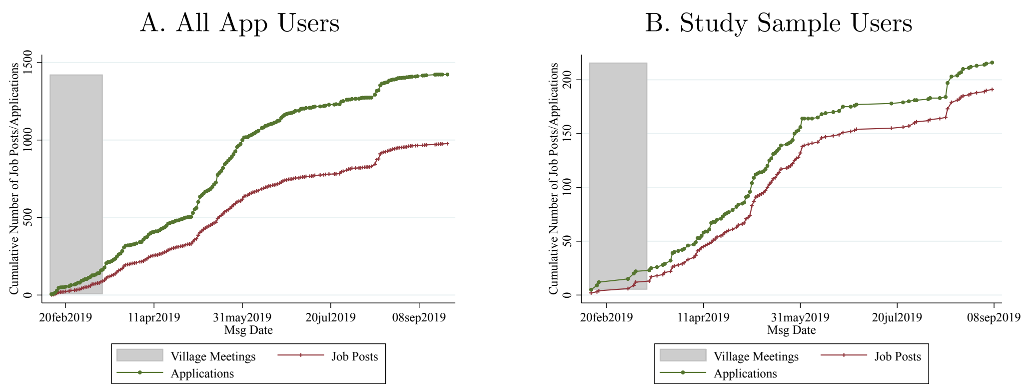
1.4.1 Take-up

The message app has been used extensively by users since its adoption in early 2019. Figure 1.2 shows the number of job posts and applications for all users as well as for the study farmers in each panel, respectively. Almost 1,000 jobs have been announced by 250 employers during 2019 crop season. Job ads were sent to more than 1,000 unique workers during this period. 640 workers sent back almost 1,500 job applications. The large and persistent usage suggests that this technology is simple to use and useful for farmers. The results are noteworthy given that labor demand was quite low in 2019 season due to low rainfall.

I report the take-up of the messaging app for the village meeting sample as well as the study sample. Table 1.4 shows that on average 64 farmers attended the meeting, among them 29 employers and 35 workers. Among those who came to the meeting, 69 percent of them registered for the service. 33 percent of them posted a job as an employer or applied to a job as a worker through the app.

Among the randomly selected study sample, the administrative data from the app database indicates that 39 percent of treatment farmers registered for the service, while 16 percent used the app to find workers or jobs. According to the self-report in the endline survey, 69 percent of treatment farmers heard about the messaging app while 14 percent reported using it. While the usage among study farmers appear modest, the treatment intervention was at the village level and hence the study farmers can be affected without using the app directly.

Figure 1.2: Job Posts and Applications



Notes: Almost 1,000 jobs have been announced by 250 employers. And the job ads were sent to more than 1,000 workers during this period. Among those, 640 workers sent almost 1,500 job applications as of the end of September 2019.

Table 1.4: Take-up of The Search App

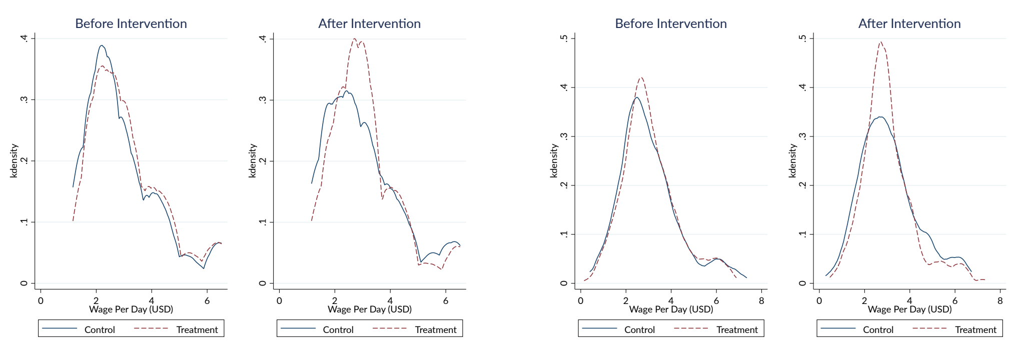
	Mean	SD	N
A. Village Meeting Sample			
Meeting Turnout	64.17	44.81	40
Meeting Turnout: Employer	28.60	23.28	40
Meeting Turnout: Worker	35.58	28.41	40
Proportion Registered	0.69	0.41	40
Proportion Used	0.33	0.24	40
B. Study Sample			
Proportion Registered (Admin Data)	0.39	0.49	370
Proportion Used (Admin Data)	0.16	0.37	370
Proportion Heard About The App (Self-Report)	0.69	0.47	324
Proportion Used (Self-Report)	0.14	0.34	324

1.4.2 Wage Dispersion And Search Costs

Before I formally analyze wage dispersion using regression analysis, I first examine wage dispersion visually in Figure 1.3. Using the daily wage reported by employers, the wage distributions before the intervention in Panel A are quite similar between control and treatment group. However, after the intervention, the wage distribution of treatment villages is more compressed than the distribution of control villages. I further regress the raw wage on crop, task, production stage, and village fixed effects, and plot the distribution of the residualized wage in Panel B. The same pattern is observed – the two distributions are similar for control and treatment villages before intervention, but the wage distribution of treatment group is less dispersed after the intervention. As shown in Appendix Table A.6, I cannot reject that the variances of the two distributions are equal in pre-period, while this hypothesis is rejected in post-period.

Figure 1.3: Reduction in Wage Dispersion

A. Raw Wage B. Residualized Wage



Notes: Raw wage is the salary paid to workers. Residualized wage is the residual from the regression of raw wage on crop, task, production stage, and village fixed effects. The residuals are plotted after centering at the median of the raw wage.

Table 1.5 confirms the wage compression by estimating Equation (1). Column 1-2 explore the standard deviation in wage, while columns 3-4 explore the standard deviation in residualized wage, which is the residual from the regression of raw wage on crop, task, season, production stage, and village fixed effects. Odd columns use raw wage to calculate the standard deviation and even columns use the combined wage and benefits, which is the sum of the wage and food and transportation payments. The result shows that there is a large and significant reduction in the standard deviation. As explained in section 1.3, I require at least three reported wages within a village-production stage when calculating standard deviation and coefficient variation, and this drops four villages. To avoid this exclusion, I collapse all pre- and post-treatment time periods and run a Analysis of Covariance (ANCOVA) estimation as in McKenzie (2012). Appendix Table A.7 shows that the results are qualitatively similar.

Table 1.6 explores the wage dispersion using other dispersion measures, as the standard deviation can be sensitive to outliers. Using the coefficient of variation, p50-p10 percentile wage ratio, and mean-minimum wage ratio, the treatment villages experience 16-30 percent reduction in wage dispersion. In all regressions reported in Table 1.5 and 1.6, wage is winsorized at p5 and p95 to make sure that results are not driven by a few outliers.

Table 1.5: Reduction in Wage Dispersion (Village-Stage Level)

	SD in Wage		SD in Residualized Wage	
	(1)	(2)	(3)	(4)
	Wage	Wage&Benefits	Wage	Wage&Benefits
TREAT	0.113	0.0937	0.0999	0.0787
	(0.111)	(0.105)	(0.104)	(0.0984)
TREAT × Post	-0.467***	-0.435***	-0.458***	-0.426***
	(0.170)	(0.162)	(0.166)	(0.160)
Stage FE	X	X	X	X
Observations	268	268	268	268
Villages	66	66	66	66
Control Mean	1.133	1.151	1.123	1.147

Notes: Raw wage is the salary paid to workers. Wage and benefits include wage payment, food, and transportation payment. Residualized wage is the residual from the regression of raw wage on crop, task, season, production stage, and village fixed effects. Standard errors clustered at the village level.

While the analysis at the village level is intuitive, the raw wage is collected at the hiring event level. Table 1.7 further confirms the wage compression result by estimating Equation (2). In columns, I control for production stage, job task type, and crop type fixed effects. I also report results with and without winsorization of wages. Across specifications in columns 1-4, the treatment villages have a lower wage dispersion than

Table 1.6: Reduction in Wage Dispersion - Various Measures of Dispersion

	Coefficient of Variation		p50-p10 Ratio		Mean-Min Ratio	
	(1)	(2)	(3)	(4)	(5)	(6)
	Wage	Wage&Benefits	Wage	Wage&Benefits	Wage	Wage&Benefits
TREAT	0.00849	0.0100	0.0704	0.0565	0.0643	0.0769
	(0.0319)	(0.0297)	(0.0957)	(0.0922)	(0.0921)	(0.0878)
TREAT × Post	-0.126**	-0.121***	-0.260**	-0.260**	-0.333***	-0.352***
	(0.0479)	(0.0440)	(0.126)	(0.124)	(0.117)	(0.114)
Stage FE	X	X	X	X	X	X
Observations	268	268	269	269	269	269
Villages	66	66	66	66	66	66
Control Mean	0.379	0.369	1.627	1.621	1.788	1.753

Notes: Raw wage is the salary paid to workers. Wage and benefits include wage payment, food and transportation payment. All numbers are in USD. Standard errors clustered at the village level.

the control villages roughly by 20 percent.[17] In columns 5-6 of the same table, I report the treatment effect on wage. Interestingly, the average wage in level in the treatment villages is not significantly different from the average wage in the control villages. Not only insignificant, but also the magnitude of the wage change is quite small (i.e., 4 percent of the control mean wage before the intervention).

Table 1.7: Reduction in Wage Dispersion (Farmer-Hiring Event Level)

	No Winsorization		Winsorized at p5 and p95		Wage In Level	
	(1)	(2)	(3)	(4)	(5)	(6)
	Wage	Wage&Benefits	Wage	Wage&Benefits	Wage	Wage&Benefits
TREAT	-0.0154	-0.00884	-0.0120	-0.00797	0.202	0.139
	(0.0254)	(0.0255)	(0.0219)	(0.0216)	(0.131)	(0.134)
TREAT × Post	-0.0643**	-0.0715**	-0.0636**	-0.0641**	-0.116	-0.0176
	(0.0298)	(0.0306)	(0.0271)	(0.0268)	(0.160)	(0.168)
Stage FE	X	X	X	X	X	X
Task FE	X	X	X	X	X	X
Crop FE	X	X	X	X	X	X
Observations	1613	1613	1613	1613	1613	1613
Farmers	439	439	439	439	439	439
Villages	66	66	66	66	66	66
Control Mean	0.323	0.319	0.300	0.293	2.946	3.095

Notes: This is regression at a farmer-hiring event level data. Outcomes are individual percent deviations from the village mean wage and/or benefits in USD. The results are robust to using the deviation from the village median wage as opposed to village average wage. Standard errors clustered at the village level.

[17] The results are similar with and without fixed effects.

1.4.3 Mechanism: Heterogeneous Effects Across Employers

To reconcile the wage compression result along with the null effect on average wage, I explore the characteristics of farmers who are affected differently by the treatment. I first look at the treatment effect by the initial wage paid by employers. If search frictions are symmetric between workers and employers, search theory predicts that initially high-paying employers reduce the wage, while initially low-paying employers raise the wage. Using the pre-period wage data, I categorize employers into terciles: initially low-paying employers, medium-paying, and high-paying. Because labor markets are at the village level, I define an individual percent wage deviation from the village average wage to determine the categories using pre-intervention wage data.

Column 1 in Table 1.8 presents the results. It shows that initially high-paying employers reduce the wage significantly relative to initially medium-paying employers (in comparison to the control group). On the other hand, I do not find evidence that an increase in wage by initially low-paying employers is significantly different from the medium-paying employers. I also standardize the initial individual wage deviation to explore the result in a continuous fashion. Column 2 suggests that one standard deviation increase in initial wage deviation is associated with a 12 cent reduction in wage only in the treatment group, although insignificant.[18] The asymmetric result found in the table implies that the average wage level does not change because initially low- and medium-paying employers together raise the wage while initially high-paying employers reduce the wage.

The app is designed to promote a competitive market environment by integrating fragmented markets. Table 1.9 further explores if the mechanism of wage compression is due to increased competition. I use three proxies of competition: (i) whether an employer hires a new worker whom the employer had not hired previously in a hiring event, (ii) the fraction of new workers from the employer's labor force, and (iii) the ratio of the number of applicants to the number of hired workers. By reducing search frictions, employers are more likely to consider new workers outside of an exisiting network and

[18] Note that I lose a few farmers in columns 1 and 2 because they did not report hiring pre-period and their initial wage level is undefined.

Table 1.8: Heterogeneous Effects on Wage

	Dep.Variable: Wage In Level	
	(1)	(2)
TREAT × Post	0.317*	-0.0669
	(0.185)	(0.142)
TREAT × Post × Pre-Period Wage: Low	-0.288	
	(0.314)	
TREAT × Post × Pre-Period Wage: High	-0.721**	
	(0.287)	
TREAT × Post × Pre-Period Std(Wage Deviation)		-0.122
		(0.163)
Observations	1567	1567
Farmers	409	409
Villages	64	64
Control Mean	2.946	2.946

Notes: Some farmers are dropped in columns 1 and 2 because they did not report hiring pre-intervention period. Initial low vs. high wage is defined at the individual farmer level. All specifications include crop, task, and production stage fixed effects as well as the full interaction variables on the triple differences. All numbers are in USD. Standard errors clustered at the village level.

face higher job competition measured by the number of applications per vacancy. While some coefficients suffer from a lack of power, I find modest evidence that initially high-paying employers are more likely to hire a new worker with an increased fraction of new workers in the labor force. Moreover, they experience an increased job competition, consistent with the reduction in wage in Table 1.8. Overall, the intervention seems to have improved competitiveness of the labor markets.

1.4.4 The Effect of The Wage Bidding

One feature of the messaging app is the ability to disclose wage information which may change the bargaining behaviors as well as the belief on the distribution of wage offers. To isolate these channels from the reduction in search cost, the wage disclosure was randomized in a subset of treatment villages. I explore the difference in the treatment effect between villages with and without wage disclosures in Table 1.10 by estimating regression equation (3).

In the table, the coefficient of $TREAT\_BID \times Post$ measures the additional treatment effect of the bidding feature relative to the villages where the wage information was not

Table 1.9: Heterogeneous Effects on Labor Supply

	𝟙(Hired a New Worker)		New Worker Ratio		Job Competition	
	(1)	(2)	(3)	(4)	(5)	(6)
TREAT × Post × Pre-Period Wage: Low	-0.0703		-0.00184		0.124	
	(0.0886)		(0.0475)		(0.163)	
TREAT × Post × Pre-Period Wage: High	0.133		0.0781		0.277*	
	(0.116)		(0.0526)		(0.163)	
TREAT × Post × Pre-Period Std(Wage Deviation)		0.0896*		0.0300*		0.0965
		(0.0455)		(0.0173)		(0.0717)
Stage FE	X	X	X	X	X	X
Crop FE	X	X	X	X	X	X
Task FE	X	X	X	X	X	X
Observations	1567	1567	1567	1567	1567	1567
Farmers	409	409	409	409	409	409
Villages	64	64	64	64	64	64
Control Mean	0.395	0.395	0.161	0.161	1.595	1.595

Notes: Some farmers and villages are dropped because they did not report hiring pre-intervention period. 𝟙(Hired a New Worker) is a dummy that indicates whether a hiring event included a new worker that the employer did not hire previously. New Worker Ratio is the number of new workers divided by the number of hired workers in a hiring event. Job Competition is the number of applicants over the number of hired workers. All specifications include crop, task, and production stage fixed effects as well as the full interaction variables on the tripple differences. Standard errors clustered at the village level.

Table 1.10: The Effect of The Bidding Feature

	SD in Wage		Coefficient of Variation		p50-p10 Ratio		Mean-Min Ratio	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Wage	Wage&B	Wage	Wage&B	Wage	Wage&B	Wage	Wage&B
TREAT	0.147	0.131	0.00490	0.00969	0.0320	0.0171	-0.0347	-0.0121
	(0.176)	(0.165)	(0.0441)	(0.0427)	(0.149)	(0.154)	(0.129)	(0.132)
TREAT_BID	-0.0456	-0.0501	0.00481	0.000422	0.0515	0.0529	0.133	0.119
	(0.182)	(0.174)	(0.0451)	(0.0444)	(0.155)	(0.158)	(0.133)	(0.136)
TREAT × Post	-0.494**	-0.473**	-0.138**	-0.129**	-0.158	-0.134	-0.258**	-0.266*
	(0.225)	(0.217)	(0.0594)	(0.0558)	(0.159)	(0.166)	(0.129)	(0.134)
TREAT_BID × Post	0.0365	0.0514	0.0131	0.00982	-0.129	-0.158	-0.104	-0.116
	(0.198)	(0.195)	(0.0502)	(0.0485)	(0.155)	(0.161)	(0.121)	(0.134)
Stage FE	X	X	X	X	X	X	X	X
Observations	268	268	268	268	269	269	269	269
Villages	66	66	66	66	66	66	66	66
Control Mean	1.159	1.175	0.387	0.374	1.635	1.626	1.792	1.758

Notes: Raw wage is the salary paid to workers. Wage and benefits include wage payment, food, and transportation payment. Standard errors clustered at the village level.

disclosed. The evidence suggests that displaying the wage information in the job ad and/or in the worker application does not contribute more to the wage compression. Appendix Table A.8 also shows the results at the farmer-hiring event level. Again the additional bidding feature does not reduce wage dispersion more than the regular treatment group without the bidding feature. Overall, the results seem to suggest that announcing the job availability among a large number of workers alone is sufficient to compress the wage.

1.5 Efficiency of Labor Markets

1.5.1 Labor Allocation

The main wage compression results used wages reported by employers only. While the main results are at the village level, the results might suffer from the endogenous selection of employers if the intervention caused some farmers to become employers or caused employers to hire more workers. Table 1.11 examines the changes in hiring outcomes by estimating Equation (4). All specifications use farmer-crop season level data, including all study farmers regardless of whether the farmers are in the labor market or not. Column 1 shows that treatment farmers are no more likely to become an employer.

In columns 2-5 of Table 1.11, various types of labor input are examined. The results are conditional on cultivating in a given rainy season, and hence have a smaller number of observations than column 1. Overall, the intervention did not seem to have affected the average labor input. Treatment farmers are no more likely to use family labor, hired labor, and exchange labor than control farmers. The total labor input amount is also similar between control and farmers.

While there is no treatment effect on the different types of labor input measures, the wage compression result has allocative implications. Agricultural production theory predicts that the marginal product of labor must be equal across households if markets are complete and prices and productivities are controlled for (Benjamin 1992; LaFave and Thomas 2016; Dillon et al. 2019). However, the existing wage dispersion due to

Table 1.11: Treatment Effects On Labor Allocation

		Types of Labor Input In Person Days			
	(1)	(2)	(3)	(4)	(5)
	=1 if Hired	On-Farm Labor	Family Labor	Hired Labor	Exchange Labor
TREAT	-0.006	-3.770	-6.170	-0.299	1.558***
	(0.039)	(9.991)	(7.864)	(3.018)	(0.585)
TREAT × Post	0.040	4.244	5.427	2.118	-0.620
	(0.057)	(6.243)	(3.906)	(2.624)	(0.689)
Observations	1698	1139	1139	1139	1139
Households	566	555	555	555	555
Villages	66	66	66	66	66
Season FE	X	X	X	X	X
Region FE	X	X	X	X	X
Control Mean	0.36	71.70	45.06	18.79	2.53

Notes: A crop season fixed effect is included. Standard errors clustered at the village level.

market imperfections implies that employers face different prices for labor which contributes to the dispersion in labor input. In other words, if lower search cost reduces market frictions, then it is predicted that the dispersion in labor input also decreases.

I test this prediction in Table 1.12 using the dispersion in log labor days per acre as an outcome. For three out of the four dispersion measures, I find that the labor input dispersion is lower in treatment villages by 17 to 30 percent. The result offers suggestive evidence that the messaging app helps to reduce the misallocation of labor in rural labor markets.

1.5.2 Harvest Output

The improved labor allocation is predicted to increase the aggregate output level in theory. This section explores this downstream effect on harvest output. One challenge of estimating the effect on harvest level is that many farmers reported that their entire crops were wasted due to various shocks including low rainfall, resulting in zero harvest. About 16 percent of farmers indicated that they cultivated and ended up harvesting nothing in a given season. To retain the farmers with zero harvest output, I convert the output using an inverse hyperbolic transformation.

Since some farmers grow multiple crops, the value of each crop is evaluated at the prevailing market price and aggregated across crops to compute the total harvest output

Table 1.12: Dispersion In Labor Input

	Dep. Var: Dispersion in Log($\frac{\text{Labor Days}}{\text{Acre}}$)			
	SD	CV	p50-p10	Mean-min
TREAT	-0.038	-0.014	0.005	-0.068
	(0.064)	(0.034)	(0.140)	(0.160)
TREAT $\times$ Post	-0.134*	-0.077**	-0.360*	-0.202
	(0.075)	(0.035)	(0.181)	(0.252)
Observations	169	169	173	173
Villages	66	66	66	66
Season FE	X	X	X	X
Region FE	X	X	X	X
Dep.Var. Mean	0.81	0.27	1.56	1.69

Notes: A rainy season fixed effect and a region fixed effect is included. Standard errors clustered at the village level.

value. As a robustness check, I also show the result using physical output in kg, given that the large proportion of harvests comprises maize and/or beans only. The regressions also include other controls such as the use of fertilizer or seeds and agricultural shocks. Columns 1-4 in Table 1.13 show inconclusive evidence that treatment villages had more harvest than control villages. While the coefficients are positive, the harvest data is extremely noisy and the treatment effects are not distinguishable from zero.

Another measure of farmer welfare is consumption. Columns 5-6 show that treatment farmers are less likely to skip a meal due to food shortage in the past 3 and 6 months. One explanation might be that it is now easier for workers to find a job and to cope with food shortage with the help of the app. But since the village-level employment did not increase as shown in Table 1.11, it is difficult to conclude whether the increase in consumption is driven by the imbalance at the baseline between treatment and control groups. This outcome is measured at the endline only and therefore the possible baseline difference is not controlled for. Also, recall that two people per village were randomly selected to get 10 USD if they used the app during the 2019 agricultural season. The robustness check controlling for winning a raffle prize of cash \$10 is shown in Appendix Table A.9, and the results are similar.[19]

[19] Note that the random selection includes all users (not just study farmers who were interviewed).

Table 1.13: Treatment Effect on Output Level

	Harvest Output (HH-Season Level)				Skip Meals (HH Level)	
	(1)	(2)	(3)	(4)	(5)	(6)
	Kg	Kg	USD	USD	Past 6m	Past 3m
TREAT	0.468	0.445	0.377	0.367	-0.034*	-0.031*
	(0.316)	(0.287)	(0.274)	(0.252)	(0.020)	(0.018)
TREAT × Post	0.361	0.370	0.196	0.204		
	(0.478)	(0.440)	(0.385)	(0.357)		
Observations	1069	1069	1069	1069	566	566
Households	554	554	554	554	566	566
Villages	66	66	66	66	66	66
Season FE	X	X	X	X		
Input Controls	X	X	X	X		
Shock Controls		X		X		
Control Mean	6.534	6.534	5.187	5.187	0.075	0.051
Control Mean (before IHS)	1001.85	1001.85	224.21	224.21		

Notes: Harvest values are computed by evaluating the crop harvest at market prices and adding them up across crops. Harvest output measures in columns 1-4 are converted using inverse hyperbolic transformation. HH Endowment means the dummies of the number of household members in gender-age bracket (e.g. male from 0-14 years old, male from 15-19 years old, etc). Input controls include the use of fertilizer, seeds, herbicides, insecticides, and irrigation. Shock controls include whether the harvest was affected by low rainfall, flood, crop diseases, insects, birds or animals, thefts, or lack of casual workers. The last row for columns 1-4 shows the mean of the control group before intervention and before inverse hyperbolic transformation. Standard errors clustered at the village level.

1.6 Conclusion

Understanding the sources of inefficiency in labor markets is crucial to improving market outcomes. Labor is the most abundant input factor in rural economies of developing countries. More importantly, misallocation of labor implies that some farmers use too much or too little labor on their farm relative to what is optimal in a frictionless environment. Therefore, simply correcting the misallocation of labor can improve aggregate output without any technological innovations. In this paper, I find evidence that search frictions are a constraint for efficient labor allocation even in tightly connected rural economies. Offering a cost-effective SMS app technology can connect a large number of workers and employers and compress the dispersion of prevailing wages and labor input.

The use of digital technology in African agriculture is becoming increasingly common with hopes to improve agricultural productivity and farmers' welfare. For example, *Hello Tractor* and *Trotro Tractor* connect smallholder farmers with nearby tractor owners using a mobile app so that farmers can hire a tractor even if they cannot afford to own. Additionally, several companies offer frequent updates on weather and market prices and provide tips on farming and financial management via SMS.[20] In particular, *WeFarm* formulates a farmer-to-farmer digital network where farmers can ask and answer questions through text messages, just like an online forum for farmers in developed countries.

The messaging app I developed to create digital rural labor markets is an example of this trend. Feedback survey presented in Appendix Table A.10 suggests that there is enough demand for the app and scope for the profitability of this service. Among the study sample app users, 93 percent indicated that the app service was useful. Most of them reported that they were able to find workers and jobs faster and it required less effort and lower costs. Furthermore, 50 percent of treatment farmers indicated that they plan to use the app in the next season. 74 percent of those who plan to use the app are willing to contribute an average of 1.5 USD per season. A back-of-the-envelope

Among the study sample, six farmers won the phone during the village meeting, and seven farmers won the 10 USD. This is 2 percent of the study sample (13/584 = 0.02).

20 A list of selected companies include *AgroSpaces*, *AgroCenta*, *Farmerline*, *iShamba*, *MFarm*, *Sokopepe*, and *WeFarm*.

calculation implies that the payments from 630 users are enough to cover the cost of the service to make it sustainable.[21] Given that the labor demand was particularly low in the year of the intervention due to low rainfall, it seems that there is a potential for this messaging app to be scaled up to benefit a large number of farmers.

[21] The annual maintenance cost of the app is $950, which can be covered by 630 users if each pays 1.5 USD per season.

Part 2

Market Access, Trade Costs, and Technology Adoption: Evidence from Northern Tanzania

2.1 Introduction

It is widely believed that poor access to markets – due mainly to poor transportation infrastructure – limits agricultural productivity in rural areas of developing countries, by making it harder to access productivity-enhancing inputs like fertilizer and to obtain high prices for harvest output World Bank Group (2007, 2017).[22] However, while remoteness no doubt limits market access and, by extension, input adoption and agricultural productivity, there is little research to quantify its effect.

In this paper, we rigorously document market access for farmers in two regions – Kilimanjaro and Manyara – of Northern Tanzania, which together comprise 6 percent of the land area and population of the country. Our data collection exercise spans the entire supply chain of maize (output) as well as of fertilizer (input) in all 1,183 villages in these two regions, including (1) surveys with a random sample of 2,845 farmers in 246 randomly selected villages; (2) surveys with 532 agro-input retailers ("agrovets"), effectively spanning the universe of input retail locations; (3) a retrospective panel of buying and selling prices of maize from a sample of maize-sellers in each of the 226 markets in the area; (4) surveys with transportation operators which measure road quality, travel times, and travel costs; and (5) driving times and distances from Google Maps API.

We make three main contributions. First, we precisely document spatial price dispersion for input and output prices, inclusive of trade costs. To do this, we use our extensive travel cost data to estimate travel costs to every destination, and then take

[22] Transportation infrastructure is particularly underdeveloped in Africa. The continent has only 137 kilometers of roads per 1000 square kilometers of land area, with only a quarter paved. In contrast, the average for developing countries outside the region is 211 kilometers of roads per 1000 square kilometers, with more than half paved (Foster and Briceño-Garmendia, 2009). For comparison, the US has 679 kilometers per 1000 square kilometers, with nearly 2/3 paved.

the most favorable prices for farmers. We find clear evidence of large and economically meaningful spatial heterogeneity in both input and output prices. For both, we find that the price difference between the 90th and the 10th percentile of delivered input and output prices is equivalent to about 50% of the mean.

Second, we conduct a reduced-form investigation of the correlation between usage and remoteness on the input side, and sales and remoteness on the output side, where the remoteness of any location is proxied by two measures: (a) the population-weighted distance from a set of 5 major urban centers, and (b) the elasticity-weighted trade cost from the same set of hubs (calculated analogously to Donaldson and Hornbeck, 2016). We find that a standard deviation increase in remoteness is associated with a 9-20 percentage point reduction in the probability of using fertilizer and a 4-9 percentage point reduction in the probability of selling maize. These effect sizes are meaningful: input usage in the most remote villages is only a third of that in the least remote villages, while maize sales are only half as high.

While we find clear evidence of reduced market access in more remote villages, and while it is intuitive that this reduced access will affect the choice set and decisions of farmers, it is not possible to quantify these effects in the reduced form alone, since remote villages and villagers may differ from proximate ones in other econometrically unobservable ways not directly related to access to markets. To evaluate the effect of market access on input adoption, our third contribution is to develop a quantitative spatial model of fertilizer adoption, in which the decision to adopt fertilizer is based on local output prices, innate farmer productivity, the distribution of delivered input prices and retailer quality, and idiosyncratic shocks. Transportation costs affect the distribution of prices by increasing the costs for farmers to reach a particular agrovet to buy inputs, as well as costs to reach the local market to sell their harvest.

On the input side, the structure of the model (which is similar to Eaton and Kortum, 2002) facilitates a decomposition of choosing an agrovet into three components: (1) the decision whether to adopt; (2), the decision of which location to buy from; and (3) the decision of which retailer to pick within that location. Farmer surveys record (1) and (2) and thus allow us to calibrate local factors that may affect adoption, as well as the implied trade costs incurred while sourcing from each agrovet location. To estimate trade costs, we derive a structural multinomial logit specification that estimates the implied iceberg trade costs to each location as a function of distance. The results suggest that transportation costs are large: our preferred specification yields estimates of local iceberg costs that are approximately 4% ad-valorem per kilometer of travel, which translates to an average of approximately 30% when buying from the closest agrovet. When comparing this estimate to data collected in our surveys, pecuniary costs make up approximately 43% of this overall travel cost, suggesting that there are significant non-pecuniary costs of travel (which may include the opportunity cost of the time to travel, risk-aversion related to potential stock-outs, or information frictions). After estimating trade costs, we use the model to build a market-clearing condition for fertilizer for each agrovet, which is a function of the expected spatial distribution of fertilizer expenditures by each farmer and the probability that a farmer at each location adopts at a given agrovet. We balance these market clearing conditions by finding a vector of agrovet "amenities" that exactly rationalize the market-shares of each agrovet. After doing so, we are able to calculate a precise measure of market access for fertilizer.[23] The estimated measure of market access for fertilizer falls approximately 50% per standard deviation increase in remoteness.

We use the estimated parameters from the model to simulate market access counterfactuals. For input market access, our primary counterfactual is reducing trade costs incurred to reach retailers by 50%, which is similar to the expected reduction in travel time if roads were upgraded (Casaburi et al. 2013). This policy roughly doubles adoption relative to baseline, and also reduces the remoteness gradient by 39% for a binary measure of using fertilizer, and 59% for total fertilizer expenditures. We also leverage our detailed transport surveys to assess a counterfactual in which transport improvements are targeted toward main roads and rural roads separately. While improving both types of roads increases adoption, it is only the improvement of main roads which reduces the remoteness gradient (surprisingly, improving rural roads only has no effect

[23] This is similar in spirit to Redding and Venables (2004) and Redding and Sturm (2008), Head and Mayer (2011), and Donaldson and Hornbeck (2016).

on the gradient). The reason for this is that there are not many retailers located in remote areas, and so farmers in remote areas typically must travel on main roads to reach a retailer. We also study hypothetical entry counterfactuals, and we find that agrovet entry in remote areas has a larger effect on adoption, but entry into more remote areas is less profitable. Finally, we halve transportation costs for reaching output markets, and find a slightly smaller change in adoption, though a slightly larger reduction in the remoteness gradient (relative to input market access).

This paper sits at the intersection of trade and development economics. On the development side, we contribute to a literature examining why sub-Saharan Africa has lagged behind the rest of the developing world in agricultural technology adoption. Many studies find evidence of large *yield* increases due to using improved inputs (i.e. Duflo et al. 2008; Beaman et al. 2013; Stewart et al. 2005; Udry and Anagol 2006), though the evidence is much more mixed on whether using these inputs is *profitable* (i.e. Duflo et al. 2008; Beaman et al. 2013). Our results quantify the extent to which profitability, and thus adoption, will tend to be lower in more remote locations, due to less favorable input and output prices for farmers. Our work is closely related to Suri (2011), who shows that many Kenyan farmers with high gross returns to hybrid seeds choose not to adopt them because the fixed costs of obtaining seeds are too high, presumably due to travel costs. Our paper is differentiated by focusing on heterogeneity in market access, rather than on heterogeneity in returns. Related work in Minten et al. (2013) also focuses on remoteness and profitability, documenting significant farmer-to-retailer transaction costs to reach price-controlled input cooperatives in a rugged region in northern Ethiopia.

Our paper is related to a growing literature about the effect of transportation infrastructure improvements on development outcomes and on the spatial distribution of economic activity,[24] which includes outcomes other than just prices, such as consumption, farm and human capital investments, migration, and occupational choice. In our

[24] A partial listing of papers includes Aggarwal (2018, 2019), Alder (2019), Adukia et al. (2020), Asher and Novosad (2020), Bird and Straub (2014), Brooks et al. (2019), Bryan and Morten (2019), Gertler et al. (2019), Ghani et al. (2016), Khanna (2019), Morten and Oliveira (2016), Shamdasani (2020), and Storeygard (2016). See Donaldson (2015) for a review.

paper, we focus narrowly on the specific effect of transportation costs on goods market access (i.e. transportation costs and the presence of intermediaries and the prices they charge) in isolation, without changing other margins.[25]

Our work is related to a voluminous trade literature which attributes spatial price differentials to three primary components – marginal trade costs (e.g. Donaldson 2018; Eaton and Kortum 2002; Shiue and Keller 2007; Sotelo 2019), spatially varying mark-ups (Atkin and Donaldson 2015; Asturias et al. 2019), and the organization of intermediaries (Allen and Atkin 2016; Dhingra and Tenreyro 2017; Bergquist and Dinerstein 2019; Casaburi and Reed 2019; Chatterjee 2019). Our work is particularly related to Atkin and Donaldson (2015), who estimate trade costs in a setting where an intermediary buys products at wholesale prices, transports them to distant markets, and sells directly to consumers. In contrast, we are interested in how trade costs affect the decisions of producers (in this case, farmers) regarding buying intermediate goods through their access to retailers and output markets.

Other general equilibrium trade models also assess the link between trade costs, price gaps and technology adoption, though most existing work focuses on trade between larger cities and markets, typically using trucks or trains (Atkin and Donaldson 2015; Porteous (2019, 2020)) whereas our paper focuses on last-mile costs to farmers (which are usually incurred using smaller private vehicles or simply on foot). Consequently, the per-unit travel costs we study are much larger than estimated in prior work. For example, we measure the cost of transport at \$4.74 per ton-km, compared to \$0.29 per ton-km in Porteous, while our estimated ad-valorem travel costs to retailers is 30% over just 6.8 kilometers, whereas Atkin and Donaldson (2015), estimate ad-valorem costs of 10-20% over a distance of approximately 720 km. Our paper is also differentiated by developing a new approach to estimating implied iceberg costs for farmers to reach retailers, as revealed by sourcing decisions that are measured through surveys.

The paper proceeds as follows. Section 2.2 provides background and context on our study region. Section 2.3 explains the data, and documents summary statistics.

[25] Technological advances may make it possible to decouple market access from traditional road infrastructure. For example, Rwanda has a "droneport" already under construction. Drones capable of transporting cargo of up to 20 kilos over a distance of 100 kms already exist.

Section 2.4 presents our main results. We put our findings in the context of a spatial model, which is presented and calibrated in Section 2.5, and is used for running policy counterfactuals in Section 2.6. Section 2.7 concludes.

2.2 Background on Input and Output Markets and Study Regions

This study took place in the Kilimanjaro and Manyara regions[26] of Northern Tanzania. The two regions are a combined 57,000 square-kilometers (6% of the land mass of Tanzania), contain 1,183 villages, and had a population of 3.1 million in 2012 (National Bureau of Statistics, 2013). Compared to developed countries, the quality of roads in Kilimanjaro and Manyara is poor: the paved road density is 2.2% in Kilimanjaro (i.e. 2.2 kilometers of paved roads per 100 square kilometers of area), 0.15% in Manyara, and 0.7% in Tanzania overall (TanRoads and PMO-RALG, 2014), compared to 68% in the US and 134% across the OECD.[27]

The main crop grown in this area is maize. There are two growing seasons in this area: a longer, more productive "long rains" season, from March to June, and a less productive "short rains" season from October to January. Input usage tends to be much higher in the long rains, and some farmers do not plant in the short rains. Our main outcomes are based on behavior in the long rains.

As in much of Sub-Saharan Africa, production capacity of fertilizer is virtually non-existent and almost all of what is used is imported via the port at Dar es Salaam (FAO 2016; Hernandez et al. 2011), and then transported throughout the country over surface roads. In all of these respects, the study area is fairly similar to other countries throughout East Africa that predominantly grow maize and import fertilizer, such as Kenya, and perhaps a little bit better than landlocked countries such as Malawi and Uganda, that can receive fertilizer only after it has traversed the distance between a

[26] Tanzania has 31 regions in all, including 5 in Zanzibar.

[27] Information compiled from various resources. The Roads Act, 2007 (No. 13 of 2007) defines a a trunk road as one that is primarily (i) a national route that links two or more regional headquarters or (ii) an international through route that links regional headquarters and another major or important city or town or major port outside Tanzania. A regional road is a secondary national road that connects (i) a trunk and district or regional headquarters; (ii) a regional headquarters and district headquarters.

neighboring coastal nation's port and their shared border, and then must travel further inland to reach farmers in the destination country. Urea is the most widely used and sold fertilizer, with about a third of the farmers using it and 84% of the agrovets selling it, followed by DAP, which half of the agrovets sell and 10% of the farmers use. For the agrovets in our sample, procurement almost universally happens at wholesalers located in major cities or towns (agrovets that procure from hubs account for about 94% of market revenue), and the vast majority of agrovets (90%) travel to the wholesalers themselves for procurement.

On the output side, farmers can sell to itinerant buyers – "agents" – who visit the farmgate soon after harvest and therefore buy at the low post-harvest prices. Alternatively, a farmer could travel to a market with their maize and find a buyer, or engage in informal sales near their homestead.

2.3 Data and Summary Statistics

We have four main sources of data: agrovet surveys, farmer surveys, transport surveys, and maize price surveys. All were collected from January 2016 to December 2017 in Kilimanjaro and February to May 2018 in Manyara.

2.3.1 Agrovet surveys

We conducted a census of all agricultural input retailers (known as "agrovets" locally) in the two study regions, finding a total of 585 that sold either fertilizer or seeds. We then revisited these agrovets to conduct a longer survey which took about 2 hours to complete. Of the 585, we did surveys with 532 of them (see Appendix Table B.1 for survey compliance and attrition), asking questions about varieties of fertilizer sold, and their prices, quantities, and the wholesale costs of acquiring stock from the distributor. The survey took care to differentiate fertilizer varieties by distributor, brand, and type – thus the level of granularity should be akin to the barcode-level. The survey also included a number of questions about costs of travel to the distributor, as well as some background characteristics about the business and its owner.

2.3.2 Farmer surveys

We conducted farmer surveys in 246 randomly selected villages in three waves. The first wave covered 115 villages in Kilimanjaro in early 2016, the second wave 97 villages in Kilimanjaro in 2017, and the third wave 50 villages in Manyara in 2018. The surveys included questions on input usage and prices, transport costs and agrovet choice, maize sales, harvest output, and household and demographic information. Though the exact questions varied from survey to survey, the general format was similar across rounds. The main difference across rounds was the sampling procedure and the number of farmers enrolled per village: in round 1, households were selected through a random walk procedure[28], while in rounds 2-3 households were pre-identified from a listing exercise conducted with village leaders. In Wave 1, we sampled only 5 households per village for budgetary reasons, while in Waves 2-3 we selected 18 households per village. We find no qualitative difference in results from the two methods, and thus we pool all surveys together in the analysis.[29]

2.3.3 Measuring transport costs

One of the primary contributions of this work is to carefully document transport costs incurred by farmers. We measured transportation costs in several ways. First, we collected the GPS location for every village,[30] from which we calculated driving times and distances using the Google API (via the statistical program R). Second, we conducted surveys of transportation operators in every village in our sample, which were either motorbike taxis ("Boda Bodas"), or consumer van taxis ("Dala Dalas"). In each village, we asked up to 3 operators how much it cost to travel to the major towns in Kilimanjaro (Arusha and Moshi), the capital city (Dar es Salaam), and importantly, the market

[28] In particular, enumerators were instructed to first find a landmark within the village. These landmarks included a primary/secondary school (1st choice), local church (2nd), and boda stand (3rd). Once the landmark was identified, the enumerators randomly picked a direction to begin their fieldwork, and selected every third homestead, or the next homestead after five minutes of walking, whichever came first.

[29] Results disaggregated by survey method are available on request.

[30] We cross-checked these GPS coordinates, and filled in a handful of missing values, using a dataset of postal geocodes from www.geopostcodes.com.

center as defined for the sampling procedure.[31]

Third, enumerators recorded information on road quality and travel times as part of their field work. To get to a market center and village from a major hub, enumerators took the standard routes, which entailed travel for some distance along a major trunk road, and then turning off onto unpaved feeder and village roads. Costs were measured on these routes. To measure travel times, field officers recorded their GPS location at the point at which they turned off the main road, and then recorded the travel time, distance, and road quality on the road to the market center associated with the village. On reaching the market, they took a second mode of transportation to the village, recording again cost, distance, travel time, and road quality. We use this data to correlate costs of travel with road quality, and to estimate the percentage of roads which are paved versus gravel or dirt.

2.3.4 Maize prices

To measure maize prices, we first identified the local market for each village.[32] These markets are typically located some distance from farmers, and market activity occurs on pre-specified days. Enumerators visited these markets in September and October of 2017 in Kilimanjaro and February to May of 2018 in Manyara. During these visits, enumerators sampled up to 3 maize sellers per market and collected pre- and post-harvest selling prices for maize during recent seasons. These data allow us to compare prices across markets at the same point in time, though they are not intended to be used in a panel analysis.

2.3.5 Summary statistics on villages

A map of Kilimanjaro and Manyara is shown in Figure 2.1. Summary statistics on villages are provided in Table 2.1. The average village has 480 households (see table notes), and is located 6.5 kilometers from the nearest market center. It takes about 40

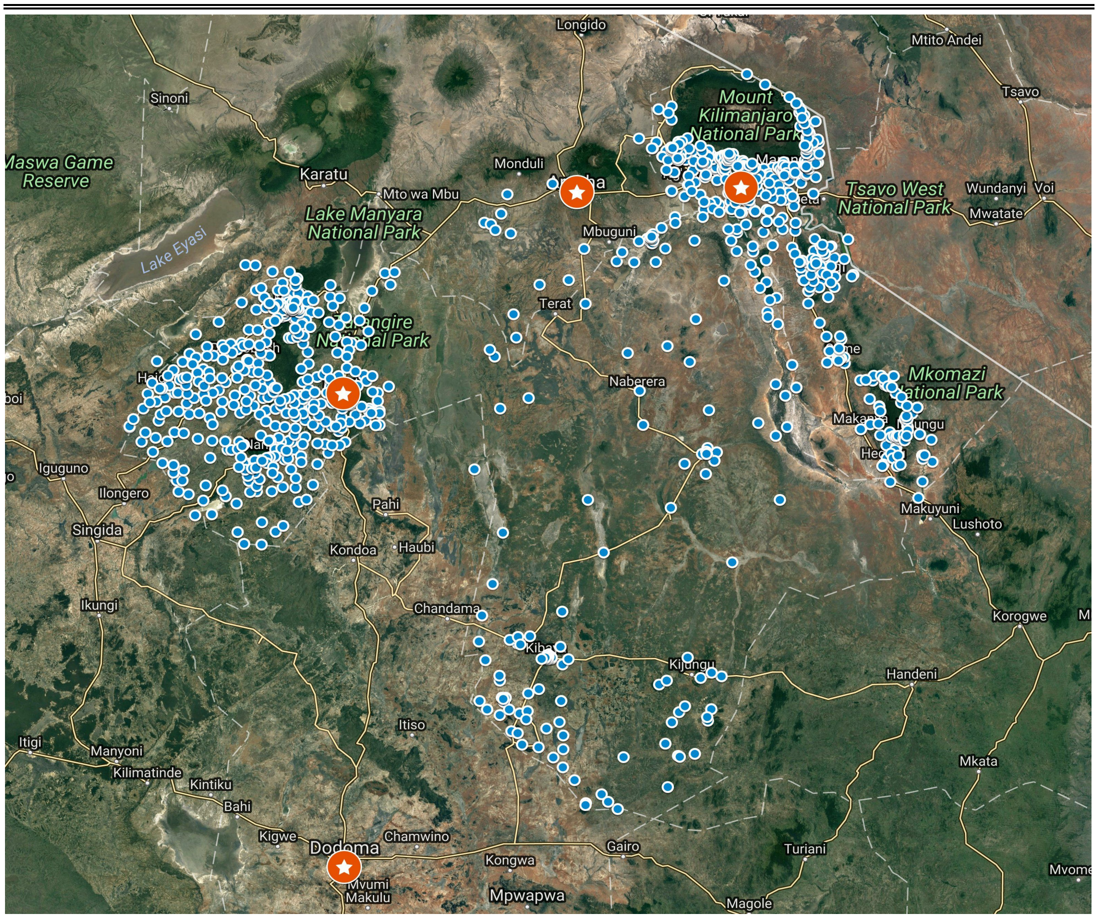
[31] In Manyara, we also asked about trip costs and times to Babati, Dodoma, and Tanga.

[32] This was done by visiting ward offices (the ward is the second-lowest administrative level, just above the village) and asking the ward officer to list the market that people from each village frequented.

minutes of driving to reach the market and return, and a round-trip costs about \$1.90 on average. The average village is over 70 km away from the nearest major hub, and a round-trip to the hub would take about 3 hours and cost \$6.[33] Some villages are extremely remote – the standard deviation of time to a hub is about 2 hours.

Panel B shows information on the quality of the rural roads connecting markets and villages. Roads are about 20% paved, 40% dirt, and 40% gravel, and travel times according to Google are fairly slow: 36.7 km/hour on rural roads compared to 46.1 km/hour on the main roads.

Figure 2.1: Map of Survey Region and Villages



Notes: Blue dots represent all villages in the Kilimanjaro and Manyara Regions. The star signs represent the five major hubs that are used to construct our market access proxies in Section 4.1. They are Moshi, Arusha, Babati, Dodoma, and Tanga.

33 The average income of a farmer from all non-farming and farming sources is about \$610 (Table 2.3).

Table 2.1: Summary Statistics on Villages

	(1) Mean
Panel A. Travel costs to markets and major hub towns	
Distance to nearest market center (km) - Google maps	6.52
	(9.94)
Time for round-trip journey to nearest market center - surveys	40.8
	(39.30)
Cost of round-trip from village to nearest market center (USD) - surveys	1.92
	(2.43)
Cost of round-trip from market center to village (paid by enumerator)	2.53
	(3.14)
Distance to a major hub (km) - Google maps	72.8
	(56.10)
Round-trip travel time to a major hub (mins) - Google maps	171.5
	(115.10)
Round-trip cost of travel to a major hub (USD) - surveys	5.72
	(5.33)
Panel B. Road quality	
Field Measurement of roads from market centers to villages	
Percent of road that is:	
Paved	0.20
Dirt	0.42
Gravel	0.38
Travel speed on feeder roads and rural roads - km/hr (GPS surveys)[1]	21.6
	(11.80)
Google estimates	
Travel speed on feeder roads and rural roads - km/hr (Google)	36.7
	(15.7)
Travel speed on major roads - km/hr (Google)[2]	46.1
	(12.7)

Notes: The average village had approximately 480 households in the 2012 census and ranged in size from 48 to 3241. Table includes 1,168 villages in the Kilimanjaro and Manyara regions of Tanzania. There are 1,183 total villages in the area but several were not visited. Standard deviations in parentheses.

[1]Feeder roads and rural roads are routes from villages to a nearest market.

[2]Major roads are routes from markets to a nearest city.

2.4 Results

In this section, we examine dispersion in input and output prices, and document the relationship between remoteness and prices, adoption, yields, and other related outcomes.

2.4.1 Travel cost-adjusted price dispersion

For each village, we assume that farmers are free to travel to any agrovet/market to buy inputs or sell output, but must incur a transportation cost, which we calibrate using information from transport surveys and Google distances. Specifically, using Google

API, we calculate the route from every village to every agrovet/market. This route will involve either (1) traveling only on local roads over a relatively short distance, or (2) using local roads to connect to trunk roads. We calibrate the costs of local and trunk roads using our transport operator surveys, and information collected by enumerators during their own travel. We present these results in Table 2.2.

Table 2.2: Calibrating Travel Costs

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Main roads (Market centers to hub towns) (Transport Operator Surveys)			Rural roads (villages to market)					
				Enumerator's Trips			Transport Operator Surveys		
	Cost	Cost	Hours	Cost	Cost	Hours	Cost	Cost	Hours
Panel A. Costs from Markets									
Google maps: kilometers to destination	0.02***								
	(0.00)								
Google maps: hours to destination		1.26***	1.00***						
		(0.03)	(0.03)						
Number of markets	201	201	201						
Number of observations	900	900	893						
Panel B. Costs from villages									
Google maps: kilometers to destination				0.12***			0.09***		
				(0.01)			(0.01)		
Google maps: hours to destination					3.54***	0.72***		2.61***	0.84***
					(0.27)	(0.07)		(0.25)	(0.08)
Number of villages				1127	1033	1036	1133	1133	1027
Number of observations				1127	1033	1036	1133	1133	1027

Notes: Data is constructed from interviews with transportation operators, and from travel costs and times incurred by enumerators. There are 226 market centers in our sample. In both regions, transportation operators were asked about the 3 most important hubs (Moshi, Arusha, and Dar es Salaam); in Manyara, they were also asked about 3 additional hubs (Tanga, Dodoma, and Babati). The unit of observation is the market-hub level for Panel A, while it is the village-market pair level for Panel B. Cost is for one-way trip for a given route. Standard errors in parentheses (clustered by market in Panel A).
*, **, and *** indicate significance at 10%, 5%, and 1% respectively.

Columns 1-3 show the costs of traveling from market centers to hub towns, which involves primarily traveling on trunk roads. We find a cost of about \$0.021 per km, or \$1.26 per hour of travel. The remaining columns present figures for rural roads (i.e. villages to market). As expected, we find higher costs for rural travel: \$0.088 per km, or \$2.61 per hour of travel. We use these estimates to calibrate costs from every village to every retailer, depending on whether the route includes only local roads or involves travel along trunk roads as well.[34]

[34] The specific methodology is as follows. For travel from a village to another village using only local roads, we do not have direct survey measures of travel costs, and so we imputed the travel costs using the Google distance between the two villages and the average travel cost per km on a rural road. For non-local travel which involves connecting to a larger trunk road, we calculate costs more directly. In this case, getting from an origin village to a destination involves going from the origin to the trunk road, traveling along the trunk road for some distance, and then turning off the trunk road to travel to the destination location. Our surveys contain a direct measure of the cost of going from any village to the main road (via the market), since we asked transport operators

With these costs, we calculate a travel cost-adjusted price of fertilizer for every village in two ways. First, we define the minimum travel cost-adjusted price that is available to villagers as follows:

$$r_v^{min} = \min_j \{r_j + c_{jv}\} \tag{5}$$

where r_j is the price at agrovet j and c_{jv} is the cost of traveling to agrovet j, and returning to village v with a bag of fertilizer. Farmers must therefore make a round-trip for themselves, and a one-way trip for the bag of fertilizer. To calibrate these costs, we use survey questions which asked those farmers who traveled to retailers about travel costs for themselves and the fertilizer (Appendix Table B.2). We do this for a 50 kg bag of fertilizer, the modal amount purchased by farmers. We find that transporting a 50 kg bag of fertilizer costs about 69% as much as transporting a person for the same amount of time, implying therefore that a farmer must make 2.69 trips to buy a bag (2 for the farmer and 0.69 for the bag).

Equation (1) assumes that farmers have information on prices for every location, that they are free to travel to any location, and that they choose the lowest price from this menu. While we argue that this is the appropriate benchmark, some readers might argue that farmers make decisions using a simpler decision rule. While it is impossible to characterize all possible alternative decision rules, the most extreme possibility is that farmers only travel locally and prices in all other locations are irrelevant (i.e. costs beyond the nearest retailers are effectively infinite). Accordingly, we also conduct our analysis using the following travel cost adjusted price to the *nearest* agrovet.

$$r_v^{nearest} = r_{nearest} + c_{nearest,v} \tag{6}$$

On the output side, we construct the *maximum* travel cost-adjusted selling price for

this question. Most markets are directly on the main road; for those that are not on the road, we calibrate any remaining distance using the cost of local travel.

maize using a similar approach:

$$p_v^{min} = \max_m \{p_m - c_{mv}\} \tag{7}$$

Here, p_m is the price of maize post-harvest for market m, and c_{mv} is the cost of traveling from village v to market m. We use a 120 kg bag for this calculation, and assume that the cost of transporting the bag is proportional to the weight. Thus, a trip to the market and back to sell 120kg of maize requires 3.7 trips (2 for the farmer and 1.7 for the bag). Finally, as motivated above, we also calculate the price if farmers only transact at the nearest maize market.

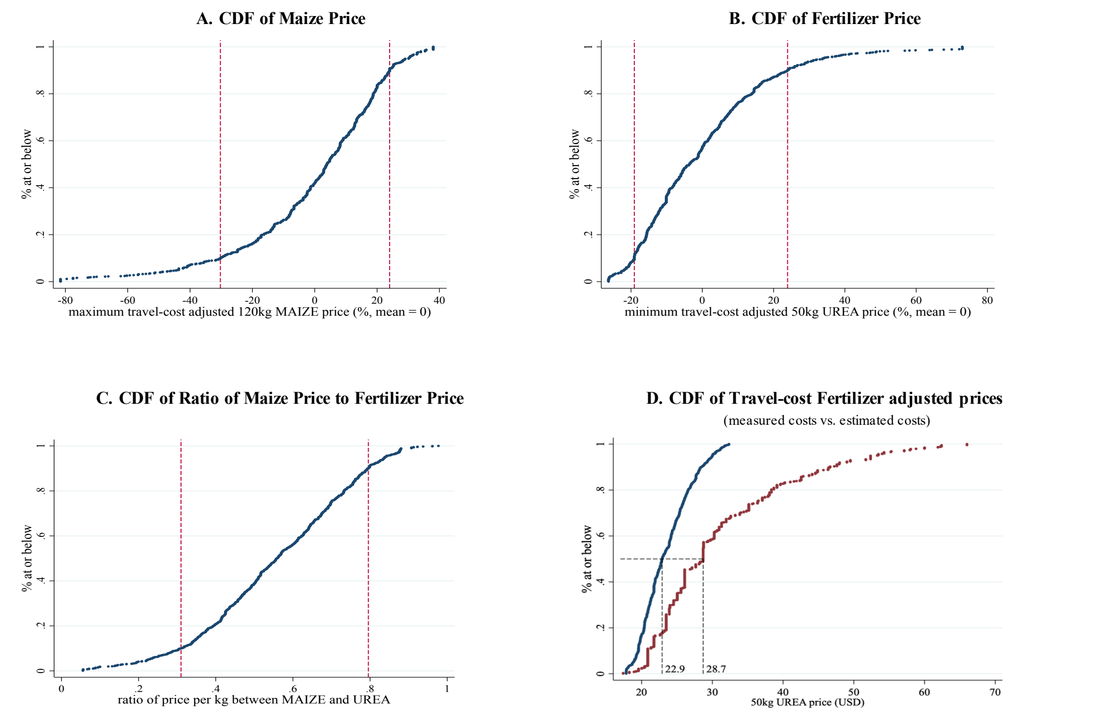
$$p_v^{nearest} = p_{nearest} - c_{nearest,v} \tag{8}$$

We calculate these prices for every village-agrovet and village-market pair. Figure 2.2 plots CDFs of village-level best prices of inputs and output, adjusting for travel costs, and shows tremendous heterogeneity in prices across villages. In Panel A, the difference of the travel cost-adjusted price for maize between the 90th and the 10th percentile is about 54% of the mean, while the standard deviation is about 23% of the mean. In Panel B, the 90-10 difference for the travel cost-adjusted price for fertilizer is about 43% of the mean, and the standard deviation is about 19% of the mean. To give a sense of the variation in profitability in using fertilizer, Panel C of Figure 2.2 calculates the ratio of the best travel-cost-adjusted maize price (per kg) to the best travel-cost-adjusted urea price (per kg). The 90-10 gap is 88% of the mean and the standard deviation is 34% of the mean. Web Appendix Figure B.1 shows analagous figures for prices at the nearest location, and figures look similar.

2.4.2 Reduced form analysis

In this subsection, we explore the relationship of these best prices (along with other outcomes like input usage, selling behavior, and access to retailers and markets) to the remoteness of the market.

Figure 2.2: CDF of Travel-cost Adjusted Prices Across Villages



Notes: Each observation represents a village. Travel-cost adjusted prices are calculated through observed prices from an agrovet survey, a maize price survey at markets and transport cost information collected from interviews with transport operators. In Panels A-C, the vertical dotted lines represent the 10th and 90th percentile. In Panel D, the vertical lines represent the median.

2.4.2.1 Specification

When evaluating the relationship between market conditions and remoteness for every village in the two study regions, the primary specification is,

$$m_{vt} = \beta_r \cdot R_v + \varepsilon_{vt} \tag{9}$$

where m_{vt} is a measure of market conditions (or a related outcome) at location v in year t, and R_v is a measure of remoteness.

For the measures of village-level market conditions estimated in (5), we include no controls. However, for farmer outcomes such as input adoption, it is clear that usage will depend not only on market access but also farmer-specific characteristics. Therefore, these are estimated as:

$$m_{fvt} = \beta_r \cdot R_v + \beta_X X_{fvt} + \varepsilon_{fvt} \tag{10}$$

where subscript f refers to farmer and X_{fvt} is a vector of other controls. These controls include information from the survey, such as land ownership, income, assets, education and other demographic characteristics, as well as soil information from the FAO-GAEZ. All farmer-level results are presented both with and without these controls.

2.4.2.2 Defining remoteness

To measure the remoteness of each village v, we focus on its proximity to selected "hubs" that are within or near the study regions: Arusha, Babati, Dodoma, Moshi, and Tanga. These locations are chosen because distributors for both maize and fertilizer are commonly located here.[35]

We use two measures that are motivated by the market access measure from Donaldson and Hornbeck (2016), but differ in requirements to estimate travel costs and distance elasticities. In the first, we define the remoteness of village v as a simple population weighted distance to each hub:

$$remoteness_v = \sum_h d_{hv} pop_h \tag{11}$$

where pop_h is the (relative) population of hub h (i.e. the population of that hub divided by the population of all hubs) and d_{hv} is distance from village v to that hub. In this measure, relative population is used as a proxy for the importance of each city in terms of availability of goods and average prices. Unlike Donaldson and Hornbeck, we use distance to measure proximity to hubs, rather than calculate the ad-valorem costs of travel and estimate a distance elasticity. We do this because it simplifies the construction

[35] Appendix Table B.3 presents input- and output-distributor locations, showing that nearly all of them are located in the towns of Arusha, Moshi, and Babati. We extend this list to also include the regionally important cities of Tanga and Dodoma, and our complete set of hubs are marked with "stars" in Figure 2.1. We do not include Dar es Salaam in the remoteness measure as its high relative population leads it to overwhelm all the other hubs.

of the measure substantially.[36]

The second measure engages on Donaldson and Hornbeck more directly, and calculates the market access of each village using the following formulation:

$$MA_v = \sum_h \tau_{hv}^{-\theta} pop_h \tag{12}$$

MA_v includes population weights as measures of the relative importance of each hub. These weights are adjusted by their elasticity-adjusted trade costs of reaching each hub, $\tau_{hv}^{-\theta}$. The cost term τ_{hv} is calculated as

$$\tau_{hv} = 1 + \frac{2.69 * cost_{hv}}{avgprice}$$

where $cost_{hv}$ is the estimated cost to get from village v to hub h, 2.69 is the number of one-way trips required to travel to a destination and return with a 50kg bag of fertilizer (see section 4.1), and $avgprice$ is the average price of fertilizer in the sample (measured at agrovets). We choose fertilizer as the benchmark good to measure ad-valorem costs, since it is the focus of the paper, and also because agrovets commonly report traveling to hubs to stock fertilizer. To measure the elasticity term, $-\theta$, we appeal to estimation later in the paper where the substitution elasticity across agrovets is estimated to be approximately -7.5, though the results are robust to other estimates.

We standardize both measures to have mean 0 and standard deviation 1 (and put a negative sign in front of MA_v, so that it measures remoteness rather than market access). The distributions of these variables are illustrated in Appendix Figure B.2. Both measures feature two modes, and for both measures the least remote areas (near major hubs) are approximately -2 standard deviations from the mean, while the most remote are +3 standard deviations away. The difference between these (5 standard deviations) is useful for benchmarking differences in outcomes between the most and least remote areas (similar to the approach taken in Atkin and Donaldson 2015).

[36] Despite this simplification, we show in the technical appendix that there exists a first-order approximation that links the two measures.

2.4.2.3 Summary statistics and correlations with remoteness

Table 2.3 presents summary statistics, and shows how these variables vary with remoteness. From Panel A, we see a number of differences: farmers in more remote areas are less educated, own fewer assets, have less access to finance, and earn less income from sources outside of farming. These farmers also tend to have larger families and larger farms.

Table 2.3: Remoteness and Farmer Characteristics

	(1)	(2)	(3)
	Mean	(Standardized) coefficient from regression of dependent variable on remoteness measure based on (population-weighted):	
		Distance to hubs	Elasticity-adjusted travel costs to hubs
Panel A. Demographic and background characteristics			
Age	49.76	-0.98*	-1.45***
	(15.23)	(0.52)	(0.50)
Female	0.45	-0.02	-0.02
		(0.02)	(0.02)
Married	0.76	0.00	0.01
		(0.01)	(0.01)
Household size	4.95	0.26**	0.36***
	(2.78)	(0.11)	(0.10)
Years of education	6.58	-0.31***	-0.47***
	(3.56)	(0.11)	(0.12)
Home has thatch roof	0.17	0.03	0.04**
		(0.02)	(0.02)
Has cell phone	0.89	-0.03***	-0.03***
		(0.01)	(0.01)
Has bank account	0.15	-0.05***	-0.05***
		(0.01)	(0.01)
Has mobile money account	0.77	-0.08***	-0.08***
		(0.02)	(0.01)
Acres of land	5.46	1.37**	2.65***
	(13.89)	(0.57)	(0.68)
Has market business	0.28	-0.05***	-0.06***
		(0.01)	(0.01)
Annual total income from non-farming (USD)	408.9	-74.72**	-87.91***
	(772.60)	(30.25)	(28.88)
Panel B. Production Capacity (in kg/acre)[1]			
FAO-GAEZ production capacity for low input level	788.3	70.07***	53.84***
	(290.70)	(21.21)	(19.10)
FAO-GAEZ production capacity for high input level	3325	-296.16***	-291.20***
	(876.00)	(57.38)	(58.96)
FAO-GAEZ production difference between high and low	2536	-366.23***	-345.04***
	(744.90)	(46.71)	(49.04)
Panel C. Harvest Output			
Total harvest output in 2016 long rains (kg)	928.7	-16.62	136.62**
	(1360.00)	(51.18)	(54.35)
Harvest output per acre		-85.11***	-82.61***
		(17.46)	(15.60)
Value of harvest output at average regional post-harvest price	201.9	-3.61	29.70**
	(295.60)	(11.13)	(11.82)

Notes: N = 2,845 farmers in 246 villages. In Column 1, standard deviations are in parentheses. Columns 2 and 3 show regression coefficients from separate regressions of the dependent variable on a measure of remoteness (equations 5 and 6 in the paper). See text for further discussion of these measures. In those columns, standard errors in parentheses, clustered at the village level.

*, **, and *** indicate significance at 10%, 5%, and 1%.

[1] Regressions for production capacity are at village level.

Panel B shows production capacity, based on GIS data from the FAO-GAEZ database, which provides information on counterfactual yields with and without inputs. At the mean, the FAO estimates that using inputs would more than quadruple yields. There is some evidence that more remote areas have lower returns to inputs – a 1 standard deviation increase in remoteness is associated with 14% lower increases in yields. However, we note that yield increases remain very large even in the most remote areas – for the most remote villages (3 standard deviations away), yields with inputs are still 140% higher than without inputs. We control for these measures in our main regressions, and develop a strategy to absorb factors like these within the spatial model. Panel C shows harvest output from the most recent long rains. While the relationship between total yields and remoteness depends on the measure used, yield per acre is lower in areas located farther away from hubs. In particular, a standard deviation increase in either measure of remoteness is associated with a reduction in harvest output per acre of about 20%. This is consistent with lower input usage in rural areas, or with differences in other factors such as soil quality.

In conclusion, Table 2.3 makes clear that it is difficult to pinpoint the role of input prices on outcomes, since access to roads is correlated with so many other characteristics. Ultimately, this motivates the use of an economic model to conduct counterfactuals.

2.4.2.4 Access to input markets

Table 2.4 shows how the two remoteness measures correlate with access to input markets. We first tabulate access to retailers within 10 km, which is a distance that is reasonably traveled by farmers.[37] Panel A shows several measures, including a dummy for retailer presence within 10 km, the number of retailers within 10 km, and the minimum distance to a retailer. On each measure, we find clear evidence of reduced access to retailers in more remote villages, all significant at 1%.

[37] Appendix Figure B.3 shows a CDF of the distance farmers travel to access inputs, conditional on purchase. We find that approximately 70% of purchases are made within 10 km of a farmer's village, and 85% within 20 km.

Table 2.4: Remoteness, Access to Input Markets and Retail Price Heterogeneity

	(1)	(2)	(3)
	Mean	(Standardized) coefficient from regression of dependent variable on remoteness measure based on (population-weighted):	
		Distance to hubs	Elasticity-adjusted travel costs to hubs
Panel A. Summary measures of access to input retailers			
Has at least 1 agrovet within 10 km of village	0.75	-0.14***	-0.13***
which sells fertilizer or seeds		(0.01)	(0.01)
Number of agrovets within 10 km of village	7.79	-2.93***	-4.18***
which sells fertilizer or seeds	(8.96)	(0.18)	(0.21)
Distance to nearest agrovet	6.79	3.17***	2.46***
which sells fertilizer or seeds	(15.15)	(0.73)	(0.58)
Distance to the second nearest village with an agrovet	15.52	5.38***	5.93***
which sells fertilizer or seeds	(23.97)	(1.02)	(0.86)
Panel B1. Travel-cost adjusted prices faced by farmers			
Minimum travel-cost adjusted price	24.19	2.33***	2.41***
for 50 kg of Urea (USD)[1]	(4.66)	(0.14)	(0.12)
Decomposition of price between retail price and cost of transportation			
Retail price at the location with the lowest	19.82	1.09***	1.27***
travel-cost adjusted price (USD)	(2.63)	(0.07)	(0.06)
Cost of travel to obtain minimum travel-cost	4.372	1.24***	1.14***
adjusted price (USD)	(4.39)	(0.14)	(0.12)
Panel B2. Travel-cost adjusted prices at the nearest agro-input shop			
Travel-cost adjusted price at the nearest input seller	26.55	2.37***	2.14***
for 50 kg of Urea (USD)[1]	(6.10)	(0.20)	(0.17)
Decomposition of price between retail price and cost of transportation			
Retail price at the nearest input seller (USD)	23.35	1.30***	1.29***
	(3.39)	(0.10)	(0.08)
Cost of travel to the nearest input seller (USD)	3.21	1.07***	0.85***
	(4.58)	(0.17)	(0.14)

Notes: The unit of observation is the village. Data is from the universe of villages in Kilimanjaro and Manyara regions (N = 1,183). Travel costs imputed from transport surveys and Google maps. In Column 1, standard deviations are in parentheses. Columns 2 and 3 show regression coefficients from separate regressions of the dependent variable on a measure of remoteness (equations 5 and 6 in the paper). See text for further discussion of these measures. In those columns, standard errors in parentheses, clustered at the village level.
*, **, and *** indicate significance at 10%, 5%, and 1%.

[1]We assume farmers buy a 50 kg bag in one trip (enough for 1 acre), and must incur the cost of a round-trip for herself, plus the cost of carrying the bag of fertilizer, equivalent to 0.7 trips.

Panel B1 of Table 2.4 shows our preferred measure of input market access, the minimum travel cost-adjusted price that is available to farmers. We find that one standard deviation of remoteness raises prices by \$2.33-2.41, equivalent to about 10% of the mean. This implies a difference in prices of approximately 50% between the most and least remote villages in our sample. We then decompose this price difference into differences in the retail price itself, and those in the travel cost. We find that the retail prices (at the optimal location) and transportation costs (to the optimal location) are approximately equal in their contribution toward the increase in minimum delivered prices.[38]

Panel B2 of Table 2.4 presents our secondary measure of access, the travel cost-adjusted price at the nearest shop. By definition, the travel cost-adjusted price is higher than in B1 (by about 10%); in particular, because farmers do not shop around, the retail price is higher and the travel cost is lower. As before, we find roughly equal contributions of each to the remoteness gradient.

These numbers give us a sense of the (pecuniary) ad-valorem equivalent transport costs for buying fertilizer. For the minimum price (Panel B1), transport costs are about 22% of the optimal purchase; for the nearest retailer (Panel B2), they are about 13%. While these costs are already substantial, when we examine farmer behavior more carefully in Section 5.2, we find that farmers behave as if costs are even larger than these pecuniary costs. [39]

[38] In Appendix Table B.6, we present 2 robustness checks. First, since we only surveyed retailers within the regional boundaries, we have no information on retailers in neighboring regions. It is possible, therefore, that there exist lower-priced retailers just across the border, causing us to potentially overstate travel cost-adjusted prices. To address this, in Panel A1, we drop all villages within 10 km of regional boundaries – the results are actually stronger. Second, while we had high survey completion rates among agrovets (91% – see Appendix Table B.1), we nevertheless do not have the universe of retail options. This suggests that retail price heterogeneity may be understated. To address this, we conduct a bounding exercise in Appendix Table B.6, Panel B, where we estimate the distribution of prices within regions. We then assign prices in the tails of this distribution (the 10th or 90th percentile) to missing agrovets in a way that attenuate our regression results – for example, in remote areas, we assign agrovets low prices. This exercise lowers the coefficient marginally, but the qualitative results are unchanged.

[39] How much price heterogeneity can be explained by retailer pricing behavior? While causal identification is challenging (since entry is endogenous), we provide descriptive evidence in Appendix Table B.5. From Panel A, we find some evidence that more remote shops sell different products (further complicating inference) – remote shops are less likely to sell fertilizer but more likely to sell seeds. We find strong evidence that retailers face higher costs of procuring supply from wholesalers; in this setting, retailers typically travel themselves to wholesalers to purchase inventory, and so it is intuitive that these procurement costs are higher in remote areas. In Panel B, we examine wholesale and retail prices. We find evidence that remote retailers charge higher prices but they also face higher wholesale prices (perhaps because competition is weaker among available wholesalers). Ultimately we find that mark-ups are no higher in remote areas. This descriptive evidence suggests that pricing behavior is likely a secondary factor in explaining higher prices.

2.4.2.5 Access to output markets

Table 2.5 performs a similar analysis, but on the output side. As before, Panel A shows that more remote villages are less likely to have a market within 10 km, and the nearest market where maize is sold is located farther away. Panel B1 shows travel cost-adjusted prices for maize. Since there are large seasonal price fluctuations in rural Tanzania (as in much of rural Africa),[40] we use a price for the single point in time which is most relevant for farmers: immediately post-harvest (our surveys show that most farmers who sell do so shortly after harvest). We find that across both remoteness measures, travel cost-adjusted prices of output are lower in remote areas. As before, we decompose this into the retail price and the travel costs, finding that while retail maize prices rise modestly with remoteness, transport costs to their best maize market rise by $3.9 with each standard deviation in remoteness, overwhelming the increase in the price of maize. In Panel B2, we repeat the analogous exercise from the input market to evaluate the impact of remoteness when farmers simply choose the *closest* weekly maize market to sell their harvest. By definition, average travel cost adjusted sales prices are lower, and empirically the magnitude is large (about 50%). As in Panel B1, we find that this price declines with remoteness, and in fact the point estimate is similar. However, the decomposition between the retail price and the travel cost is very different: for the nearest price, the retail price falls and the travel cost rises.

Finally, we show one other measure of price, in this case measured at the village level. First, in Panel B3, we report coefficients from farmers' self-reported "going price" of maize after the last harvest, regressed on measures of remoteness. Consistent with the above, the going price in the village is decreasing in remoteness. This is intuitive if maize agents are traveling from the larger population centers (which are used to construct our remoteness measures), and offering lower selling prices to compensate for the higher costs of travel. Overall, whether searching for the best market, or selling locally, the returns from selling maize are clearly lower in more remote regions.

[40] Aggarwal et al. (2018) document an average price increase of about 46% over the season for the years 2006-16 in Kisumu market in neighboring Kenya; Burke et al. (2019) document increases in the range of 15-30% for a sample of markets in the east African region.

Table 2.5: Remoteness, Access to Output Markets and Output Price Heterogeneity

	(1)	(2)	(3)
	Mean	(Standardized) coefficient from regression of dependent variable on remoteness measure based on (population-weighted):	
		Distance to hubs	Elasticity-adjusted travel costs to hubs
Panel A. Summary measures of access to output markets			
Has at least 1 maize seller within 10 km of village	0.67	-0.16***	-0.16***
		(0.01)	(0.01)
Number of maize sellers within 10 km of village	1.89	-1.06***	-1.48***
	(2.48)	(0.07)	(0.08)
Distance to nearest output market with maize sellers (km)	8.67	5.65***	4.39***
	(14.17)	(0.71)	(0.43)
Panel B1. Maximum imputed travel-cost adjusted price if farmers were to sell in a local market			
Market survey: maximum travel-cost adjusted price	30.30	-3.08***	-3.05***
immediately after 2017 harvest (USD)[1]	(7.24)	(0.22)	(0.19)
Decomposition of price between retail price and cost of transportation			
Retail price at the location with the highest	39.34	0.80***	0.23**
travel-cost adjusted price (USD)	(3.17)	(0.08)	(0.09)
Cost of travel to obtain the highest travel-cost	9.05	3.88***	3.28***
adjusted price (USD)	(7.06)	(0.21)	(0.18)
Panel B2. Travel-cost adjusted otput price at the nearest maize selling market			
Travel-cost unadjusted 120 kg bag of maize price	20.83	-3.26***	-3.16***
immediately after 2017 harvest (USD)[1]	(8.98)	(0.29)	(0.24)
Decomposition of price between retail price and cost of transportation			
Retail price at the nearest maize selling market (USD)	26.67	-1.37***	-1.77***
	(5.95)	(0.19)	(0.15)
Cost of travel to the nearest maize selling market (USD)	5.840	1.89***	1.39***
	(5.88)	(0.21)	(0.17)
Panel B3. Price available within village by maize-buying intermediaries immediately after last season's harvest			
Farmer surveys: average "going price" in local	25.86	-1.31**	-2.60***
village immediately after previous harvest[2]	(6.24)	(0.52)	(0.48)

Notes: The unit of observation is the village. Data is from the universe of villages in Kilimanjaro and Manyara regions (N = 1,183). Travel costs imputed from transport surveys and Google maps. In Column 1, standard deviations are in parentheses. Columns 2 and 3 show regression coefficients from separate regressions of the dependent variable on a measure of remoteness (equations 5 and 6 in the paper). See text for further discussion of these measures. In those columns, standard errors in parentheses, clustered at the village level.
*, **, and *** indicate significance at 10%, 5%, and 1%.

[1]We assume farmers sell a 120 kg maize bag in one trip, and must incur the cost of a round trip for herself and the cost of carrying the maize that is equivalent to 1.7 trips.

[2]Data is from the farmer surveys (2,171 farmers in 137 villages).

2.4.2.6 Farmer decisions

The results so far show clear evidence of reduced market access in more remote areas for both inputs and output, and of higher prices for inputs, lower (travel cost-adjusted) prices for output, and lower "going" prices for output within the village. These results lead us to expect lower input usage and maize sales in more remote areas. We investigate this in Table 2.6, where we present results with and without a full set of farmer controls. In Panel A, we present the extensive and intensive margin of input use, for both seeds and fertilizer. In all specifications, these relationships are strong (significant at 1%) and large. We find that use of fertilizer is 9-20 percentage points lower in villages 1 standard deviation away, and that of hybrid seeds is 5-11 percentage points lower. Since the distance between the least and most remote regions is about 5 standard deviations, the regressions predict at least 45 percentage point lower usage of fertilizer in the most remote villages, which translates to about 80% of the mean in the least remote areas. The effect for seeds is smaller but still evident.

Similarly, in Panel B, we see strong evidence that sales are lower in remote areas, especially when using the simple weighted-average distance measure of remoteness. While the regression predicts that 44% of farmers will sell in the least remote areas, this declines to only 14% in the most remote areas. This is predominantly coming from a decline in sales to agents (since agents are by far the most common way to sell maize), but there are declines in sales at the market as well.

Consistent with this, Panel C shows buying behavior. Remote farmers are more likely to buy maize and to be net buyers of maize. Interestingly, we find a lot of heterogeneity in net buying behavior - 37% of farmers buy maize but sell none, 24% sell maize but buy none, and only 8% buy and sell maize (the other 30% do not transact at all).

2.5 Model

In this section, we quantify the impact of access to input and output markets by developing a spatial model of fertilizer adoption. In the model, we develop a rigorous

Table 2.6: Remoteness and Input Market Access and Adoption

	(1)	(2)	(3)	(4)	(5)
		(Standardized) coefficient from regression of dependent variable on remoteness measure based on (population-weighted):			
	Mean	Distance to hubs		Elasticity-adjusted travel costs to hubs	
		No controls	Controls for soil and farmer characteristics	No controls	Controls for soil and farmer characteristics
Panel A: Input usage					
Used chemical fertilizer in previous long rains	0.39	-0.17***	-0.09***	-0.20***	-0.13***
		(0.03)	(0.03)	(0.03)	(0.03)
Quantity of chemical fertilizer used (kg)	19.84	-13.06***	-6.46***	-14.42***	-9.33***
	(31.63)	(2.15)	(1.74)	(1.91)	(1.88)
Used improved seeds in previous long rains	0.66	-0.07***	-0.05**	-0.11***	-0.10***
		(0.02)	(0.02)	(0.02)	(0.03)
Quantity of improved seeds used (kg)	6.29	-1.30***	-1.21***	-1.09***	-1.03**
	(8.21)	(0.36)	(0.44)	(0.32)	(0.43)
Panel B. Maize sales					
Sold maize after previous long rains	0.32	-0.09***	-0.06**	-0.07***	-0.04*
		(0.02)	(0.03)	(0.02)	(0.02)
Total quantity sold (kg)	388.1	-97.86***	-112.16**	-5.90	-19.23
	(1142.00)	(35.11)	(48.86)	(39.86)	(47.71)
Sales to agents at home					
Agent visited homestead	0.31	-0.14***	-0.09***	-0.12***	-0.07*
		(0.03)	(0.03)	(0.03)	(0.04)
Sold maize to agent after previous long rains	0.17	-0.07***	-0.04**	-0.05***	-0.02
		(0.02)	(0.02)	(0.01)	(0.02)
Quantity sold to agents (kg)	142	-46.39***	-39.11**	-13.14	4.06
	(433.70)	(13.80)	(18.61)	(12.51)	(19.12)
Sales at market					
Sold maize at market after previous long rains	0.06	-0.03***	-0.03***	-0.02***	-0.02**
		(0.01)	(0.01)	(0.01)	(0.01)
Quantity sold at market (kg)	34.42	-14.61***	-15.26**	-9.53	-11.26
	(197.10)	(5.32)	(7.72)	(5.98)	(7.40)
Panel C. Maize purchases					
Farmer ever buys maize	0.48	0.11***	0.08***	0.11***	0.09***
		(0.02)	(0.02)	(0.02)	(0.02)
Quantity purchased in typical year (kg)	152.3	75.67***	65.05***	90.29***	77.49***
	(315.50)	(15.79)	(17.92)	(16.06)	(14.37)
Net buying					
Farmer buys maize but sells none	0.37	0.11***	0.08***	0.11***	0.08***
		(0.02)	(0.03)	(0.02)	(0.02)
Farmer sells maize and buys none		-0.09***	-0.06***	-0.08***	-0.06***
		(0.02)	(0.02)	(0.01)	(0.02)
Farmer buys and sells maize	0.08	-0.00	0.00	0.01	0.02
		(0.01)	(0.01)	(0.01)	(0.01)
Net buyer (quantity bought > quantity sold)	0.32	-0.09***	-0.06**	0.11***	0.08***
		(0.02)	(0.03)	(0.02)	(0.03)
Net seller (quantity bought < quantity sold)	0.17	-0.07***	-0.04**	-0.07***	-0.05*
		(0.02)	(0.02)	(0.02)	(0.03)

Notes: N = 2,845 farmers in 246 villages. See text for sampling details. Standard deviations are in parentheses in Column 1. Columns 2-5 show regression coefficients from separate regressions of the dependent variable on a measure of remoteness (equations 5 and 6 in the paper). See text for further discussion of these measures. In those columns, standard errors in parentheses, clustered at the village level.
*, **, and *** indicate significance at 10%, 5%, and 1%.

framework of retailer choice, while including other factors that affect adoption but are not related to input access. Ultimately, using a model specified measure of input access that we estimate, we run counterfactuals that study the role of transportation costs in the adoption decision.

2.5.1 Model Preliminaries

Production and Inputs

We begin the model by presenting the two technologies available to farmers, and the role of retailer choice in affecting farmer productivity. For farmer i, the production function *without* fertilizer is:

$$Y_{i0} = \widetilde{\theta}_{i0} K_i^{\alpha_0} L_{i0}^{1-\alpha_0} \tag{13}$$

Here, $\widetilde{\theta}_{i0}$ is baseline productivity without fertilizer, K_i is land held by farmer i, and L_{i0} is labor hired/used by farmer i. If the going wage rate for i is w_i and the selling price of maize is p_i, holding land fixed, profits can be derived as

$$\begin{aligned} \Pi_{i0} &= \alpha_0 (1-\alpha_0)^{\frac{1-\alpha_0}{\alpha_0}} \widetilde{\theta}_{i0}^{\frac{1-\alpha_0}{\alpha_0}} p_i^{\frac{1}{\alpha_0}} w_i^{-\frac{1-\alpha_0}{\alpha_0}} K_i \\ &= \theta_{i0} \pi_{i0} \end{aligned} \tag{14}$$

where $\theta_{i0} = \alpha_0 (1-\alpha_0)^{\frac{1-\alpha_0}{\alpha_0}} \widetilde{\theta}_{i0}^{\frac{1-\alpha_0}{\alpha_0}}$ and $\pi_{i0} = p_i^{\frac{1}{\alpha_0}} w_i^{-\frac{1-\alpha_0}{\alpha_0}} K_i$.[41] The former term, θ_{i0}, will be represented by a random variable with a village-specific mean, and the latter will be calculated as a function of observed data for farmer i and elasticities that must be estimated.

The production function *with* fertilizer has both labor and fertilizer as variable inputs, while maintaining the basic Cobb-Douglas assumption (for tractability). When using fertilizer, farmers not only have a choice of how much fertilizer to buy, but also which retailer to choose. Supposing that farmer i buys fertilizer from agrovet j in

[41] We assume for tractability that all farmers internalize a market price for maize and labor.

location v, the production function is written as:

$$Y_{ijv} = \widetilde{\theta}_{ijv} (\theta_i K_i)^{\alpha} L_{ijv}^{(1-\alpha)\beta} M_{ijv}^{(1-\alpha)(1-\beta)} \tag{15}$$

Note we are assuming that the exponents on capital and labor may be different for the technology with fertilizer, which as we will show below, allows for output prices to affect adoption decisions (while maintaining the analytical simplicity of a basic Cobb-Douglass technology).[42] Further, when using fertilizer, there are two additional productivity terms to consider. The first is the known local productivity of using fertilizer, θ_i, which in the production function, scales the effective amount of land for farmer i. The second is a productivity shock for farmer i, $\widetilde{\theta}_{ijv}$, that potentially varies by the agrovet j and location v where the fertilizer was purchased. We discuss this particular productivity shock when solving for optimal retailer choice.

Writing the delivered price of fertilizer to i from agrovet j in location v as r_{ijv}, solving for the optimal labor and fertilizer inputs (see the appendix for the derivations), profits are written as

$$\Pi_{ijv} = \theta_{ijv} \pi_i r_{ijv}^{-\sigma} \tag{16}$$

where $\sigma \equiv \frac{1-\alpha}{\alpha}(1-\beta)$, $\pi_i = \theta_i p_i^{\frac{1}{\alpha_0}} w_i^{-\beta \frac{1-\alpha_0}{\alpha_0}} K_i$, and $\theta_{ijv} = \kappa_2 \widetilde{\theta}_{ijv}^{\kappa_1}$.[43] Here, the profitability of fertilizer is a function of the productivity shock, θ_{ijv}, the (delivered) price of fertilizer itself, r_{ijv}, and deterministic profits based on local factors and technology, π_i.

Input and Agrovet Choice

Farmers choose whether to purchase fertilizer, and if so, how much and from where. These decisions are affected by prices for fertilizer at each agrovet location, the productivity shock received in buying from a particular location, and the round-trip travel costs. Suppose that the set of villages that contain an agrovet is defined as $\mathcal{V}$, where

42 The Cobb-Douglas framework to model agricultural production shares similarities with recent work in Chatterjee (2019) and Gollin and Udry (2019), though unlike the latter, we do not model allocations across different plots within the household.

43 κ_1 and κ_2 are constant functions of model parameters.

the price charged at location $v \in \mathcal{V}$ by agrovet j is r_{jv}. The per-unit cost to the farmer i, inclusive of transport costs, will be written as $r_{ijv} = r_{jv}\tau_{iv}$, where τ_{iv} is an iceberg trade cost for farmer i in traveling to v and back. The assumption of iceberg trade costs will facilitate a decomposition of the model that aids estimation and calibration.

We assume that θ_{ijv} is a random variable that measures the benefit of i purchasing at agrovet j in location v. These latter benefits could represent other inputs purchased in location v (hybrid seeds, for example), availability of extension services at location v, or perhaps other networking and information that is acquired at location v that may affect profitability. Further, it may represent the probability of getting bad or adulterated inputs at a given retail location, or given the functional form of (16), measurement error in the price at a retail location. Whatever the interpretation, for analytical convenience we assume that θ_{ijv} is distributed according to a Fréchet distribution with location parameter T_{jv} and dispersion parameter ε. Precisely:

$$\Pr\left(\theta_{ijv} < \theta\right) = \exp\left(-T_{jv}\theta^{-\varepsilon}\right)$$

That is, while each farmer may get a random draw from this distribution, its central moments are specific to the retail location itself. Using this distributional assumption, the unconditional distribution of profits for farmer i buying from agrovet j in location v is written as:

$$\Pr\left(\Pi_{ijv} < \pi\right) = \exp\left(-T_{jv}\pi_i^{\varepsilon} r_{ijv}^{-\varepsilon\sigma}\pi^{-\varepsilon}\right)$$

We also assume that the outside option of not buying fertilizer is random. Specifically, θ_{i0} is distributed Fréchet with location parameter T_{i0} and the same dispersion parameter ε. Thus, the distribution of profits without fertilizer is written as:

$$\Pr\left(\Pi_{i0} < \pi\right) = \exp\left(-T_{i0}\pi_{i0}^{\varepsilon}\pi^{-\varepsilon}\right)$$

Here, we allow for the average productivity of the outside option of not buying fertilizer to vary by village i through the location parameter T_{i0}. This may reflect difficulties in using or adopting fertilizer that are specific to a location (poor soil quality, lack of training, existing norms, etc.).

Farmer i chooses among locations $v \in \mathcal{V}$ and agrovets $j \in \mathcal{J}_v$ at each location to find the most profitable option. Solving the standard discrete choice problem (which is derived in the technical appendix), the probability that farmer i buys from agrovet j at location v is written as:

$$\lambda_{ijv} = \frac{T_{jv}\pi_i^{\varepsilon} r_{ijv}^{-\varepsilon_a}}{T_{i0}\pi_{i0}^{\varepsilon} + \sum_{v' \in \mathcal{V}} \sum_{l \in \mathcal{J}_{v'}} T_{lv'}\pi_i^{\varepsilon} r_{ilv'}^{-\varepsilon_a}} \tag{17}$$

Here, we have imposed $\varepsilon_a = \varepsilon\sigma$, with ε_a being a critical elasticity to estimate. Summing across all agrovet options, the probability that farmer i adopts in any location is written as:

$$\mu_i = \frac{\sum_{v' \in \mathcal{V}} \sum_{l \in \mathcal{J}_{v'}} T_{lv'} r_{ilv'}^{-\varepsilon_a}}{T_{i0}\left(\frac{\pi_{i0}}{\pi_i}\right)^{\varepsilon} + \sum_{v' \in \mathcal{V}} \sum_{l \in \mathcal{J}_{v'}} T_{lv'} r_{ilv'}^{-\varepsilon_a}} \equiv \frac{\Phi_i}{\Phi_{i0} + \Phi_i} \tag{18}$$

In (18), we define the two terms that fully characterize the adoption decision for each farmer i. First, we define farmer i's *market access* to inputs as $\Phi_i = \sum_{v' \in \mathcal{V}} \sum_{l \in \mathcal{J}_{v'}} T_{lv'} r_{ilv'}^{-\varepsilon_a}$, which after imposing the iceberg assumption and simplifying, can be written as

$$\Phi_i \equiv \sum_{v \in \mathcal{V}} \tau_{iv}^{-\varepsilon_a} \phi_v$$

where $\phi_v = \sum_{l \in \mathcal{J}_v} T_{lv} r_{lv}^{-\varepsilon_a}$. Here, market access is a function of the elasticity-adjusted iceberg to a given village, $\tau_{iv}^{-\varepsilon_a}$, and a local index, ϕ_v, which is the sum of elasticity-adjusted local prices weighted by the local input-quality. Second, we define the *outside option* to buying fertilizer as

$$\Phi_{i0} = T_{i0}\left(\frac{\pi_{i0}}{\pi_i}\right)^{\varepsilon}$$

which is the relative profitability of using fertilizer (compared to not using fertilizer), adjusted for local productivity factors that are unrelated to market access. To be clear, we will not be able to disentangle each component of Φ_{i0} lacking additional data and identifying variation in output markets or the micro-foundations of farm production.

However, for purposes of calibrating the model-derived measure of market access and its relationship to the remoteness of villages from population centers, Φ_{i0} will be useful in absorbing all other variation in adoption as a residual.

2.5.2 Calibrating the Farmer's Problem

In linking the model to the data, we will make use of the novel cross-sectional surveys as described in sections three and four, and match this data to adoption and location-choice probabilities as specified by the model. To make clear what we have, and what we do not, we note that (17) can be broken up into the probability of adoption for i, μ_i; the probability i buys somewhere at location v conditional on adopting at all, $\lambda_{iv|adopt}$; and finally, conditional on adopting from an agrovet at location v, the probability that agrovet j is chosen, $\lambda_{j|adopt\,at\,v}$:

$$\begin{aligned}\lambda_{ijv} &= \underbrace{\frac{\Phi_i}{\Phi_{i0}+\Phi_i}}_{\mu_i} \cdot \underbrace{\frac{\tau_{iv}^{-\varepsilon_a}\phi_v}{\sum_{v'\in\mathcal{V}}\tau_{iv'}^{-\varepsilon_a}\phi_{v'}}}_{\lambda_{iv|adopt}} \cdot \underbrace{\frac{T_{jv}r_{jv}^{-\varepsilon_a}}{\sum_{l\in\mathcal{J}_v}T_{lv}r_{lv}^{-\varepsilon_a}}}_{\lambda_{j|adopt\,at\,v}} \\ &= \mu_i\cdot\lambda_{iv|adopt}\cdot\lambda_{j|adopt\,at\,v}\end{aligned}$$

The farmer surveys collect data to calculate all three probabilities, though only the first two reliably (since some farmers could not recall the name of the agrovet at which they purchased). However, the first two probabilities contain a significant amount of information that is useful to calibrating the farmers problem, and we use this data extensively below.

To calibrate terms important for the farmer's problem, we proceed in four steps. First, we use $\lambda_{iv|adopt}$ as reported in our "trips" surveys to estimate a functional form for $\tau_{iv}^{-\varepsilon_a}$, via multinomial logit. Second, we use the model to solve for a value of $T_{jv}r_{jv}^{-\varepsilon_a}$ for each agrovet that exactly equates observed agrovet revenues with expected expenditures. Together with the trade costs from step one, this will yield a measure of Φ_i for each farmer. Third, we use remaining variation in μ_i, to solve for the outside option residual, Φ_{i0}. Finally, we decompose $T_{jv}r_{jv}^{-\varepsilon_a}$ into its components using a IV strategy and information from our agrovet surveys, at which point we can calculate λ_{ijv} for all farmer-agrovet combinations. We now detail each step in order.

Estimating Transport Costs through Location Choice

In the first step, we focus on the choice probability for location v, conditional on adopting anywhere:

$$\lambda_{iv|adopt} = \frac{\tau_{iv}^{-\varepsilon_a}\phi_v}{\sum_{v' \in \mathcal{V}} \tau_{iv'}^{-\varepsilon_a}\phi_{v'}} \tag{19}$$

To estimate equation (19), we need a dataset that identifies when each farmer i chooses location v to purchase fertilizer. Thus, defining I as the set of farmers who adopt, and $\mathcal{V}$ as the set of locations with an agrovet, we construct a $I \times \mathcal{V}$ dataset of bilateral visit indicators. For each bilateral combination, we will also measure the routed distance in kilometers between the farmer's village and the potential purchase location, $dist_{iv}$ (similar to Section 3).

Exponentiating the village share equation, and re-writing $\log(\phi_v)$ into a location v fixed effect, d_v, we can write:

$$\lambda_{iv|adopt} = \frac{\exp(d_v - \varepsilon_a \log(\tau_{iv}))}{\sum_{v' \in \mathcal{V}} \exp(d_{v'} - \varepsilon_a \log(\tau_{iv'}))}$$

As the main objective from this section is to assess the role of trade costs in agrovet choice (and consequently, adoption), we need to specify a functional form for trade costs, τ_{iv}. As a starting point, we will estimate a simple linear relationship between the elasticity adjusted log trade cost and distance, $-\varepsilon_a \log(\tau_{iv}) = \beta_{dist} dist_{iv}$.

To allow for a potentially non-linear cost of travel for farmers by distance (for example, if transport technologies or non-pecuniary costs differ at longer distances), we will also use distance bins D_{iv}^b, which are equal to one if the distance between i and v is in bin b, and zero otherwise. With the distance bins, the multinomial logit is written as:

$$\lambda_{iv|adopt} = \frac{\exp\left(d_v + \sum_b \beta_b D_{iv}^b\right)}{\sum_{v' \in \mathcal{V}} \exp\left(d_{v'} + \sum_b \beta_b D_{iv'}^b\right)} \tag{20}$$

Equation (20), and its linear alternative, can be estimated by McFadden's alternative-specific conditional logit. The results from doing so are presented in Table 2.7. In the first column, we present the linear specification of distance. Assuming $\varepsilon_a \approx 7.5$ (which will be supported by later results), the results suggest that iceberg transport costs for fertilizer are 2.3% ad-valorem per kilometer.

As technologies may change discretely depending on the distance to each agrovet (walking short distances, taking transit for long distances), our preferred specification using distance bins is presented in Column 2, while Column 3 reports the ad-valorem equivalent per kilometer when evaluated at the farthest distance that defines each bin, and with trade costs compounded each kilometer.[44] The estimates suggest costly travel for farmers acquiring fertilizer. To interpret the coefficients, we take two approaches. In the first, we compare two locations with the same "return" from fertilizer, d_v, and then focus on the reduction in probability if one is (0,5] km away rather than 0 km away (in the same village). In this case, the probability that one chooses the location (0,5] km away compared to 0 km away (in the home village) for idiosyncratic reasons that overcome trade costs is 0.25.[45]

Alternatively, we can interpret the results as log changes in trade costs via $\log(\tau_{iv}) = -\frac{1}{\varepsilon_a}\sum_b \beta_b D_{iv}^b$, where dividing the coefficient estimates by ε_a gives us log trade costs. Given the iceberg assumption, this is also interpreted as the log change in the delivered price. Thus, at the central estimate of $\varepsilon_a \approx 7.5$, the comparison is equivalent to approximately 4% ad-valorem trade cost per km for the first three bins (up to 15km). When evaluated to the typical closest agrovet (6.7 km), the ad-valorem equivalent trade cost is about 30%. Beyond this, the ad-valorem equivalent per kilometer falls modestly, which is consistent with our transport surveys and also the likelihood that longer distances require a more efficient means of travel (though still at a high overall cost, on average).

44 Precisely, the ad-valorem equivalent per kilometer is $(1+\tau_{iv})^{1/km} - 1$, where τ_{iv} is the ad-valorem equivalent for the entire trip.

45 This is calculated precisely by calculating the ratio of probabilities:

$$\frac{\lambda_{0-5km}}{\lambda_{0km}} = \frac{\exp(d_v - 1.38)}{\exp(d_v - 0)} = 0.25$$

Table 2.7: Multinomial Logit of Agrovet Choice

	(1)	(2)	(3)
	Agrovet Chosen		AVE/KM
Kilometers to agrovet	-0.171***		2.3%
	(0.009)		
Dummies for agrovet distance bin:			
between (0,5] km		-1.380***	3.7%
		(0.372)	
between (5,10] km		-2.914***	4.0%
		(0.379)	
between (10,15] km		-4.331***	3.9%
		(0.380)	
between (15,20] km		-5.367***	3.6%
		(0.400)	
between (20,30] km		-5.875***	2.6%
		(0.378)	
between (30,40] km		-7.602***	2.6%
		(0.449)	
between (40,50] km		-8.685***	2.3%
		(0.495)	
between (50,100] km		-10.625***	1.4%
		(0.560)	
over 100 km		-14.253***	-
		(0.992)	

Notes: N = 519 farmers, 119 observed locations. Omitted group is agrovet located in respondent's village. Ad-valorem equivalent per kilometer is calculated at the upper bound of each bin, and assumes that the trade cost compounds each kilometer. Standard errors in parentheses. *, **, and *** indicate significance at 10%, 5%, and 1%.

An advantage of using revealed adoption choices to estimate trade costs is that the iceberg estimates described above may include both pecuniary and non-pecuniary costs of travel. Again, when including both, we estimate that the ad-valorem equivalent to the closest agrovet is about 30%. In section 4.2.4, we summarize the ad-valorem equivalent travel cost of buying from the nearest retailer, estimating 13%. Thus, the pecuniary cost is roughly 43% of the overall cost of travel, suggesting significant non-pecuniary costs of acquiring fertilizer.

Finally, similar to Section 4.2, we calculate best trade-cost-adjusted prices for agrovets for all villages in the region, using the binned estimates of iceberg costs as described above. These results are presented in Panel D of Figure 2.2. Here, there is significantly more heterogeneity in best trade cost adjusted prices for fertilizer, again suggesting sizable non-pecuniary costs of traveling to acquire fertilizer. Precisely, at the median best travel cost adjusted price, the delivered price using our ad-valorem estimates is \$28.7 for a 50kg bag, which is approximately 25% higher than the median best-delivered price when using the pecuniary costs from transport surveys.

Model Calibration

In the conditional multinomial logit above, if enough farmers were sampled such that every location with an agrovet was chosen, we could estimate precisely a value of ϕ_v for each location (up to a standard normalization), and use this for the baseline equilibrium in resulting counterfactuals. Unfortunately, funding was not sufficient to survey such a large sample, and thus, to recover all non-price attributes of all locations that contain an agrovet, we use agrovet revenue shares from our agrovet survey, and the spatial distribution of fertilizer expenditures from the farmer survey. Specifically, for the second step of the calibration, we solve for the vector of quality adjusted fertilizer prices $T_{jv}r_{jv}^{-\varepsilon_a}$ that exactly equates supply and demand for fertilizer at each agrovet.

To derive a market-clearing condition that we intend to calibrate, we start from an equation that summarizes expected agrovet sales as aggregated from spatial farmer-level demand. Defining expected agrovet sales at j in v as $\mathbb{E}[v_{jv}]$, we have:

$$\mathbb{E}[v_{jv}] = \sum_i L_i \mu_i \lambda_{ijv|adopt} \mathbb{E}[F_i | adopt \ at \ jv]$$

where $\mathbb{E}[F_i|adopt \ at \ jv]$ is expected fertilizer expenditures by i, conditional on adopting at jv, and L_i is the village population to use as weights in the demand equation. As this conditional expectation is not observed in any practical way, we will appeal to the structure of the model to simplify to an unconditional expectation for fertilizer expenditures by i. Precisely, using the properties of the Fréchet distribution, it is straightforward to show that $\mathbb{E}[F_i|adopt \ at \ jv] = \mathbb{E}[F_i|adopt]$; that is, the expected expenditures conditional on adoption anywhere is the same as the expected expenditures at some j, conditional on choosing j.[46] Noting further that $\mu_i \mathbb{E}[F_i|adopt] = \mathbb{E}[F_i]$, and imposing the definition

[46] See technical appendix for a proof.

of $\lambda_{ijv|adopt}$, we get:

$$\mathbb{E}[v_{jv}] = \sum_i L_i \left(\frac{T_{jv}\tau_{iv}^{-\varepsilon_a} r_{jv}^{-\varepsilon_a}}{\sum_{v' \in \mathcal{V}} \sum_{l \in \mathcal{J}_{v'}} T_{lv'}\tau_{iv'}^{-\varepsilon_a} r_{lv'}^{-\varepsilon_a}} \right) \mathbb{E}[F_i]$$

Finally, we can combine the agrovet-specific non-price attributes and the price into an "agrovet-effect" ($\eta_{jv} \equiv T_{jv} r_{jv}^{-\varepsilon_a}$), and also impose the specification for transportation costs, to get:

$$\mathbb{E}[v_{jv}] = \sum_i L_i \left(\frac{\exp\left(-\sum_b \hat{\beta}_b D_{iv}^b\right) \eta_{jv}}{\sum_{v' \in \mathcal{V}} \sum_{l \in \mathcal{J}_{v'}} \exp\left(-\sum_b \hat{\beta}_b D_{iv'}^b\right) \eta_{lv'}} \right) \mathbb{E}[F_i] \tag{21}$$

To implement this equation, we use observed fertilizer revenues for each agrovet to proxy for $\mathbb{E}[v_{jv}]$, and village-level fertilizer expenditures from the farmer's survey as an unbiased estimate for $\mathbb{E}[F_i]$, i.e., for this equation, we take i to represent villages and sum up expenditures within each village.

However, we again run into a number of issues as funding was not sufficient to survey more farmers and more villages. There are two issues to consider. First, while we surveyed approximately 18 farmers per village, in some villages zero or full adoption is reported. In reality, this may be accurate, or may be biased toward the bounds by a small sample. Village adoption at the bounds complicates the calibration of the overall adoption decision (given the logit functional form). To facilitate a feasible calibration that is consistently applied across market clearing conditions and the adoption decision, we first winsorize the village adoption data to fall between 0.025 and 0.975.[47] Then, for those villages that report zero adoption in the sample, we assign a small value of $\mathbb{E}[F_i]$ that is calculated via the model using the reported land holdings of the village in the sample, the winsorized adoption share (0.025), and the 1st percentile value of fertilizer expenditures per acre of land across the entire sample of farmers who adopt, $\left(\frac{F_i}{K_i}\right)_{1st}$.[48]

Focusing on the sampling of villages, if we assume that the farmer sample captures

47 This effectively means that villages with zero adoption in the sample are assigned a level of adoption 50% lower than the lowest observed (positive) adoption share in the sample. Or, alternatively, that we would need to double the within-village sample size to find one farmer who adopts.

48 Precisely, imputed (small) values for expected fertilizer expenditures are calculated by: $\mathbb{E}[F_i] = 0.025\left(\frac{F_i}{K_i}\right)_{1st} \cdot K_i$

the entire geography of demand, there will exist agrovets in other locations that appear more remote than they actually are since no farmers were surveyed in that location. This will cause a bias in estimates of η_{jv} by assigning a large value for agrovet locations without any farmers surveyed to make-up for the incorrectly assigned remoteness. At present, the only solution to this problem is to assume that all villages within a market-catchment area share the same characteristics as the (one) surveyed village in that area. Since village selection within a market catchment area was random, this should only add random measurement error to the village i observables that are used in the calibration.

Two other empirical issues to consider are more straightforward. Since agrovet fertilizer revenues and farmer expenditures are from different surveys, and the latter aggregated from a farmer level sample, we normalize each to sum to one. After doing so, we can recover η_{jv} by solving the non-linear system of equations formed using $\mathcal{J}$ agrovets and their revenue shares, as written in (21), under the normalizing assumption that $\sum_v \sum_j \eta_{jv} = 1$.[49]

After obtaining the calibrated estimates of $\hat{\eta}_{jv}$, and estimates for transportation costs, we can calculate the model-based measure of market access as:

$$\hat{\Phi}_i = \sum_{v \in \mathcal{V}} \exp\left(-\sum_b \hat{\beta}_b D_{iv}^b\right) \sum_{l \in \mathcal{J}_v} \hat{\eta}_{lv}$$

Finally, we use residual variation in sampled (and winsorized) adoption ($\hat{\mu}_i$) in each village and estimated market-access ($\hat{\Phi}_i$) to recover the relative value of the outside option of not using fertilizer:

$$\hat{\Phi}_{i0} = \hat{\Phi}_i \frac{1 - \hat{\mu}_i}{\hat{\mu}_i}$$

The estimated (log) values of $\hat{\Phi}_i$ and $\hat{\Phi}_{i0}$ are regressed on remoteness in Panel B of Appendix Table B.7, where the relationship between standardized remoteness and market access, $\hat{\Phi}_i$, is significantly negative - a one standard deviation increase in remoteness

49 This normalizing assumption is required as the probabilities within the sum in equation (21) are homogeneous degree zero in η_{jv}. Eckert 2019, Lemma pp.28 uses a similar technique to infer services trade across locations in the US, and also provides a uniqueness proof for such a system of non-linear equations.

leads to a 0.77 reduction in log market access. In contrast, there is a positive but statistically insignificant relationship between standardized remoteness and the outside option, $\hat{\Phi}_{i0}$.

2.5.3 Estimating Elasticities for Counterfactuals

We have fully calibrated the farmer's decision to adopt as a function of quality-adjusted access to markets, and shown that market access is poorer for remote villages. The results also suggest a slightly higher outside option to using fertilizer in remote markets that may be due to local suitability for fertilizer or other market conditions. To evaluate the role of trade costs in market access and output prices in the outside option, we further estimate the fundamental parameters of the productivity distribution and the production functions, with and without fertilizer.

To begin, we estimate the composite elasticity of substitution between agrovet options, ε_a, which is a function of the native Fréchet dispersion parameter, and the land (α) and labor (β) shares in the production function with fertilizer. Taking logs of η_{jv}, we get:

$$\log(\eta_{jv}) = -\varepsilon_a \log(r_{jv}) + \log(T_{jv})$$

As η_{jv} is calibrated using revenue-expenditure market-clearing conditions, there is an endogeneity problem in estimating ε_a. To address this endogeneity, we appeal to the urban and trade literature (eg. Melitz and Ottaviano 2008; Combes et al. 2012) in which larger markets induce greater entry and competition, leading to more efficient sellers and lower prices. Accordingly, we instrument for current agrovet prices using a significantly lagged (2011) population of the market catchment area that defines the location of the agrovet.[50] The regression using this instrument as well as district fixed

[50] We experimented with other instruments that attempt to leverage the insights from Berry et al. (1995) and Hausman (1996), though each has its own drawbacks in this context. BLP requires defining the relevant product characteristics for the unit defining price, and summing up the characteristics of competitors. However, the relevant set of competitor characteristics is unclear, and also some agrovets are located in isolated areas in which they are the only retailer. The Hausman technique uses the prices of the same product in other markets as instruments, but in this context, the retailers are often buying from the same distributors, rendering this instrument inappropriate.

effects yields an estimate $\varepsilon_a = 7.5$, which we use in later counterfactuals. A full set of estimates under OLS and IV is presented in Panel A of Appendix Table B.7.

Next, we estimate production parameters that are embedded in Φ_{i0}. Recall that $\Phi_{i0} = T_{i0}\left(\frac{\pi_{i0}}{\pi_i}\right)^{\varepsilon} = T_{i0} p_i^{\varepsilon_p} w_i^{\varepsilon_w}$, where $\varepsilon_p = \varepsilon\left(\frac{\alpha-\alpha_0}{\alpha\alpha_0}\right)$ and $\varepsilon_w = \varepsilon\left(\beta\frac{1-\alpha}{\alpha} - \frac{1-\alpha_0}{\alpha_0}\right)$; thus, the effects of any output price shock are a function of the relative importance of land in the production function. Further, β is important for the wage elasticity of adoption, as well as for decomposing ε_a into the native dispersion parameter ε and production parameters. In Appendix Table B.8, we detail a simple estimator for maize production with and without fertilizer, and using data from the Tanzania Living Standards Measurement Study and Integrated Surveys on Agriculture (LSMS-ISA) produce estimates for α (0.431) and α_0 (0.570). Also using the LSMS-ISA, we use reported wages and labor and fertilizer expenditures to calculate the share of labor in variable factors; β (0.75).[51] Using these estimates, it is straightforward to calculate that ε=21.9. While this may seem high, this essentially means that there is little idiosyncratic variation in quality-adjusted prices at each agrovet around the T_{jv}'s. Practically, farmers are choosing the lowest quality-adjusted price for each agrovet, with minimal other variation that distracts from prices, quality, and transport costs.

2.5.4 Agrovet Pricing and Markups

For a farmer, adoption is a function of a quality-adjusted delivered price for fertilizer at each agrovet, as well as other terms that represent the relative incentives to abstain from using fertilizer. When evaluating trade shocks, we could hold fertilizer prices fixed. However, while this may be fine for local shocks, for a large trade shock, such as a roads program, allowing retail prices and mark-ups to change is more realistic. We now derive the pricing problem for agrovets, and describe the calibration for mark-ups (similar to Berry 1994).

51 We find a similar share in a subset of our surveys in which we collected detailed labor market information, including daily wages for different tasks.

The first order condition for an oligopolist is a mark-up over marginal cost:

$$r_{jv} = \frac{\varepsilon^d_{jv}}{\varepsilon^d_{jv}+1} c_{jv}$$

where c_{jv} is the marginal cost for agrovet j in location v, and ε^d_{jv} is the elasticity of agrovet j demand with respect to its own price. Defining ε^v_{jv} as the price elasticity of revenue, we have:

$$r_{jv} = \frac{\varepsilon^v_{jv}-1}{\varepsilon^v_{jv}} c_{jv} \tag{22}$$

Defining $s_{ijv} = \frac{\lambda_{ijv|adopt}\mathbb{E}[F_i]}{\sum_{i'} \lambda_{i'jv|adopt}\mathbb{E}[F_{i'}]}$ as the expenditure share of i within jv, in the technical appendix we derive the following:

$$\varepsilon_{jv} = -\varepsilon_a + \frac{\varepsilon - 1}{\varepsilon}\varepsilon_a \sum_i s_{ijv}\lambda_{ijv} \tag{23}$$

The elasticity equation in (23) provides clear intuition regarding the spatial distribution of demand, market power and mark-ups. For each firm, $\sum_i s_{ijv} = 1$, and thus, variation in mark-ups depends on the unconditional probability of a farmer from village i choosing agrovet j in village v. When firms are "small" within the context of the market, $\lambda_{ijv} \approx 0$ for all i and the mark-up is pinned down by the substitution across agrovets through the elasticity, ε_a.

We use the agrovet-specific elasticity to solve for the revealed marginal cost of selling fertilizer by using equation (22). The predicted markups have a mean of 14.8% (median = 13.8%), which is similar to what we find in the reduced form (13%). This is notable because estimated mark-ups do not use any marginal cost information measured for retailers.

2.6 Counterfactuals

In this section, we use the calibrated and estimated parameters to evaluate counterfactuals on input and output market access. To implement the counterfactuals, we solve for a new vector of fertilizer prices that solves the first order conditions for pricing in

(22), while taking into account equilibrium changes in the farmer's problem in response to new agrovet prices and/or trade costs.

2.6.1 Experiments on Input Access

We begin by focusing on the effects of local access to fertilizer on adoption decisions. Our general hypothesis is that farmers are disadvantaged if agrovets are not close-by. We study these issues in two ways: (1) a reduction in transport costs, both on rural and main roads; and (2) agrovet entry.

Reducing Farmer-to-Agrovet Transportation Costs

To study the role of access to inputs using a realistic counterfactual, we appeal to Casaburi et al. (2013) and evaluate the effects of a 50% reduction in iceberg costs from farmer to retailer through a hypothetical roads improvement program. Such a cost reduction can also be motivated by speeds on trunk roads in Kilimanjaro being approximately 50% lower than US speeds. Figure 2.3 displays the results of this counterfactual on adoption rates (top-left panel) and log fertilizer expenditures (top-right panel) within each village. For clarity, we have grouped villages into 20 equally-sized bins of standardized remoteness, and the points in the Figure represent average adoption within these groups. For interpretation, we have also plotted lines of best fit when regressing baseline or counterfactual adoption (or expenditures) on remoteness. For these regressions, we use the unbinned raw village data.

In the top left panel, we find a large adoption effect of 36pp, which is more than twice baseline, and which alone accounts for 39% of the baseline adoption-remoteness relationship. Expenditure counterfactuals are even more pronounced, where average log expenditures rise approximately 1.5, though from an extremely low base in many cases. Nevertheless, the log-expenditure-remoteness gradient is cut by 59% in this counterfactual scenario. Thus, we conclude that holding local factors fixed, poor access to input markets contributes substantially to the reduced adoption levels in remote areas.

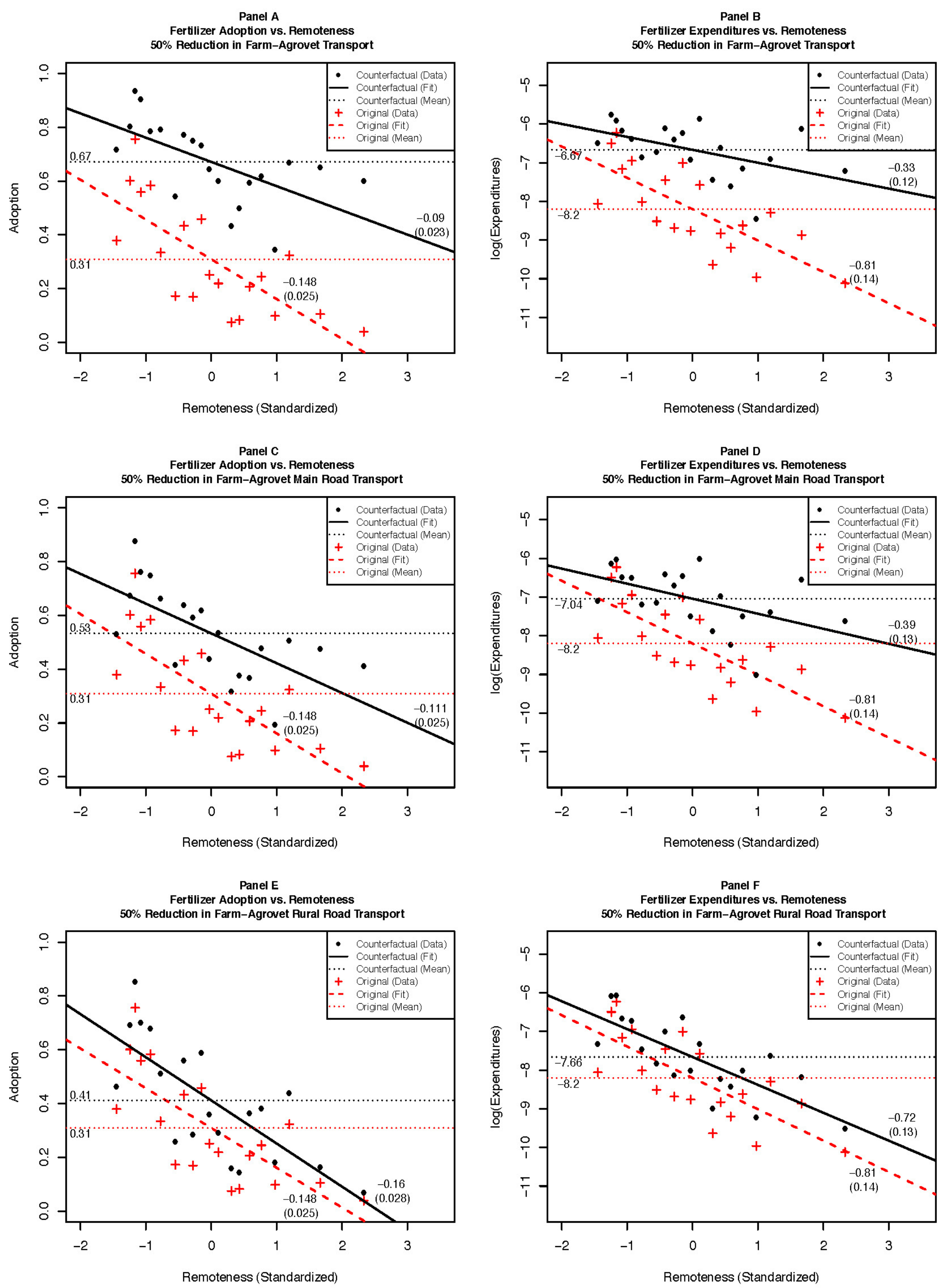
Rural roads vs. Main roads

The above counterfactual evaluates a 50% reduction in iceberg transport costs across all roads to reach agrovets, but does not distinguish between main and rural roads. This distinction may be important for a number of reasons. First, main roads may be congested, though paved, while rural roads may be uncongested but also of poor quality. Thus, both types of road improvements may be necessary to quantify. Germane to the discussion of access to retailers, critical for each farmer is whether one has to travel on rural roads to reach an agrovet, or must connect on a main road to reach an agrovet. This may be a particular problem in areas with limited entry, where farmers connect from one market area to the next using a combination of rural and main roads.

To evaluate the impact of adjusting costs on rural and main roads individually, we make an additional assumption about the iceberg trade cost, and then leverage the detail in our transport operator surveys to implement the counterfactual. We assume that the iceberg cost itself is an additive component of rural and main road costs: $\tau = 1 + t_r + t_m$. By doing so, we can derive that the iceberg cost, after reducing transport costs on rural roads by 50%, is $\tau' = 1 + \frac{t_r}{2} + t_m = 1 + (\tau - 1)\left(1 - \frac{1}{2}s_r\right)$, where $s_r \equiv \frac{t_r}{t_r + t_m}$ is the share of transportation costs incurred on rural roads. Our transport cost surveys facilitate calculating this share measure from every village to every agrovet. A similar approach can be used to isolate the impact of main road costs.

Using this approach, counterfactuals for cutting main road costs by 50%, and rural road costs by 50%, are presented in the middle and bottom rows respectively of Figure 2.3. Both counterfactuals increase adoption, but interestingly, the effect is larger when

Figure 2.3: Input Access Counterfactuals



Notes: See text for discussion of counterfactuals.

reducing main road costs. Further, rural transport costs have no appreciable effect on the remoteness gradient. The intuition for this result is that in remote areas, farmers tend to be farther from the nearest village with an agrovet, and that this travel is necessary and via a main road. Thus, to increase access to inputs, especially for rural areas, improvements in main roads are central to this policy goal. Indeed, there also appears to be a complementarity between the two effects, where the combined counterfactual from the top row of Figure 2.3 is 36pp, while the sum of the specific road-type counterfactuals is 32pp. Overall, road improvements increase adoption, but especially so for remote markets.

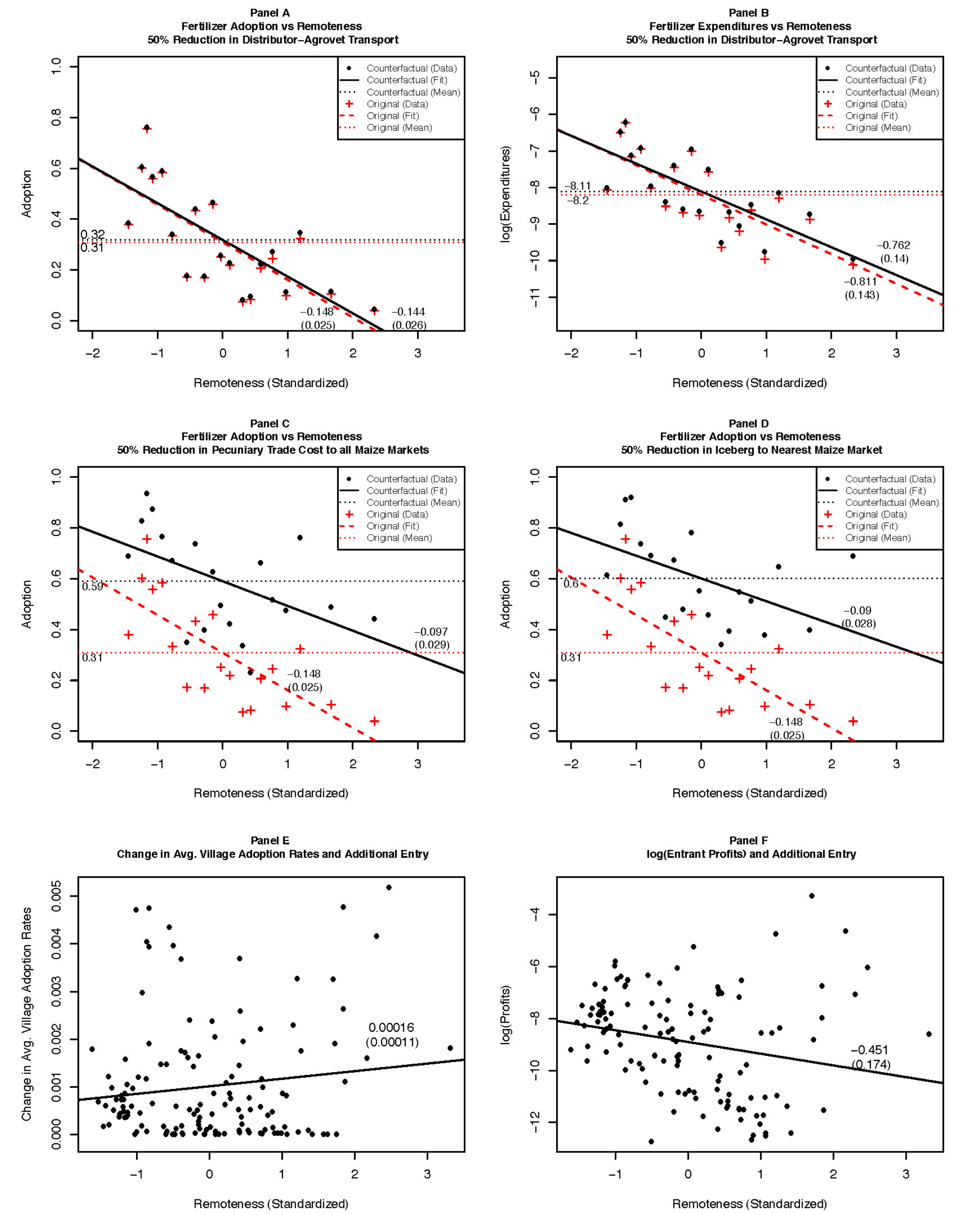
Distributor-Agrovet Costs

Next, we evaluate how the costs for retailers to source inputs from distributors affect the adoption decision. We document in our reduced-form analysis that sourcing costs rise significantly with remoteness. So, in another counterfactual that is related to input access, we halve distributor-agrovet transportation costs. The results are presented in the top panels of Figure 2.4. Adoption rises by about 1pp, or 4%, and yields a 2.7% reduction in the remoteness-adoption gradient.

Entry

An overarching question throughout the paper has been why agrovet access is worse in remote areas, and in particular, why agrovets enter intensely in other areas. While we do not present an empirical model of entry (as in papers such as Seim, 2006), we do run a simple counterfactual to examine profitability of entry and any corresponding effects on adoption rates. Specifically, we force a "median" agrovet (as defined by T_{jv} and marginal cost, within a district) to enter every village in the sample (one at a time, not simultaneously), and then measure the effects of that singular entry on adoption, and also measure the profitability of the entrant after entry. We do this for every village in the dataset, and then plot in the bottom panels of Figure 2.4 aggregate adoption (after entry) and entrant profits as a function of the remoteness of the village in which the entry took place. Clearly, profits are lower when entering more remote villages, though

Figure 2.4: Distributor, Maize Price, and Entry Counterfactuals



Notes: See text for discussion of counterfactuals.

adoption effects are higher when entering the more remote villages (though the latter is not significant). The former relationship is particularly strong, where a one standard deviation increase in remoteness reduces the profitability of hypothetical entry by 39%. Thus, while access to agrovets in remote areas would appear to improve adoption more so than entering less-remote areas, the profitability analysis supports the argument that this lack of entry in remote areas is logical.

2.6.2 Experiments in Output Access

We show in the reduced form that remote villages tend to travel farther to reach their primary market, and these travel costs can reduce the margin available to selling their maize harvest. Further, the optimally chosen "best" travel-cost adjusted selling prices are negatively correlated with remoteness. Subject to a number of caveats described below, we now examine both margins on the output side and their effects on adoption.

First, as in the reduced form, we assume that farmers optimally choose the best market to sell, while accounting for the costs of transportation. We measure the baseline best net-selling price, and then recalculate this price after halving transportation costs. The shock that is relevant to the model for each farmer is $\Phi_{i0}\left(\frac{p_{ic}}{p_{i0}}\right)^{\varepsilon_p}$, where the ratio of counterfactual prices p_{ic} to baseline output prices p_{i0} is interacted with the original calibrated parameter and raised to the price-elasticity ε_p. The results are presented in the middle left panel of Figure 2.4. We see a similar reduction in the adoption-remoteness gradient as the input market counterfactuals, and an adoption effect that is about 0.28, or 90%. Thus, this counterfactual has a smaller effect on adoption when compared with a 50% cut in farmer-retailer transport costs, but a similar effect on the gradient.[52]

[52] However, care must be taken in comparing the two. In the case of the farmer-retailer transport costs, the transport cost cut is interacted with prices and calibrated agrovet quality terms, which adds noise to the shock. That is, while a farmer might be more likely to travel to any agrovet, the transport shock is not concentrated on the agrovet that is closest or that the farmer will necessarily choose. In contrast, the best net selling price is determined by a simple calculation of the maximum net price, with no probability of choosing different option. Thus, the noise in the farmer-retailer transport shock as it relates to distance will attenuate its effect on the gradient, though still provide a sizable effect on adoption.

With this same caveat in mind, we now assume that farmers sell at their closest market, and experience a 50% reduction in iceberg costs to that market, as estimated by the agrovet choice problem. That is, the farmer must ship τ_{im} units of maize to the market to effectively sell one unit. Thus the transport-adjusted selling price that we use for each village above when calibrating adoption decisions is equal to $p_i = p_m/\tau_{im}$, where p_i is net selling price to farmer i and p_m is the price at the primary market for that village. To examine the impact of output market access on adoption, we now run an experiment cutting the iceberg costs to reach output markets by 50%. The results from this counterfactual are presented in the middle-right panel of Figure 2.4. Here, adoption almost doubles, and the adoption-remoteness gradient falls by about 40%.

Overall, access to output markets appears to be an important component of the input adoption decision, and these effects deserve more detailed attention in follow-up work.

2.7 Conclusion

We collect detailed data on transportation costs, input and output prices, input usage and maize sales along the supply chains for maize and fertilizer in all 1,183 villages in the Kilimanjaro and Manyara regions of Tanzania. We find that there is meaningful price dispersion, especially when accounting for travel costs. Access to retailers for inputs and buyers for output is much lower in remote regions, and consequently farmers in remote villages are much less likely to use fertilizer or sell output. Counterfactuals suggest that lowering transportation costs via road upgrading would substantially reduce the gradient between input usage and remoteness.

An important question is whether our results generalize to other settings. To provide some suggestive evidence on this question, we use secondary datasets and a dataset of prices we collected in our study area to examine how patterns compare between Northern Tanzania and other African countries. First, we examine how price dispersion in Northern Tanzania compares to a set of 1,512 markets in 56 African countries. Using two approaches, we find that the degree of observed price dispersion in Northern Tanzania is comparable to other countries. Second, we use data from World Bank LSMS-ISA panel surveys to study how remoteness affects fertilizer adoption in other African countries.[53] Using both measures of remoteness available in the dataset (distance to the main market, and distance to a population center), we find a negative association between remoteness and technology adoption. Finally, we compile some statistics on the state of road infrastructure in other countries in the East African region (Table B.12), and find that Tanzania is about average. The evidence therefore suggests that Northern Tanzania is not atypical of the region.

The results of our counterfactual simulations as well as the presence of similar patterns in other countries lead directly to the question of policy implications. Many African countries have experimented with input subsidies and these have had large adoption effects by directly lowering the delivered price of fertilizer even though the transport cost may have been unaffected (depending on retailer entry response to the program). However, most farmers fail to graduate out of the subsidy, perhaps in part because the market access issues remain unresolved, and therefore, inputs continue to be unprofitable at market prices. Our findings suggest that policies that lastingly affect input and output prices faced by farmers can have sustained effects. Initiatives to organize farmers into cooperative groups that enable them to defray the total costs of transportation over a large number of buyers may also be helpful.

[53] The countries included here are Ethiopia, Niger, Nigeria, Malawi, Tanzania, and Uganda.

Part 3

Is Corruption Compensation? Evidence from Local Public Office in India

3.1 Introduction

Around the world, the responsibility for administering welfare programs lies increasingly with local governments. While such decentralization may have large benefits, it may also overburden administrative capacity and lead to an under-provision of government benefits. If under-provision arises because local officials have little incentive to run the program effectively, some have argued that the opportunity to collect rents from the program could perversely realign incentives towards good service (Ravallion, 2018). Similar to classic theoretical models in the 'greasing the wheels' corruption literature, voters might grudgingly allow politicians to take informal compensation for themselves in exchange for improved service delivery (Huntington, 1968; Leff, 1964; Lui, 1985).

But testing the corruption-as-compensation hypothesis is challenging. The hypothesis assumes voters have some means to observe both corruption and the performance it compensates. Though the assumption may be plausible in many situations, there are few instances where it is directly verifiable. Even if that condition is met, applying the test in practice would also require that the researcher is able to detect what the voters observe. Corruption is always difficult to measure, and in this context even measuring "performance" is a challenge. Most prior work in the literature has either focused on other potential explanations for corruption, or has studied officials whose performance yields amorphous returns like local economic growth or perceptions of efficient local governance. As there is usually no unambiguous link from corruption to performance it is hard to confirm whether one is compensation for the other. As a result, the burgeoning literature on corruption has typically been unable to test this channel convincingly.

This paper tests the corruption-as-compensation channel in a context where both performance and corruption are measurable. We study how local politicians in the

Indian state of Uttarakhand implement the world's largest anti-poverty program. The National Rural Employment Guarantee Scheme (NREGS) funds short-term make-work jobs building public works within the village. Village council presidents play a key role in both bringing jobs to their villages and in allocating these jobs between constituents. Their success in generating public works jobs is a measure of performance. A clearly related measure of corruption in this context is the number of jobs presidents allocate to their own households. While this behavior is not illegal, such self-dealing meets the most common definition of corruption as the misuse of public office for private gain (Svensson, 2005).

The program's unusual level of transparency allows us to create a unique dataset that directly captures an individual president's self-dealing of NREGS benefits as well as her own performance in generating NREGS employment for the village. Each employment spell is published on a publicly available government website in close to real time and identifies the recipient by name and location. We scrape millions of these reports, covering over 90 percent of rural households in Uttarakhand.[54] We match these NREGS reports to the election returns for candidates competing in thousands of village council elections. This dataset shows how winning candidates allocate NREGS jobs to their own household as compared to those of typical villagers. Since election winners are likely to differ from typical villagers on a number of observable and unobservable characteristics, we compare the labor allocations of winners and runners-up in close elections decided by a few votes. Our regression discontinuity design ensures that the labor quota of the runner-up is a good counterfactual for that of the winner had she remained out of office. We also observe the NREGS employment the politician creates for the entire village, a direct measure of program performance.

Our context is well-suited to study the corruption-as-compensation channel. To be plausible, voters need to be able to observe and punish corruption, but instead allow self-dealing proportional to performance to incentivize politicians to put in more unobservable effort. In contrast to most other contexts, self-dealing in our case is plausibly visible to voters through the website at any point and in close to real time. Figure 3.1

[54] Households can be registered for the program without ever having worked on it.

Figure 3.1: Comparing Close Winners and Losers on the Official NREGS Site

Won by 3 votes

Job card

MAHATMA GANDHI NATIONAL RURAL EMPLOYMENT GUARANTEE ACT

Job card No.:	UT-[redacted]	Family-Id: 82
Name of Head of Household:	[redacted]	
Name of Father/Husband:	[redacted]	
Category:	OTH	
Date of Registration:	1/8/2008	
Address:		
Villages:	[redacted]	
Panchayat:	[redacted]	
Block:	BHAISIACHHANA	
District:	ALMORA(उत्तराखंड)	
	NO	Family-Id: 82
Epic No.:		

Details of the Applicants of the household willing to work

S.No	Name of Applicant	Gender	Age	Bank/Postoffice
1	[redacted]	Male	33	District Co-operative Bank
2	[redacted]	Female	30	UTTRANCHAL GRAMIN BANK

Signature/Thumb impression of Applicant — Seal & Signature of Registering Authority

Period and Work on which Employment Given

S.No	Name of Applicant	Month & Date from which employment requested	No of Days	Work Name	MSR No.	Total Amount of Work Done	Payment Due
1	[redacted]	21/02/2015	14	[redacted] bhetadangi se khan se mandir tak marg nirman 1.0	645	2184	0
	Sub Total FY 1415		14			2184	0
2	[redacted]	25/08/2015	14	[redacted] harju mandir se devithan tak marg nirman karya 1.0	178	2254	0
3	[redacted]	03/12/2015	14	Anganwadi/Other Rural Infrastructure,BHAISIACHHANA	404	2254	0
4	[redacted]	30/12/2015	14	Anganwadi/Other Rural Infrastructure,BHAISIACHHANA	496	2254	0
	Sub Total FY 1516		42			6762	0
5	[redacted]	03/08/2016	14	Anganwadi/Other Rural Infrastructure,BHAISIACHHANA	439	2436	0
6	[redacted]	03/08/2016	14	Anganwadi/Other Rural Infrastructure,BHAISIACHHANA	439	2436	0
7	[redacted]	23/08/2016	14	Anganwadi/Other Rural Infrastructure,BHAISIACHHANA	522	2436	0
8	[redacted]	23/08/2016	14	Anganwadi/Other Rural Infrastructure,BHAISIACHHANA	522	2436	0
	Sub Total						

Lost by 3 votes

Job card

MAHATMA GANDHI NATIONAL RURAL EMPLOYMENT GUARANTEE ACT

Job card No.:	UT-[redacted]	Family-Id: 14
Name of Head of Household:	[redacted]	
Name of Father/Husband:	[redacted]	
Category:	OTH	
Date of Registration:	1/8/2008	
Address:		
Villages:	[redacted]	
Panchayat:	[redacted]	
Block:	BHAISIACHHANA	
District:	ALMORA(उत्तराखंड)	
	NO	Family-Id: 14
Epic No.:		

Details of the Applicants of the household willing to work

S.No	Name of Applicant	Gender	Age	Bank/Postoffice
1	[redacted]	Male	35	
2	[redacted]	Female	30	

Signature/Thumb impression of Applicant — Seal & Signature of Registering Authority

Period and Work on which Employment Given

Note: From the publicly accessible official NREGS website at nrega.nic.in. Identifying information has been redacted for this image. The record of Employment Given is blank for the job card on the right because none was given.

provides an example. It shows the job card record for the council president alongside the closest runner-up in the election with household location down to the village level and a list of all adult household members. Below that information the website provides employment and wage details on all job spells under the program with their exact dates and the project name. In the example, Figure 3.1 shows that the household of the council president has received dozens of days of labor, whereas the runner-up—who lost by just 3 votes—has received no jobs at all. While voters in other contexts may be aware that money for welfare benefits goes missing, voters in Uttarakhand can directly observe how much money flows into the president's own bank account.

Even villagers that do not access the website may often be able to observe self-dealing through other means. Since NREGS is a public-works program, worksites around the village are directly visible to constituents. Villages in Uttarakhand are small, with 80 percent of villages having fewer than 1000 inhabitants. This means that villagers are very likely to know the council president personally and can monitor his actions more easily than in larger polities. Under NREGS, villagers also have the right to audit the physical muster rolls of the program. Additionally, accountability is likely to be much higher in this setting than is typical of local leaders in developing countries. Village council presidents in Uttarakhand are directly elected through competitive local elections, and by design our sample focuses on close elections determined by a small number of votes. This restriction excludes villages where elite capture or nonexistent political opposition have rendered the president unaccountable to voters.

To further establish that corruption-as-compensation is plausible in our context, we ask the village council presidents themselves using an original survey of about 200 presidents. They answer a number of questions about their work and how they run NREGS but without being told that we are studying self-dealing or monitoring their NREGS allocations. They report that NREGS takes a lot of effort to implement, and more than half of survey respondents claim that presidents who create more NREGS jobs for the village would be expected to also take more NREGS jobs for themselves. The majority of village council presidents therefore rationalize self-dealing by claiming villagers accept corruption-as-compensation.

Overall, our test of the corruption-as-compensation mechanism therefore occurs in a highly plausible context based on qualitative and descriptive evidence. We now turn to testing this channel empirically. We first establish that there is a sizeable amount of corruption. The winners of close elections receive nearly 3 times as many days of labor as losers in the year after the election. On average individuals who visit the website would therefore see that the list of job spells for the president is three times longer than their own in the year after the election, an easily detectable difference even when just casually comparing job card records. Consistent with self-dealing as compensation for good performance, we find a correlation between the extent of self-dealing and the average NREGS benefits enjoyed by other households in the village.

But the corruption-as-compensation hypothesis does not withstand more rigorous scrutiny. We find that self-dealing declines over time while benefits for other villagers remain roughly constant or even rise. Tellingly, the overall decline in corruption is entirely driven by villages best able to monitor the president. If villagers accept self-dealing that is proportional to the overall NREGS performance to incentivize presidents to put in unobservable effort, then the "corruption contract" should be most likely to arise in villages where voters can perfectly monitor performance and self-dealing. We proxy for monitoring capacity with the distance to an internet cafe, which may grant direct access to the NREGS website to villagers and may be correlated with closer scrutiny from media, local activists, and the political opposition. Stakeholders living far from a cafe may have to rely more strongly on physical muster rolls that can be more easily hidden or manipulated. We find that self-dealing falls to zero in villages close to a cyber cafe but shows no decline at all in villages far from a cafe—the opposite of what is predicted by corruption-as-compensation. By contrast, NREGS benefits for villagers other than the president are equally high in both sets of villages—more evidence that good performance is possible even in the absence of self-dealing.

Instead, the evidence is more consistent with imperfect monitoring. While monitoring NREGS in Uttarakhand is plausibly easier than in other contexts, it may not be perfect in areas without reliable access to the internet. Presidents may therefore be able to hide self-dealing in villages without internet cafes as long as it remains a small amount of the total NREGS benefits. This would explain why self-dealing appears to be proportional to the NREGS implementation quality in the village. In contrast, in areas with better monitoring capacity, pressure from villagers seems to lead to a complete eradication of self-dealing over time.

Our results contribute to the 'greasing the wheels' literature. Several theoretical models predict that corruption may be a second-best solution (Huntington, 1968; Leff, 1964; Lui, 1985). Typically, these models focus on interactions between firms and bureaucrats, showing that corruption can increase allocation efficiency if the most efficient firms are willing to pay the highest bribes. But while these models focus on the interactions between actors at the micro level, the existing empirical evidence has mostly been confined to analyzing their predictions at the macro level using cross-country datasets (Fisman and Svensson, 2007; Méon and Sekkat, 2005; Méon and Weill, 2010; Wei, 2000). Two rare exceptions are Mironov and Zhuravskaya (2016) and Weaver (2018). Mironov and Zhuravskaya (2016) reject a 'greasing the wheels' explanation, since procurement contracts in more corrupt Russian localities are allocated to less efficient rather than more efficient firms. Weaver (2018), on the other hand, finds that the allocation of health bureaucracy jobs to the person willing to pay the highest bribe leads to higher-quality hires than decisions based on a knowledge test. Our paper therefore contributes to the existing literature in two ways. First, we provide new micro-level evidence on the topic to a very small empirical literature. In contrast to Weaver (2018) and despite politicians' claims in our original survey, we find little evidence for a corruption-as-compensation channel in the empirical analysis.[55] Second, our paper extends the existing theoretical and empirical literature to interactions between citizens and local politicians and focuses on service delivery rather than allocative efficiency.

More broadly, our paper contributes to the literature on corruption (Avis et al., 2018; Bertrand et al., 2007; Campante and Do, 2014; Di Tella and Schargrodsky, 2003; Ferraz and Finan, 2008; Niehaus and Sukhtankar, 2013b; Olken, 2007; Reinikka and Svensson, 2004).[56] The existing literature tends to focus on leakages, where it may often be unclear where exactly the money goes, for example whether bureaucrats or politicians take the biggest cut. One strength of our context is that we directly observe that payments from NREGS land directly in the bank account of the politician, making it unusually clear where exactly the money disappears to, at least initially.[57] This means that specific

[55] This makes our results more consistent with the macro literature, which overwhelmingly finds no support for the hypothesis.

[56] For a literature overview see Aidt (2003); Bardhan (1997); Olken and Pande (2012); Svensson (2005).

[57] We do not find any evidence that politicians are misusing NREGS to provide preferential access to their extended family or other members of the village council. While we cannot rule out that some of the NREGS benefits politicians allocate themselves are used to pay off supporters or other officials in the system, directly providing them with NREGS benefits would seem to be an easier way of making those payments in many cases.

NREGS benefits can be directly attributed to an individual politician, and are therefore plausibly taken as a direct indication of the politician's greed. Given the nature of the self-dealing, it should also be unusually salient that the NREGS jobs that the president received could have gone to other villagers. These features also allow us to contribute to the small literature on unofficial returns to office in developing countries.[58] Our results suggest that even in a context where transparency and accountability are plausibly high, corruption can only be fully eliminated if citizens have the capacity to monitor behavior. Having an internet cafe nearby is likely to provide local stakeholders with easier access to the internet, allowing for more effective monitoring of NREGS allocations. While Muralidharan et al. (2016) have shown that the reforms that linked NREGS benefits to biometric identification information have improved targeting substantially, we show that self-dealing persists.[59]

3.2 Background

3.2.1 Village Council Elections in Uttarakhand

Village council presidents in Uttarakhand are directly elected every five years, most recently in 2014. Local elections are run by the State Election Commission of Uttarakhand, an independent body that sets the election date and monitors nominations and campaigns. Elections are widely perceived to be free and fair. Roughly 90% of elections for council president are contested, and over 90% of respondents to the 2006-2008 Rural Economic Development Survey say they feel free to vote as they desire. Uttarakhand

[58] Existing studies typically focus on constructing broad measures of wealth changes for politicians in developed countries since direct information on self-dealing is often unavailable (Albertus, 2019; Baturo, 2017; Diermeier et al., 2005; Eggers and Hainmueller, 2009; Klašnja, 2015; Querubín and Snyder, 2013; Reinikka and Svensson, 2004). But those returns could come from a variety of sources. The prestige of public office can bring customers to a politician's side business or yield invitations for paid speeches, for example, which are less inherently unethical income opportunities than self-dealing welfare benefits. For the small literature on developing countries see Asher and Novosad (2018); Fisman et al. (2014); Foltz and Opoku-Agyemang (2015).

[59] Consistent with an improvement in the implementation of NREGS, our estimates of self-dealing in a context with plausibly high levels transparency and accountability are substantially lower than the Niehaus and Sukhtankar (2013a) estimates from the early days of NREGS implementation when off-the-books corruption was much easier to pull off.

is also somewhat uniquely free of the "politics of fear" (as modeled by Padró i Miquel, 2007) that arise in other parts of India from caste and religious conflict. The state has only a tiny population of the so-called Other Backward Classes, and there have not been large attempts to create a unified political identity among more disadvantaged castes or the relatively small Muslim community.[60] The support of an ethnic community is thus less likely to insulate the president in Uttarakhand than elsewhere.[61]

Aside from elections the president faces checks on her authority while in office. She is in theory monitored by the other members of the village council (gram panchayat), who are independently elected. The voters themselves can by law file a no-confidence motion against the president if one-quarter of adults in the village sign a petition against her. A district bureaucrat then convenes a village meeting where the president can be removed from office if a majority of villagers favor her recall.

3.2.2 National Rural Employment Guarantee Scheme

NREGS, the National Rural Employment Guarantee Scheme, is the world's largest public-works program. The primary goal of the scheme is to provide a flexible safety net for rural households in times of need by offering an income transfer conditional on the willingness to perform manual labor at the minimum wage (Zimmermann, 2018). There are no further means tests (Dey et al., 2006; Government of India, 2018). Most projects are routine tasks, such as clearing bushes or digging holes, that do not create substantial public investment.

In theory, NREGS guarantees every rural household up to 100 days of public employment per year at the minimum wage, on demand whenever requested by the household.[62] But in practice the program is supply- rather than demand-driven. In Uttarakhand as

[60] See e.g. The Indian Express, 'Uttarakhand elections: Across the border; next door to UP, new caste calculus', February 15, 2017.

[61] In our survey of village council presidents in Uttarakhand, very few presidents report that they relied on a particular caste, party or religious group in their election campaign. See Figure 3.4 below. This is in stark contrast to India as a whole, where large fractions of candidates for the council presidency report in the 2006-2008 Rural Economic Development Survey (REDS) that they relied on a caste group for support.

[62] See Berg et al. (2012), Imbert and Papp (2015) and Zimmermann (2018) for analyses of the economic impacts of the program.

well as in other Indian states, excess demand for NREGS jobs is common. Households can only get employment when it is made available, rather than taking up work when they may need it most (Dutta et al., 2012; Mukhopadhyay et al., 2015). For example, many households report having to wait passively for jobs to be provided rather than actively applying for work.[63]

The necessary rationing of employment due to excess demand gives the village council president a key role in the allocation of jobs among households in the area she governs (which is also called the gram panchayat, or panchayat for short). Aside from registering households and proposing local projects to block and district officials, the village council and president also effectively control the allocation of jobs. A worker who wants NREGS labor must apply at the council office. Though in theory a joint decision by the entire council, in practice council presidents make the decisions either themselves or jointly with their spouse.[64] These allocations are then submitted to higher-level officials, who approve the wage payments. Since there are never enough jobs to meet villagers' needs the council can exercise discretion in how jobs are allocated.

To create transparency the government now requires all NREGS related information to be entered into a software application called NREGASoft. The system contains multiple modules to track different aspects of the scheme, such as employment demanded by workers and jobs allocated, proposed and approved works projects, as well as modules for managing funds and labor budgets (Government of India, 2013). To cut down on corruption the Indian government opens bank accounts for NREGS beneficiaries and directly transfers wages for completed work into those accounts, cutting out middlemen who might pocket part of the payment. Additionally, job cards are now linked

[63] Newspaper coverage on Uttarakhand is typically very negative, noting the low job availability and the absence of a well-functioning planning process at gram panchayats. Both are symptoms of a supply-driven NREGS implemenation with a centralization of decision-making power in the village council president. Only 3 to 5 percent of households in Uttarakhand get the full 100 days of employment. See e.g. Live Mint, 'MGNREGA, a cash transfer scheme?', March 18, 2013; Mainstream Weekly, 'Working of NREGA Voices from Panchayats', April 2, 2009; Financial Express, 'The state of MGNREGA performance: If inefficient states perform at par, huge gains are possible', March 14, 2017. See also India Spend, 'The Whys and Whats of India's Rural Jobs Scheme', November 4, 2014.

[64] According to our own survey of council presidents.

directly to each individual's Aadhar number, a national identification number linked to biometric markers. These changes have been shown to improve household benefits from the program, likely because it is more difficult to engage in hidden corruption through made-up work spells or underpayment of wages (Muralidharan et al., 2016).

In short, any NREGS payment must be reflected in the online system. The resulting records are fed in real-time to a publicly available website.[65] That makes off-the-books corruption, as documented in the early days of NREGS[66], more difficult than open self-dealing. Aside from letting villagers, local media, NGOs and political challengers monitor the council president, this website is also the source of data for this study.

3.3 Data and Research Design

3.3.1 Data

We use publicly available administrative data on NREGS employment that we scraped from the official NREGS website, which is maintained by the Government of India. The dataset contains digital versions of the paper trail that is mandated by the scheme, which provides us with data on NREGS employment at a highly disaggregated level. Every registered job card has an online record with the details of the job card holder, typically the household head, and his or her family members. The household's district, block, panchayat, and village are recorded. The record also includes the name, gender, and age of every household member registered to work, as well as the start date and length of each job spell, wages paid, and the name of the project they worked on. Additionally, we have information on the name of the household head's father or husband, the household's broad caste category, and the date of initial registration for the job card that made the household eligible to work under NREGS.

We merge the NREGS data to publicly available information from the local election for the president from June 2014.[67] The election dataset contains the name, closest male

[65] https://nrega.nic.in

[66] See Afridi and Iversen (2014); Niehaus and Sukhtankar (2013a,b).

[67] While the next local election took place in 2019, consistent detailed data on election results is not yet available. We therefore cannot analyze whether the presidents in our dataset participated in the 2019 election.

relative, and vote count of the winner and runner-up of each gram panchayat election. This information allows us to match the winner and runner-up to their NREGS job card profile. We attempt to match these top-two candidates for all elections decided by a margin of 7.5% or less, dropping any cases in which there is no unique match. We link this dataset to the 2011 Indian Census by collapsing statistics measured by census village to the level of the panchayat. We draw several variables from the Census, most importantly a categorical variable that reports whether the distance of the village to the nearest cyber cafe is less than 5 kilometers, more than 10 kilometers, or somewhere in between.[68] We aggregate the census data to the level of the panchayat, then merge to our linked job card-election dataset.

To better understand the wages and motives of council presidents, we surveyed a sample of them by phone. We matched the winning candidates in our sample to contact information posted on the website of the Uttarakhand Ministry of Panchayati Raj. We assigned a random ordering to this sample and hired contractors in India to work down the list making calls in the month just before the 2018 monsoon season. The contractors made as many calls as possible in this period, yielding a final sample of 207 complete or partial interviews.[69] The response rate was roughly 30 percent, where nonresponse arose mainly because our interviewer could not connect (likely because the phone was off or out of cell phone range). Conditional on someone picking up the response rate was close to 100 percent. The connection issues seem transient—several of those who could not be initially contacted were successfully interviewed when called later. We detect no statistically significant difference on observables between our survey sample and the presidents who were not surveyed, making differential non-response less likely to be a concern.

Table 3.1 reports summary statistics for four samples: all candidates that were

[68] In the appendix we also use as controls the Census reports of distance to the district and sub-district headquarters, the literacy rate, and the fraction of the population classified as scheduled caste or tribe. We also use village geo-coordinates to measure the distance to the state capital, the distance to the state's border with Uttar Pradesh, and 2011 night-time light emissions.

[69] During the first phase of the survey we had to modify the wording of some questions after our interviewers reported that respondents did not understand the original wording. As a result we do not have 207 responses for some questions.

successfully matched to their NREGS records, the subset within the bandwidth used to estimate our main specification, the set of presidents within the full matched sample, and the subset in our survey sample. The samples are broadly similar on all characteristics except those that differ by construction (e.g. all winning candidates have a positive vote margin). As noted above there are no significant differences between the survey sample and the unsurveyed presidents in the matched sample. In particular, the two samples earn almost identical amounts of NREGS labor in the year after they become president (2015).

The one difference between samples that is both clear and not mechanical is that winning candidates worked somewhat more days of labor in 2015 (after the election) than the samples that include both winners and runners-up. This difference foreshadows our main result that winners receive more labor than losers.

Table 3.1: Descriptive Statistics and Sample Sizes

	Winners and Runners-Up		Winners Only	
	Full Matched Sample	In Bandwidth	All	Surveyed
Winner	0.54	0.44	1.00	1.00
	(0.50)	(0.50)	(0.00)	(0.00)
Female	0.47	0.47	0.48	0.43
	(0.50)	(0.50)	(0.50)	(0.50)
Scheduled Caste/Tribe	0.21	0.21	0.20	0.21
	(0.40)	(0.41)	(0.40)	(0.41)
Vote Margin	0.85	-4.57	13.96	14.65
	(19.77)	(16.64)	(11.63)	(14.37)
In Bandwidth	0.82	1.00	0.66	0.65
	(0.39)	(0.00)	(0.47)	(0.48)
Surveyed	0.13	0.10	0.23	1.00
	(0.33)	(0.30)	(0.42)	(0.00)
Days of labor (2015)	40.10	36.30	57.19	57.40
	(40.66)	(38.65)	(42.15)	(39.00)
Days of labor (2013)	23.13	23.95	22.08	26.59
	(34.55)	(35.13)	(34.27)	(37.71)
Observations	1650	1351	887	207
Panchayats	1148	1003	887	207

Note: Each cell gives the mean and standard deviation of a characteristic of candidates for council president (rows) when conditioned on a specific subsample (column). “Full Matched Sample” is the set of all candidates we are able to find in the job card data. “In Bandwidth” is the subset whose vote margin falls within the bandwidth of our main specification. “All” is the subset of winning candidates within the full matched sample. “Surveyed” is the subset we were able to interview for our survey of council presidents.

3.3.2 Research Design

We estimate the causal effect of being the council president using the regression discontinuity induced by close elections. We restrict our sample to the winner and runner-up in each election. Let i be one of these two candidates in the election for panchayat p. Our running variable is the vote margin, which we define as

$$[Margin]_{ip} = \begin{cases} \text{[Winner Votes]} - \text{[Runner-Up Votes]} & \text{if } i \text{ won election in } p \\ -\Big(\text{[Winner Votes]} - \text{[Runner-Up Votes]}\Big) & \text{if } i \text{ lost election in } p \end{cases}$$

This definition generates a discontinuity at zero.[70] For our research strategy to identify a causal effect, any unobserved factors that are correlated with being council president must be continuous in the margin of votes. We therefore zoom in on a small window around the cutoff and control for a linear spline in the vote margin.

The continuity assumption holds if political candidates standing for election cannot perfectly manipulate the number of votes they receive. We verify the assumption with placebo tests using pre-determined outcomes that cannot be changed by the election.

We estimate:

$$[Outcome]_{ip} = \pi_0 + \pi_1 [Margin]_{ip} + \pi_2 [Margin]_{ip} \times [Win]_{ip} + \beta [Win]_{ip} + \nu_{ip} \tag{24}$$

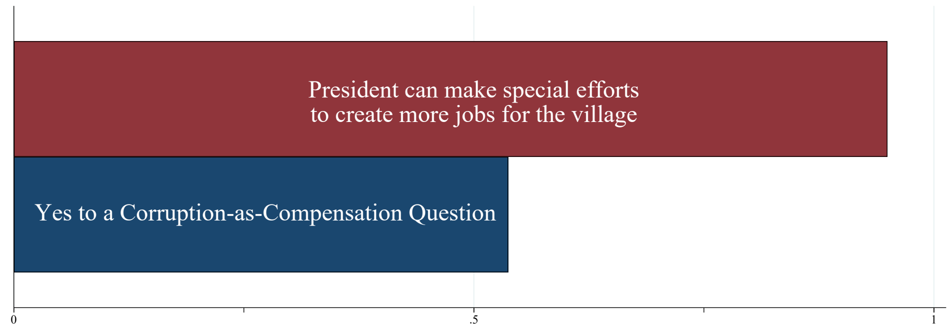
where $[Win]_{ip}$ is a dummy for whether $[Margin]_{ip} > 0$ and $[Margin]_{ip}$ is restricted within a bandwidth centered on 0. We use the method suggested in Calonico et al. (2014) to choose the optimal bandwidth for our main specification, but also explore the robustness of our results to a wide range of alternative bandwidths.

3.4 Is Corruption-as-Compensation Plausible?

Observers like Ravallion (2018) have suggested that given its many administrative challenges, NREGS may in part be sustained by the prospect of corruption. He sketches

70 In practice, the official election law for Uttarakhand breaks ties by randomly drawing the name of the winner among candidates with the same number of votes and then adding a vote to the winner's vote count in the election records.

Figure 3.2: Presidents Believe their Performance Matters and that Villagers Agree It Should Be Compensated with More NREGS Jobs



Note: Each bar shows the fraction of respondents who agree (as per the survey of council presidents). See text for details on the Corruption-as-Compensation Questions.

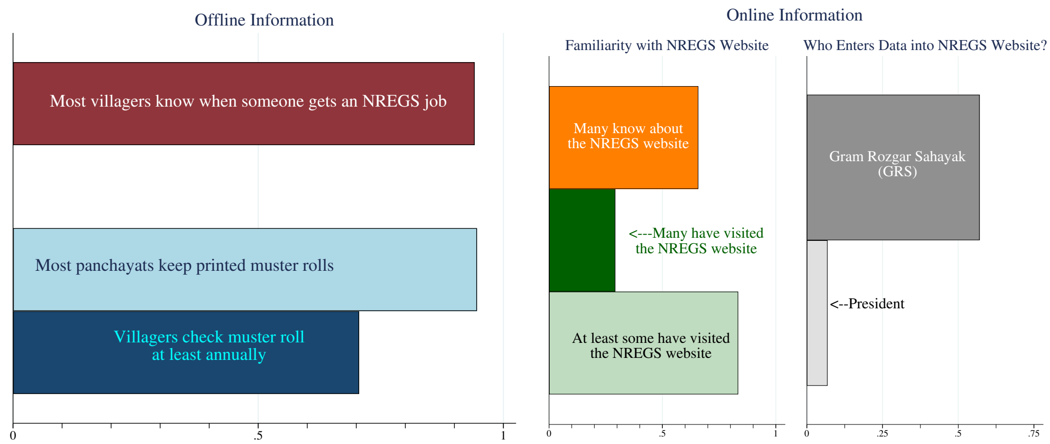
a model in which "There is a pecuniary benefit to the official that naturally depends on the level of employment. We can think of this as the official's cut on the wages paid." Our survey of council presidents gives reason to believe this hypothesis. Three questions were specifically designed to get presidents' self-reported opinion on the plausibility of a corruption-as-compensation channel. We first asked whether a president can make special efforts to "bring back more jobs for their constituents." As the top bar of Figure 3.2 shows, nearly all presidents agree with that statement. Consistent with the anecdotal accounts from Section 3.2, village council presidents are the key player in the introduction of NREGS. To measure whether these efforts should be rewarded, we measure whether a president answers yes to either of two "corruption as compensation" questions:

> Do you believe a typical person in a village in your block would agree that a Gram Pradhan who makes those efforts deserves a few more NREGS jobs than the typical household in the village?

and

> Suppose a Gram Pradhan in a typical village in your block manages to get a new worksite approved for his village. Would people in the village expect the Pradhan's household to do NREGS labor on the newly approved worksite?

Figure 3.3: Council Presidents Believe Villagers are Well-Informed about NREGS Allocations



Note: Each bar shows the fraction of surveyed council presidents whose response indicates the statement given. The "Gram Rozgar Sahayak" is a village technical assistant employed directly by the state to help village councils with the online NREGS system.

The bottom bar shows that a majority of council presidents agree with one or both of these statements.

In our survey, village council presidents also claim that NREGS is transparent and that they feel accountable to voters. Figure 3.3 shows their responses to a number of survey questions asking whether villagers could reasonably be expected to be aware of NREGS allocations. Almost all respondents tell us that most villagers would know when someone in the village gets a NREGS job (left-hand panel of Figure 3.3). Every panchayat is supposed to keep paper records of the official "muster rolls," which list all the workers who have received NREGS jobs on every project. Nearly all presidents say most panchayats in their block keep these records, and over two-thirds of presidents report that a villager will ask to check the muster rolls at least once per year (left-hand panel of Figure 3.3).

We also ask about the NREGS website. Over two-thirds of presidents believe most of the villagers in their panchayat know about the NREGS website (see right-hand panel of Figure 3.3). Although few presidents say most villagers in their panchayat have actually visited the site, over 80 percent say that at least a few have.

Lastly, we ask about the role of the Gram Rozgar Sahayak (GRS), the village technical assistant. Though the president decides how to allocate NREGS jobs, the majority of presidents report that the GRS actually submits those decisions to the NREGS online system, and almost no presidents actually enter the information themselves (see right-hand panel of Figure 3.3).[71] There is thus at least one district-level bureaucrat aware of the president's actions and in principle able to share her knowledge with villagers or district officers.

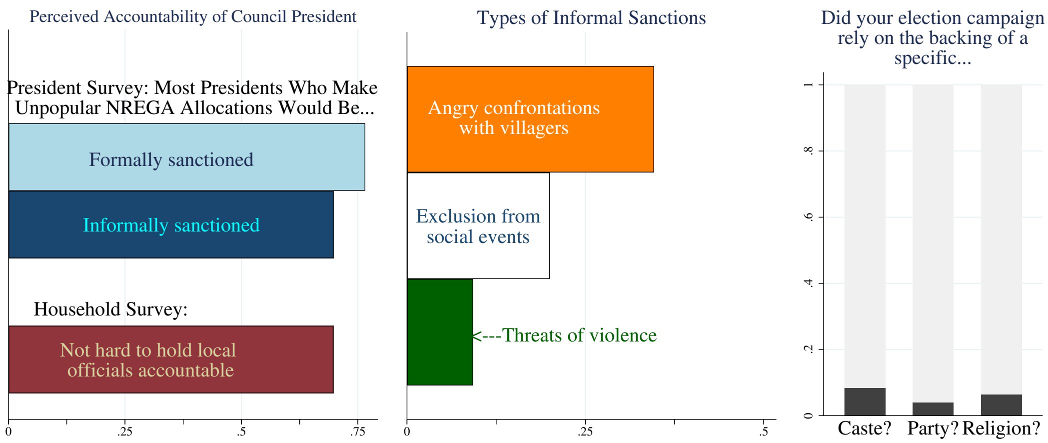
Figure 3.4 shows presidents' responses to questions about the accountability of a typical president to her villagers. The survey asked each respondent whether a typical president in her block would be formally or informally sanctioned for making NREGS allocations that are unacceptable to her constituents. The overwhelming majority said most presidents would be sanctioned (see left-hand panel of Figure 3.4). As the figure shows, presidents' beliefs on accountability are comparable to citizens' beliefs in the 2006-2008 REDS household survey that it is not difficult to hold local officials accountable. When asked what types of informal sanctions were likely, the most common answers were that the president would be confronted by angry villagers, suffer exclusion from social events, or even be threatened with violence. It is also unlikely that presidents maintain power by appealing to an ethnic group willing to ignore bad governance for the sake of keeping its own in power (e.g. Padró i Miquel, 2007). Less than 10 percent of council presidents in our sample report having relied on the support of a caste, religion or political party in their campaign for office.[72]

Overall, the survey responses of the village council presidents therefore suggest that a corruption-as-compensation channel is plausible in the implementation of NREGS. Respondents claim that NREGS allocations are transparent, that presidents who make unpopular NREGS allocations would be held accountable, and that presidents who create more jobs would be expected to also take more jobs for themselves. At face value

71 Most of the other presidents report that the decisions are submitted by the village secretary or some other assistant.

72 This is in stark contrast to India as a whole, where large fractions of candidates for the council presidency report in the 2006-2008 Rural Economic Development Survey (REDS) that they relied on a caste group for support.

Figure 3.4: Council Presidents Feel Accountable to Villagers and Do Not Rely on Specific Subgroups for Support



Note: Each bar shows the fraction of respondents who agree with the statement. Household-level responses are from the household module of the 2006–2008 Rural Economic Development Survey. All other responses are from the survey of council presidents.

these responses imply that voters are aware of NREGS allocations and could punish a politician who is too greedy, but will not do so as long as self-dealing is proportional to the total program benefits created for the village.

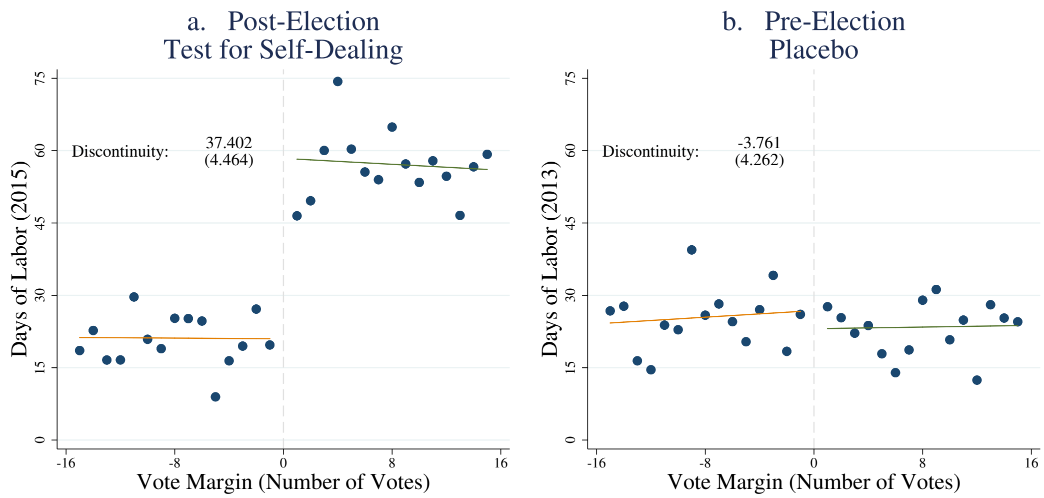
3.5 Main Results: There is Self-Dealing

3.5.1 How Much of Compensation Comes from Self-Dealing?

To test for corruption-as-compensation we must first test whether there is corruption. We estimate Equation 24 on candidates whose vote margin is within a bandwidth of 15 votes.[73] As the election was in mid-2014 we test for a discontinuity in the total days of labor allocated to the household of the candidate in 2015. The left-hand panel of Figure 3.5 shows the regression line of best fit alongside the average days of labor earned by households whose candidate had each possible winning margin. The figure shows a large discontinuity when the margin switches from negative to positive—that is, when a candidate switches from barely losing to barely winning. The winner receives an extra

[73] Unless otherwise specified we use this same bandwidth as we test other outcomes or specifications to avoid conflating the effect of changing specifications with the effect of changing the bandwidth. But the results are qualitatively similar when we vary the bandwidth.

Figure 3.5: Winners of Close Elections Receive 3 Times as Much Labor



Note: Standard errors are clustered by panchayat. The bin size is 1 vote. Each dot shows the average of the outcome within the bin.

37 days of labor—nearly 3 times as many as the loser—suggesting she heavily favors her own household over others.

Panel A of Table 3.2 shows this estimate (in Column 1) together with several robustness checks. In some panchayats we were unable to match both the winner and runner-up to their job card record. These observations are included in the main specification, but in Column 2 we verify that the result is robust to including only panchayats for which we are able to match both candidates. As noted in Section 3.3.2 we generally define the running variable as the margin of votes in levels. Column 3 verifies that defining the margin as a proportion of all votes cast does not qualitatively change the results.[74] Columns 5—7 estimate Equation 24 for other outcomes. Column 5 shows that winners receive 3 more jobs than losers (who receive 2). Column 6 shows that winners are 37 percentage points more likely to have gotten a job at all in 2015. Column 7 shows that their NREGS payments are nearly 6000 rupees higher on average. According to both our survey of council presidents and newspaper reports from Uttarakhand,[75] the

[74] Since this new running variable is on a completely different scale we calculate a different optimal bandwidth using the method of Calonico et al. (2014).

[75] National Herald, accessed on 26 July 2019. https://www.nationalheraldindia.com/national/5000-gram-pradhans-resign-after-ukhand-slashes-gram-sabha-funds

median annual salary is 9000 rupees. Column 7 thus implies the president earns excess NREGS returns equal to nearly two-thirds of the official salary. Finally, as we show in Appendix C.2.3, we cannot reject that presidents elected in constituencies reserved for women or members of lower castes (Scheduled Castes and Scheduled Tribes) self-deal the same amount as presidents elected in unreserved seats.[76]

Table 3.2: Main Results

Panel A: Main Results

	Days of Labor				Other Outcomes		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Basic	Dual Matches	Vote Share as RV	Test: SUTVA	# of Jobs	Any Job?	NREGS Payments
RD Estimate	37.402***	39.935***	24.781***	-0.043	3.305***	0.373***	5957.528***
	(4.464)	(5.635)	(6.777)	(2.736)	(0.419)	(0.053)	(708.960)
Outcome at Disc.	20.99	21.17	23.90	20.99	1.97	0.54	3333.89
Observations	1105	696	400	1105	1105	1105	1105
Panchayats	757	348	283	757	757	757	757

Panel B: Placebo and Specification Tests

	(1)	(2)	(3)	(4)	(5)	(6)
	Labor (2013)	SCT Cand.	Female Cand.	Name Length	Name Length (M. Rel.)	Matched?
RD Estimate	-3.761	0.012	0.010	-0.287	0.216	0.028
	(4.262)	(0.042)	(0.044)	(0.312)	(0.409)	(0.032)
Outcome at Disc.	26.85	0.18	0.47	10.46	8.61	0.32
Observations	1105	1105	1105	1105	1105	2400
Panchayats	757	757	757	757	757	1200

Panel C: Robustness to Bandwidth

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$h = 25$	$h = 22.5$	$h = 20$	$h = 17.5$	$h = 15$	$h = 12.5$	$h = 10$	$h = 7.5$
RD Estimate	36.565***	36.094***	34.875***	36.583***	37.402***	37.615***	36.110***	32.689***
	(3.519)	(3.791)	(3.954)	(4.186)	(4.464)	(5.088)	(5.386)	(6.625)
Outcome at Disc.	22.13	21.10	20.99	20.89	20.99	19.73	20.22	20.24
Observations	1467	1380	1336	1216	1105	898	752	472
Panchayats	1009	949	916	834	757	623	518	331

Note: "Outcome at Disc." gives the estimate of the counterfactual outcome at the cutoff in the absence of treatment (that is, the left limit at the cutoff). Standard errors are clustered by panchayat. See text for description of each specification.
*p=0.10 **p=0.05 ***p=0.01

3.5.2 Are the Identification Assumptions Valid?

Our main result shows that the winner of the election gets more days of labor than the loser, but is it possible the difference arises only because the loser is given fewer days of

[76] We also do not find evidence that presidents allocate excess NREGS benefits to extended family or other village council members (Appendix C.2.2).

labor than other households? Though punishing a political rival is clearly misconduct, it does not earn any financial return for the president. In jargon the question is whether the Stable Unit Treatment Value Assumption is violated. We test for a violation by reassigning every *winning* candidate the number of days earned by the average household in the panchayat (excluding both winner and loser). If our estimates are driven by harm to the loser rather than benefit to the winner, this estimate should be similar to the estimate in Column 1 of Table 3.2.A. But Column 4 shows that the estimate is close to zero, suggesting losers are treated no differently than the typical household. This implies that one year of self-dealing by the politician amounts to roughly the same amount of NREGS benefits a typical villager receives in three years.

We then test the key assumption behind the regression discontinuity design, that the households of candidates who barely lose are similar to those who barely win in all ways except that they lost the election. Like much of the literature, we test the assumption by testing for discontinuities in pre-determined outcomes. Since the election was in 2014, winning or losing should not affect outcomes determined before 2014—for example, the number of days of labor allocated in 2013. Any discontinuity would suggest the type of household that received more labor in 2013 was able to sort itself onto the winning side of the cutoff (say, by manipulating the vote count).

The right-hand panel of Figure 3.5 estimates and plots Equation 24 in exactly the same way as was done to construct the left-hand panel, but using as the outcome the days of labor in 2013. There is no sign of a discontinuity. Columns 1—5 of Table 3.2.B report applying the same procedure to several other pre-determined outcomes. Column 1 is the same as Panel B of Figure 3.5. Column 2 tests for differences in whether the winner is a member of a scheduled caste or tribe (SCT), both historically disadvantaged groups. Column 3 tests for whether the winner is more or less likely to be a woman. Ideally we would also test other measures of income or social status, but the job card data are relatively sparse. One very rough measure of social status is the length of the candidate's name, as higher caste candidates are likely to have a last name related to their caste (e.g. Kothari) whereas lower caste candidates tend to have "default" names that hide their caste (e.g. Devi). Columns 4 and 5 test for differences in the length of the winner's own name and that of the closest male relative (husband or father). None of these placebo tests show a difference that is statistically or economically significant.

It is also common in the literature to apply a test for discontinuities in the empirical density of the running variable. But the density of vote margins is continuous (and actually symmetric) because every winner to the right of the cutoff has a loser to the left. Then a discontinuity in our matched dataset can only arise if it is systematically easier to make a match between the election records and the job card data for winners. That is especially a concern if losers are less likely to get a NREGS job card, without which they would not even appear in the job card data. We test for whether there is a discontinuity in the match rate by taking the full set of candidates we attempted to match, restricting to the bandwidth of our main specification, and estimating Equation 24 on a dummy for whether the candidate was matched. Reassuringly, Column 6 suggests there is no discontinuity.[77]

Finally, we verify that the results are not sensitive to the choice of bandwidth. Table 3.2.C estimates Equation 24 for bandwidths ranging from as wide as 25 votes to as narrow as 7.5.[78] The estimates are all similar.

3.5.3 Is it Necessarily Self-Dealing?

Is it possible that there is a more innocent explanation for why the president gets more days of NREGS labor than anyone else? For example, the president might be supervising the projects to make sure they are completed properly, and thus needs to be on nearly every project. But each NREGS project has an official work site supervisor, the "Mate," and thus does not need an unofficial supervisor. The Mate is supposed to be chosen based on technical expertise that most presidents lack. Over 80 percent of presidents who answered our survey confirm that neither they nor any member of their household has served as a mate since the election. In any case, mates are paid directly for their labor through the project budget, not through NREGS labor.

[77] Though this is the most direct test for the underlying source of any discontinuity in the density, we also find no evidence of bunching in the final matched running variable (results available on request).

[78] Since the number of votes is discrete a fractional bandwidth is effectively rounded down.

The other innocent explanation is that the president is stepping in to keep work on NREGS projects continuing at times of the year when no one else needs employment. As noted in Section 3.2, demand for NREGS jobs generally far outstrips what is available. It is unlikely that there would have been a lack of interest in NREGS labor in 2015 when Uttarakhand suffered poor rainfall.[79] But we can test this hypothesis directly by checking whether presidents take less excess labor during the season when NREGS demand tends to be highest. Not surprisingly, the overwhelming majority of presidents (83 percent) report that NREGS demand is highest during the dry season (rabi). But when we estimate Equation 24 separately on labor in the dry season and labor during the monsoon season, we find very similar estimates that lie within a single standard error of one another.[80] There is no evidence to suggest the president's own NREGS allocation varies by season.

3.6 Potential Mechanisms

3.6.1 Corruption-as-Compensation Appears Plausible in the Cross Section...

While the presidents in our survey give us a coherent explanation for the self-dealing, they may have wanted to present us with what they believe to be the "correct" or least problematic response. Our interviewers made it clear that they were not affiliated with the government (and respondents seemed to accept that, as evidenced by their tendency to make unsolicited complaints about how the government runs NREGS). They also did not know that we were linking their NREGS allocation records from the website with their election results and other information. Nevertheless, presidents may have been unwilling to honestly report behavior that could be deemed unethical or embarrassing, such as feeling unaccountable to their citizens or self-dealing without a good reason. We therefore turn to testing the plausibility of the corruption-as-compensation channel with our data.

[79] Uttarakhand experienced a 71 percent deficit of post-winter rainfall in 2015. See e.g. Hindustan Times, '9 Uttarakhand Districts Declared Drought Affected', April 15, 2016.

[80] For the monsoon season the discontinuity is 17.9 days with a standard error of 2.6, and for the dry season it is 16.0 with a standard error of 2.6. Regression output is available on request.

The most straightforward prediction of this mechanism is that there should be a correlation between the level of self-dealing and the total NREGS employment provided in the village. Unfortunately our sample size is too small to directly test whether the regression discontinuity estimates are larger in villages with larger average allocations. We instead define 3 measures of corruption, which we use as the dependent variable in simple OLS regressions. Let D_{it} be the days of NREGS labor allocated to household i in year t, I_p the set of households in panchayat p, $j \in I_p$ be the household of the president of p, and $\hat{j} \in I_p$ that of the runner-up. Define

$$[\text{Corruption 1}]_p = D_{j,2015} - \frac{\sum^{i \in I_p \setminus \{j\}} D_{i,2015}}{|\{I_p \setminus \{j\}\}|} \tag{25}$$

$$[\text{Corruption 2}]_p = D_{j,2015} - D_{\hat{j},2015} \tag{26}$$

$$[\text{Corruption 3}]_p = D_{j,2015} - D_{j,2013} \tag{27}$$

Corruption 1 is the difference between the president's labor allocation and the average allocation to *all other households* in the panchayat. Corruption 2 is the difference between the allocation of the president and the runner-up in the election, which is only defined for panchayats where we are able to match both winner and runner-up to their job card record. Corruption 3 is the change in the president's allocation from the year before the election (2013) to the year after (2015). When we restrict our sample to elections won by 12 or fewer votes, the means of these three measures are 38.2 for Corruption 1, 37.6 for Corruption 2, and 35.0 for Corruption 3. All are similar to the estimated discontinuity in the main specification reported in Table 3.2.

We test the corruption-as-compensation channel by measuring the correlation between these three measures and the aggregate per household labor generated for the panchayat. Since the prediction applies specifically to differences in labor that are not driven by observable aggregate factors we control for block fixed effects. Since blocks are relatively small, weather and access to markets will be similar within a block. More importantly, the NREGS budget is fixed for each block and distributed between panchayats by a block-level program officer. A president who manages to bring home more jobs

Table 3.3: Performance Predicts Higher Corruption

	(1) Corruption 1	(2) Corruption 2	(3) Corruption 3
Avg. Labor (2015)	0.391**	-0.009	0.911***
	(0.172)	(0.316)	(0.211)
Mean Outcome	38.3	37.9	35.1
Observations	478	274	478

Note: The panchayat is the unit of observation. "Avg. Labor" is average NREGS labor for all households in the panchayat *excluding the pradhan*. Standard errors are robust to heteroskedasticity. All regressions control for block fixed-effects.

relative to other presidents appealing to the same officer might be seen as performing well.

We estimate an OLS regression of each measure of corruption on a set of block fixed effects and the average labor for all households in the panchayat *excluding the president's own household* (otherwise there would be a mechanical correlation). Though these are not regression discontinuity estimates, we nevertheless restrict to elections won by no more than 12 votes to exclude panchayats where elections are uncompetitive because, for example, one family monopolizes power.[81] Columns 1—3 of Table 3.3 show that Corruption 1 and 3 both show a significant positive correlation. Corruption 2 does not follow the pattern, but that may be because as noted earlier we cannot compute it for many panchayats, forcing us to drop roughly half the sample. These results are at least somewhat supportive of the prediction of corruption as compensation.

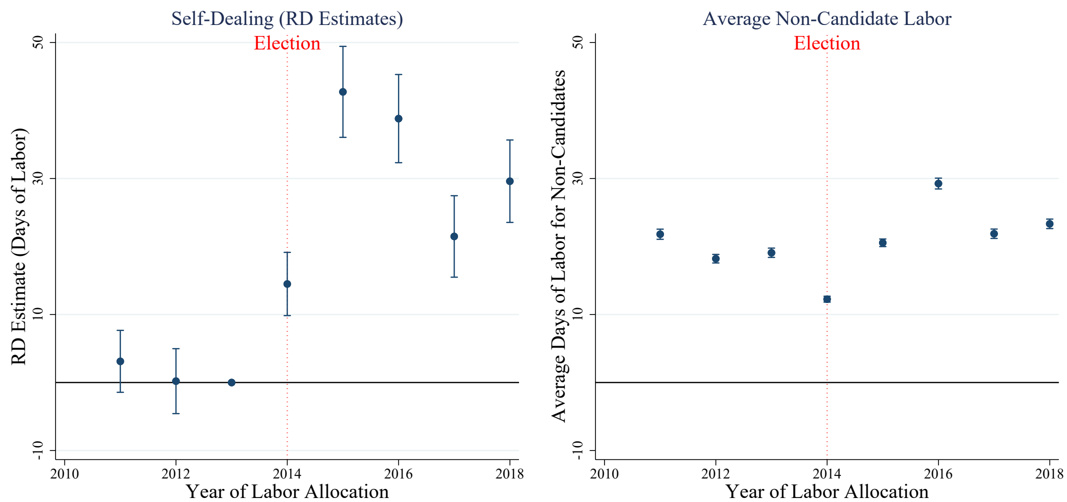
One may wonder whether block fixed-effects are adequate to control for aggregate factors other than the president's performance. In unreported results we run a placebo test for whether corruption in 2015 is positively correlated with average labor in 2013. Since 2013 is reflective of the previous president's performance it should not impact the pay of the current president, which is what we find.

3.6.2 ...but Does Not Withstand Closer Scrutiny in the Panel

But focusing solely on 2015 could be misleading because the NREGAsoft system, introduced in 2013, was still relatively new. The presidents elected in mid-2014 were the first

[81] We restrict to an even closer set of elections because unlike before we are comparing simple means rather than local linear estimates of the mean at the discontinuity. The results are broadly similar for other restrictions.

Figure 3.6: Self-Dealing Declines over Time, But Average Allocations Do Not



Note: RD estimates of main specification calculated separately for each calendar year. Standard errors are clustered by panchayat.

whose NREGS allocations could be monitored through the website from the first day in office. Since local stakeholders may not have much prior experience using the system to check on their leaders, it is worth asking whether they tamp down on corruption in later years.

Figure 3.6 suggests they do. The left-hand panel shows difference-in-discontinuities estimates that measure the size of the discontinuity relative to its size in 2013. Since presidents only took office in mid-2014, 2015 is the first true post-election year (which is why it is our focus in Section 3.5). But 2015 is the high-point in self-dealing. By 2018 it has fallen by roughly one quarter.

If the corruption-as-compensation hypothesis were true, the decline in self-dealing should be matched by a similar decline in performance. But the right-hand panel, which shows yearly averages of NREGS labor to regular households in the village actually rises from 2015 to 2018. The pattern seems inconsistent with the idea that voters face a trade-off between corruption and performance.

If the decline in corruption is somehow related to voters learning to monitor corruption, it should be stronger in villages with better infrastructure to monitor corruption. We proxy for better monitoring capacity with the distance to the nearest cyber cafe.

Villagers with access to a cyber cafe might be more likely to access the NREGS website to check actual program benefits, for example because they can access the internet in the cafe, because having a cyber cafe is correlated with a stronger internet connection than may be available elsewhere, or because cyber cafes are likely to be located near a tech-savvy local population that creates more demand for such a service.

The availability of a cyber cafe, which is based on information from the Indian Census, is positively correlated with presidents in our survey reporting that most villagers know of the NREGS website and that most villagers have visited the NREGS website, although given the smaller survey sample size the estimate is noisy. In either case, it seems unlikely that the proxy would exclusively capture villagers' direct monitoring of NREGS implementation. It may also proxy for the presence of journalists and local activists who can monitor a president's behavior and publicize their findings among the village population and beyond. Similarly, the political opposition will likely find it easier to scrutinize the president's behavior in office and should have an incentive to do so since it only narrowly lost a competitive election.

But our proxy does not necessarily have to imply that villagers themselves are monitoring the president's behavior online.[82]

Panel a of Figure 3.7 shows that average NREGS benefits were similar in villages close to and far from a cyber cafe and remain roughly constant over the studied time period. But Panel b, which shows the RD estimates of self-dealing, shows a very different pattern. Self-dealing is lower in villages closer to a cyber cafe and disappears entirely over time. The estimate for 2018 is even negative, implying that close election winners take fewer NREGS days than election losers. In villages far away from an internet cafe, on the other hand, self-dealing remains high and, with the exception of a temporary drop in 2017, at roughly the same magnitude. The divergence between performance and corruption in villages near cyber cafes contradicts the idea that corruption is compensation.

The year-by-year RD discontinuity graphs (Appendix Figures C.1 and C.2) confirm

[82] While we cannot directly test whether presidents actually show up to work on the NREGS projects or just receive the benefits, we can rule out that presidents are able to hide self-dealing by making up phantom projects with no or very few other workers (Appendix Table C.3).

that these estimates are not a fluke. And we show in Appendix Table C.1 that the decline in corruption near cyber cafes is robust to including a large number of control variables such as night lights, literacy, the fraction of low-caste (SC/ST) individuals, the distance to the state capital, the border, or district- and sub-district headquarters. This suggests that the heterogeneity by distance to internet cafe does not just proxy for other factors such as remoteness, better living conditions or higher economic growth.

Though not causal, this pattern suggests corruption arises through imperfect monitoring rather than to compensate performance. Corruption-as-compensation predicts that corruption should be no lower in villages with better monitoring because voters choose not to punish self-dealing to induce the president to put more unobservable effort into running NREGS. If anything the link between corruption and performance should be stronger in these areas because better monitoring enables stricter adherence to the corruption contract. Figure 3.7 is inconsistent with this prediction.

3.6.3 A Potential Explanation of the Results

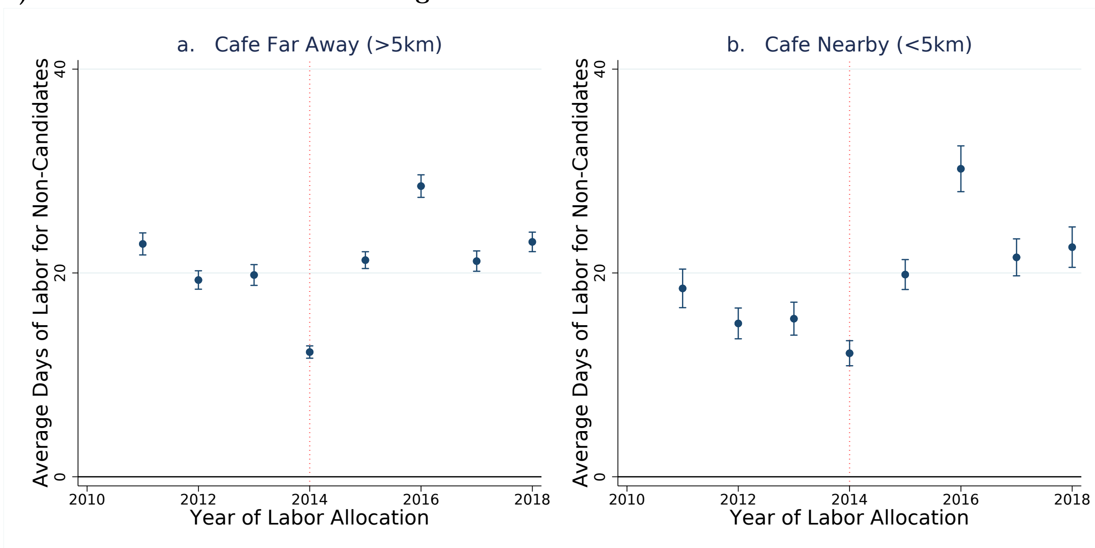
A plausible explanation of the empirical patterns needs to be consistent with three key results: First, self-dealing declines and ultimately disappears in villages close to cyber cafes, plausibly because local stakeholders can better monitor their president. By contrast, self-dealing persists in villages far from cafes where such monitoring is more difficult. Second, the average provision of NREGS jobs for villagers is the same irrespective of voters' monitoring capacity. Third, there is a positive correlation between president performance and self-dealing. Though Table 3.3 showed only the correlation in a single year, Appendix Table C.2 shows that this is robust at least in villages far from internet cafes.

These results are not consistent with a corruption-as-compensation explanation. The first key result suggests a role for imperfect monitoring of self-dealing by voters that is alleviated by a shorter distance to an internet cafe. But a simple model where voters have no ability to monitor performance or corruption cannot explain the second and third stylized facts.

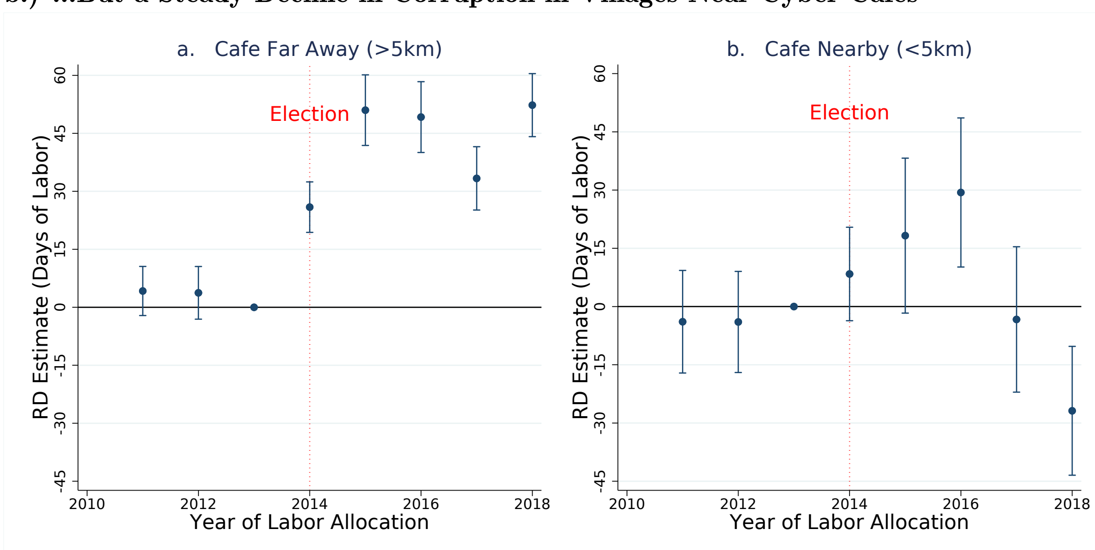
Average NREGS employment may be similar in villages both close and far from cyber

Figure 3.7: Dynamic Effects by Internet Cafe Distance

a) There is No Decline in Average NREGS Jobs for Citizens...



b.) ...But a Steady Decline in Corruption in Villages Near Cyber Cafes



Note: All standard errors are clustered by panchayat. All years are estimated simultaneously to allow for correlation in the coefficients. **a.)** Graphs show average NREGS days provided to villagers, not RD estimates. Cafe refers to the nearest internet cafe according to Census data. **b.)** Graphs show RD estimates separately for each calendar year. Cafe refers to the nearest cyber cafe according to Census data.

cafes because it is easier to monitor than self-dealing. Households perfectly observe their own NREGS allocation and can observe total employment by visiting physical worksites and talking to other villagers. Even without easy access to the NREGS website, voters can observe performance. But self-dealing may be harder to observe if the president does not actually work the jobs he is paid for and hides or manipulates the physical muster rolls. Then it would be hard to detect self-dealing without accessing the NREGS website.

But the correlation between average NREGS employment and self-dealing cannot be explained by a model where program performance is observable but self-dealing is not. Presidents should then maintain good performance while self-dealing as much as possible regardless of performance, breaking any correlation between the two. Instead, such a correlation might arise if there is an institutional constraint imposed by the project management software itself. Presidents have to create NREGS projects, assign workers to those projects, and then feed all of this information to the NREGASoft software that generates the information published on the website. Politicians can self-deal by adding their name to the muster rolls, but they can only add their name once. This means that presidents who want to self-deal more have to create more projects, leading to a correlation of self-dealing and NREGS performance.[83]

This explanation—that self-dealing is hard to monitor but the reporting technology limits the amount of self-dealing to one per muster roll—can reconcile all three stylized facts. In such a scenario, the excess allocations presidents make to themselves could well be a form of second-best corruption rather than being completely wasteful. Even in areas with imperfect monitoring, institutional constraints would incentivize presidents to put in more effort without the tacit approval from voters. This is in contrast to the corruption-as-compensation explanation where voters grudgingly accept corruption in return for better program implementation.

[83] Alternatively, the detection probability may increase rapidly when it becomes a large enough proportion of total NREGS benefits. This would incentivize presidents to keep self-dealing at a small percentage of overall NREGS benefits ('needle in the haystack').

3.7 Conclusion

Using a unique dataset, our paper tests whether local politicians use excess welfare benefits that they allocate to themselves as compensation for a better implementation of a major welfare program. While the politicians themselves claim that such behavior is reasonable and expected in their area, our results do not support such a mechanism. We find that village council presidents receive three times the benefits of a typical villager in the year after the election. While village-level jobs under NREGS remain constant or even rise over time, a better monitoring capacity in villages close to an internet cafe appears to lead to the eradication of self-dealing. In contrast, self-dealing remains high in areas where monitoring is plausibly more difficult, although most presidents do not exploit all corruption possibilities. While we cannot provide causal evidence of the exact explanation, the easiest explanation is that institutional constraints and the working of existing transparency and accountability mechanisms put an upper bound on corruption.

Our results suggest that completely eliminating self-dealing may require additional investments in monitoring capacity such as access to technology as well as in creating the socio-economic conditions that allow citizens to use those tools to effectively hold politicians accountable. Transparency and accountability are already unusually high in our context when compared to other contexts in developing countries. This is the combined result of large-scale reforms to the implementation of NREGS by the Indian government, Uttarakhand's local institutions and geographical position which creates small villages, and our focus on competitive elections determined by a few votes. But even here, the simple availability of information in close to real time and the ability to hold a politician accountable alone do not seem to be enough to eradicate corruption. Citizens and other players like the media or local NGOs may also have to be better enabled to pro-actively monitor program implementation.

Appendices

A Appendix for Chapter 1

Figure A.1: Search App: Feature Phone

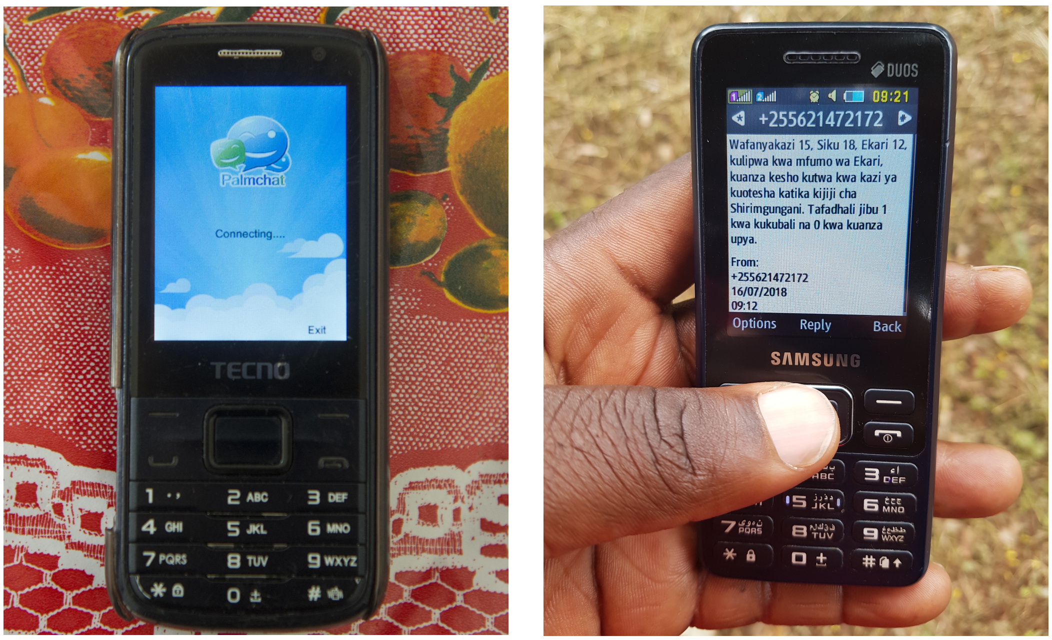


Figure A.2: Search App Development

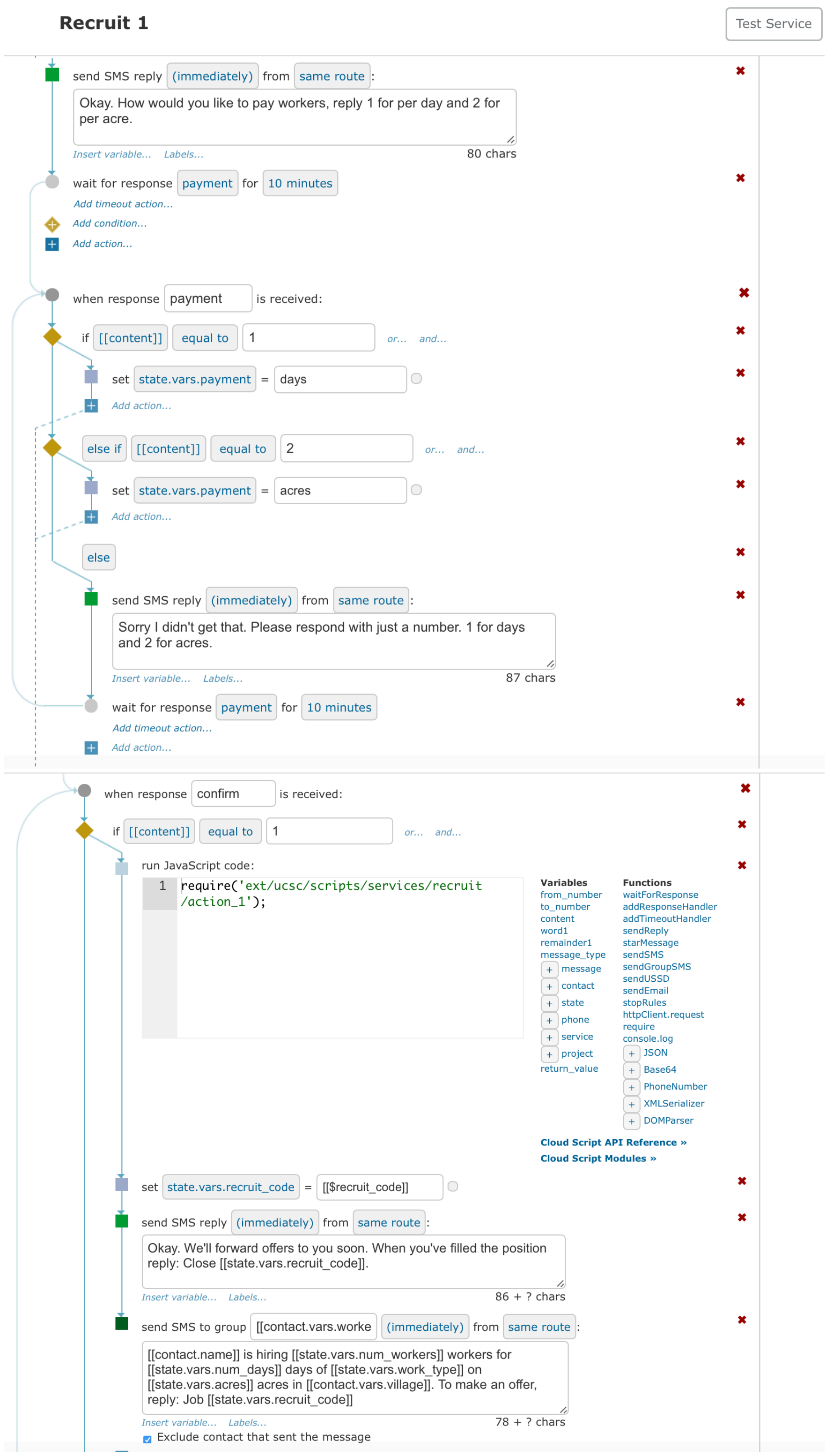


Figure A.3: How the Messaging App Works

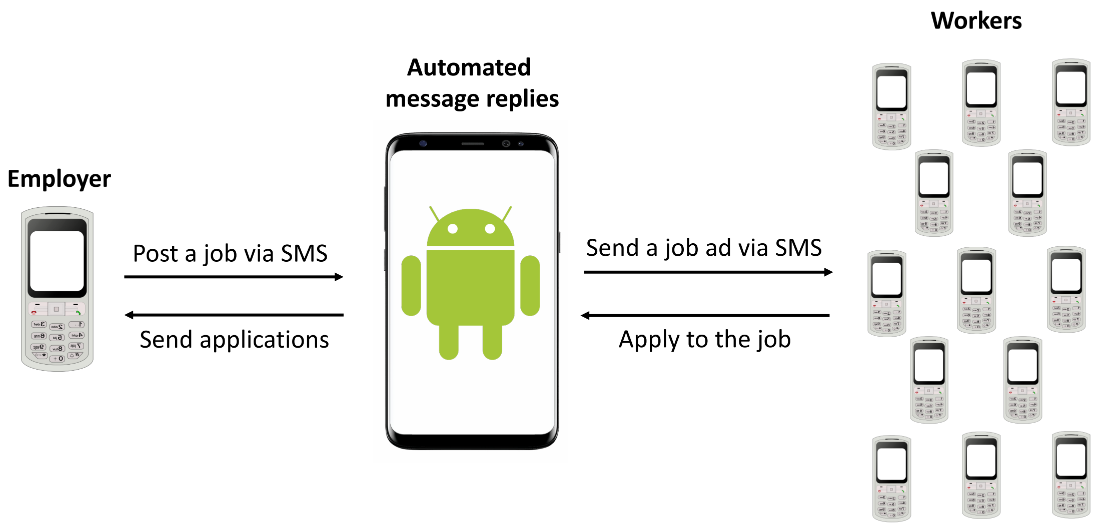


Figure A.4: Message Interactions of The Messaging App

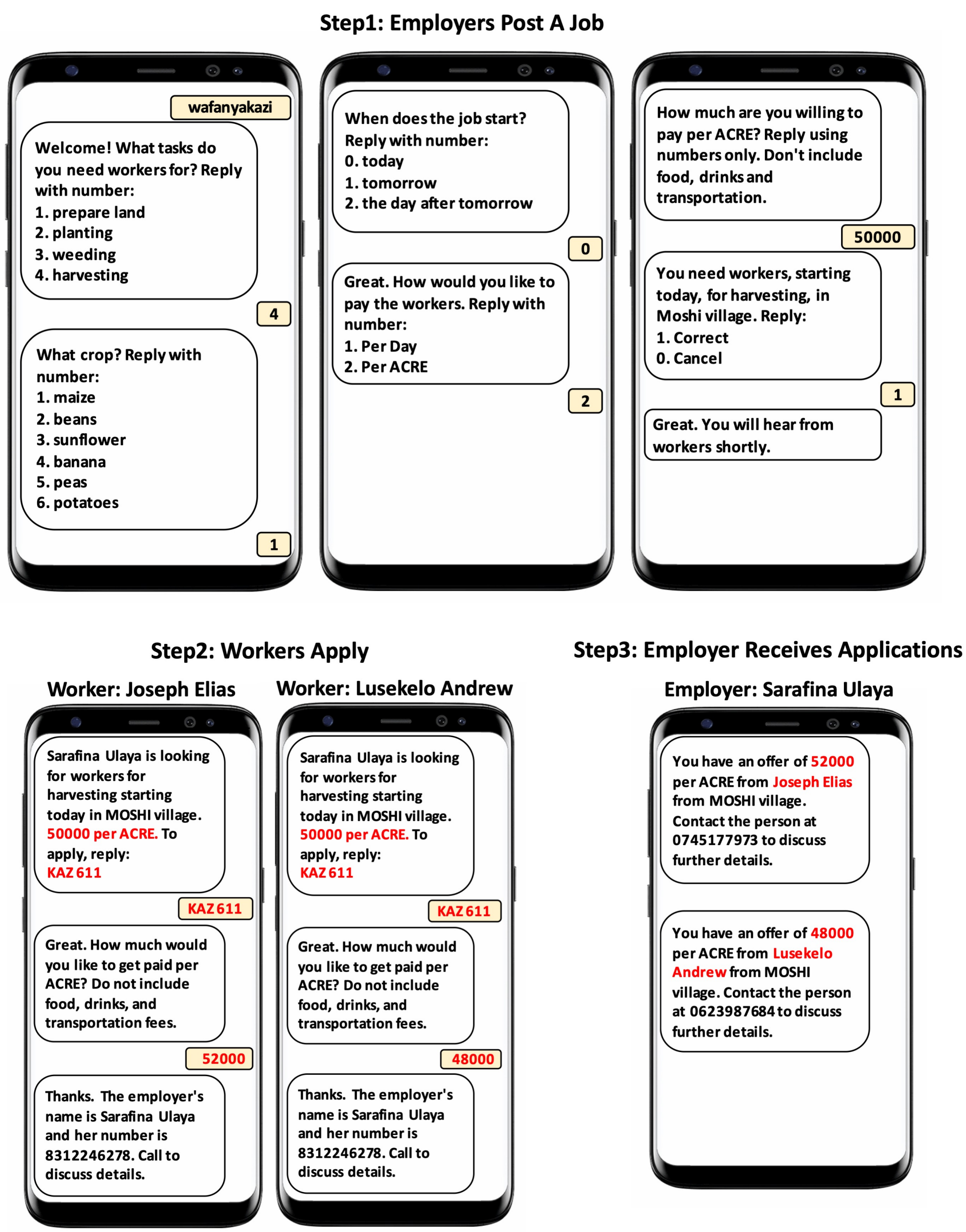
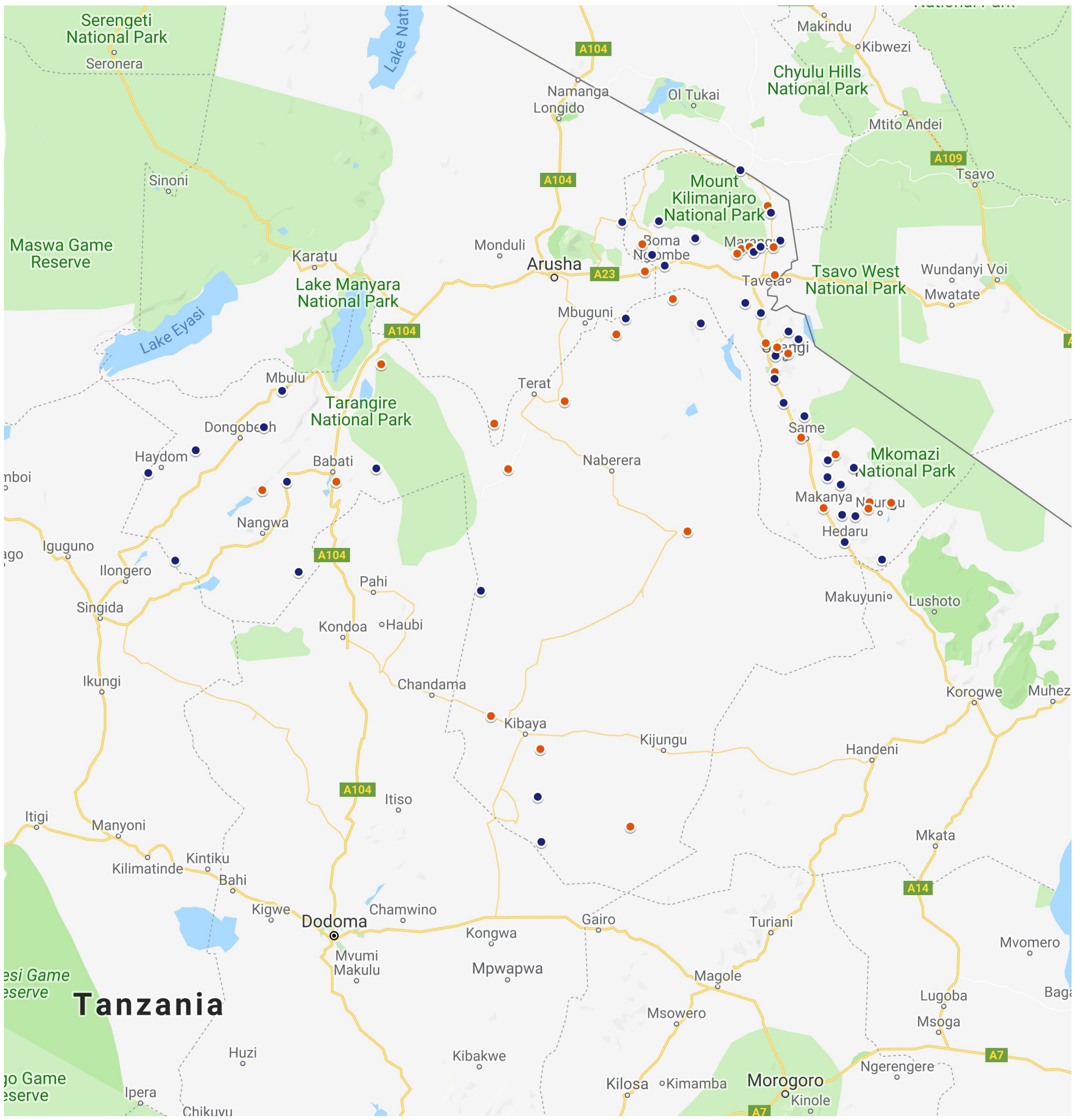


Figure A.5: Map of Study Villages



Notes: Orange dots represent 30 control villages and blue dots represent 40 treatment villages in Kilimanjaro and Manyara Region of Tanzania.

Figure A.6: Study Timeline

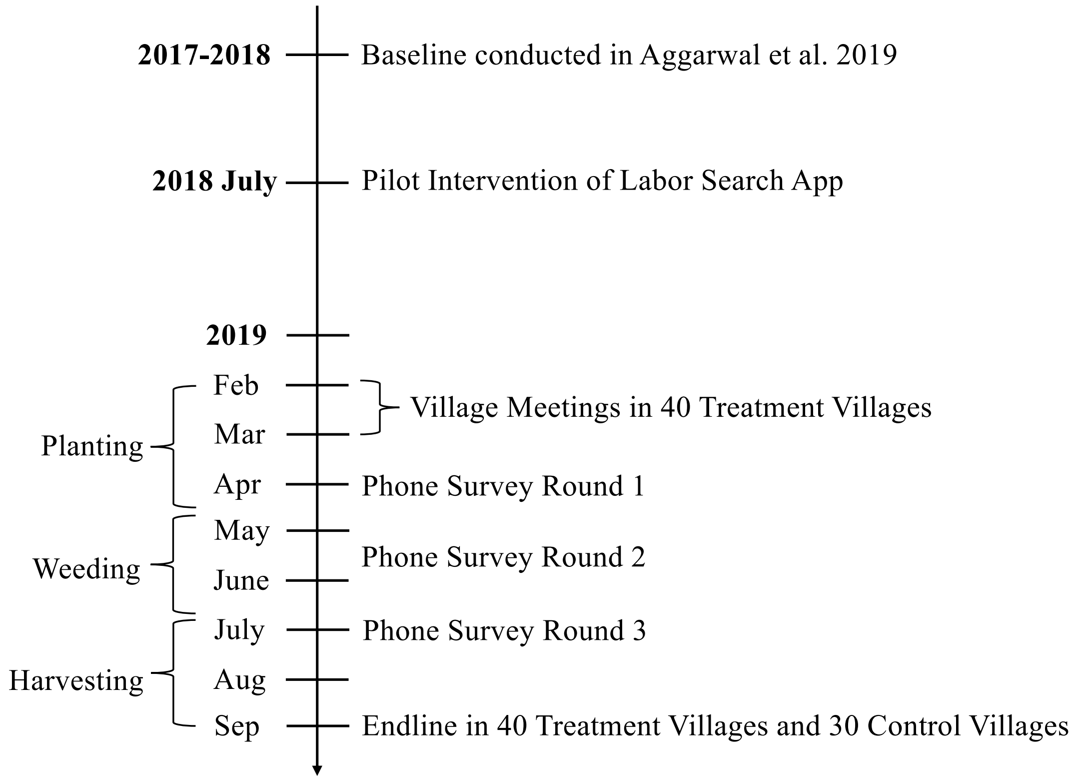


Figure A.7: An Example of Village Meeting

Figure A.8: Flyer Used For Village Meeting (1)

INNOVATIONS FOR POVERTY ACTION

SIMPLE WAY TO HIRE WORKERS FOR FARM ACTIVITIES

USE SMS SERVICE TO FIND WORKERS AND JOBS EASILY

REGISTRATION:

- Text **SAJILI** to **0746 217 484**
- You will receive text messages and follow the instructions.
- Registration and service is free of charge.

TO FIND CASUAL WORKERS

Once registered, text

WAFANYAKAZI

to

0746 217 484

and follow the instructions

TO FIND JOBS

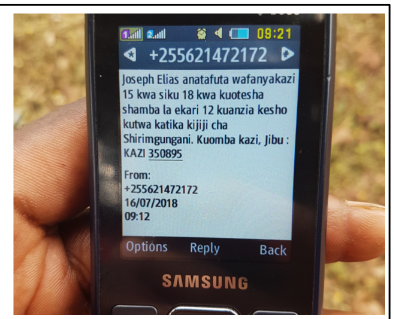
Once registered, to see if there are available jobs in your area, text:

KAZI

to

0746 217 484

Also, you will instantly receive job announcements whenever they are requested by employers.



FOR REGISTRATION AND OTHER QUESTIONS:

- **Call** Joseph Kissiri **0745 177973** or **0785 043635**
- **Do not call 0746 217484**. This number is not answerable.

Figure A.9: Flyer Used For Village Meeting (2)

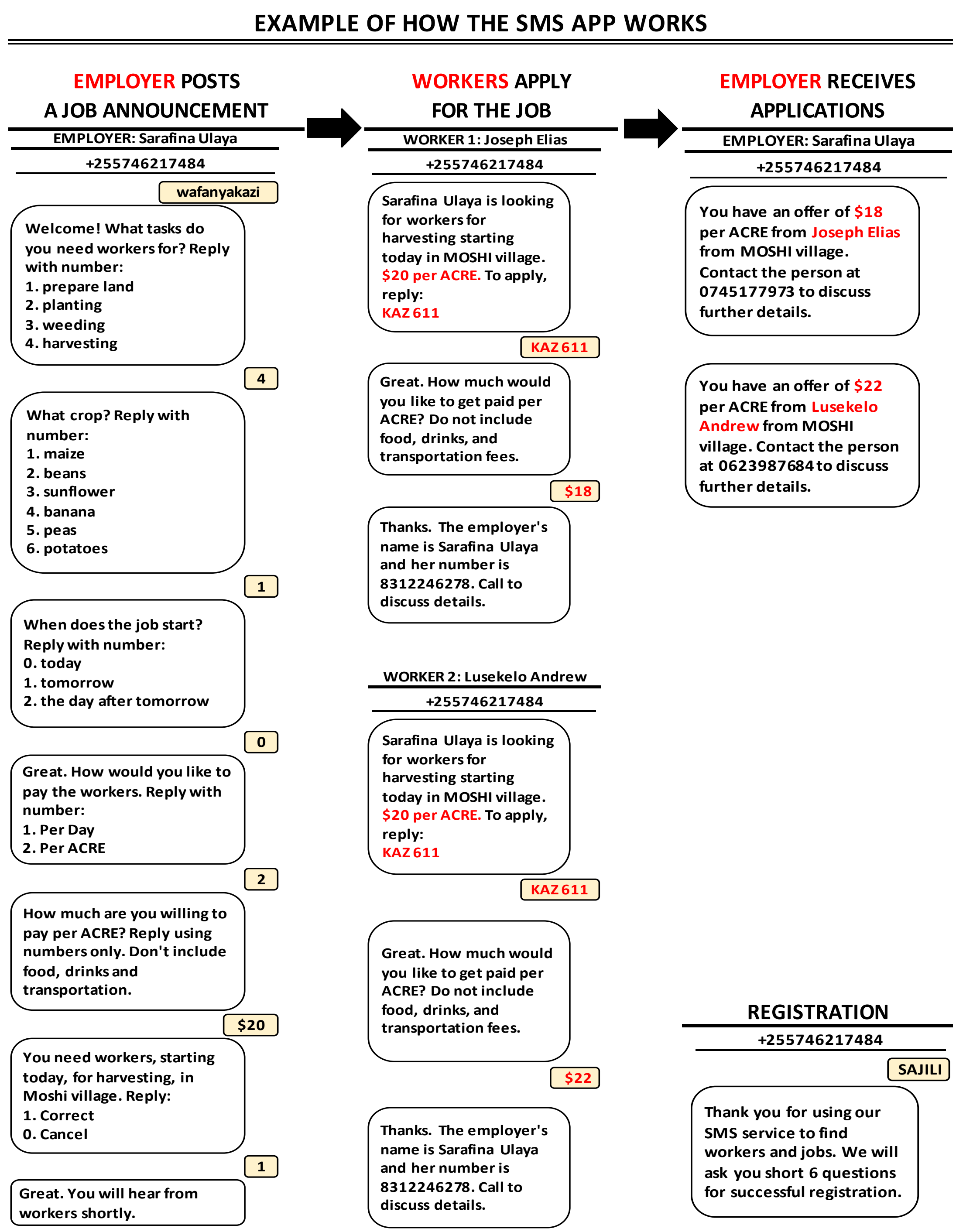
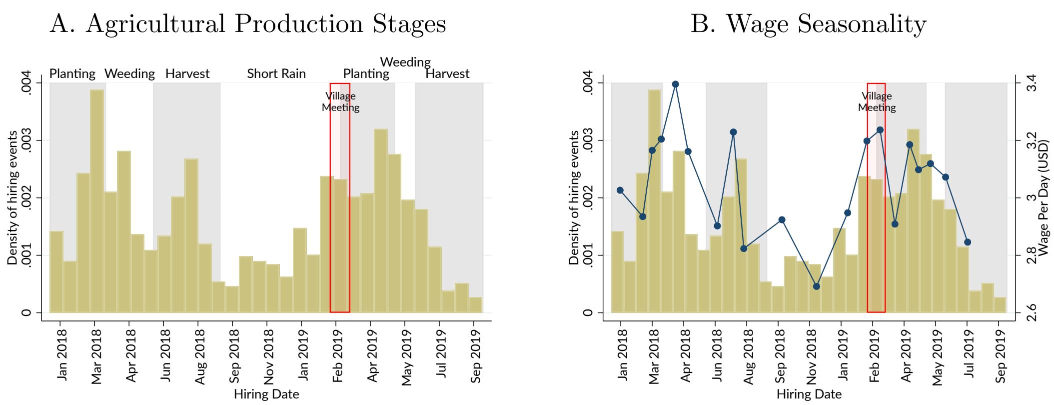


Figure A.10: Production Stages and Seasonality



Notes: Panel A is a binned scatter plot with 66 hiring events per bin on average.

Figure A.11: Wage Trajectory By Initial Wage Level

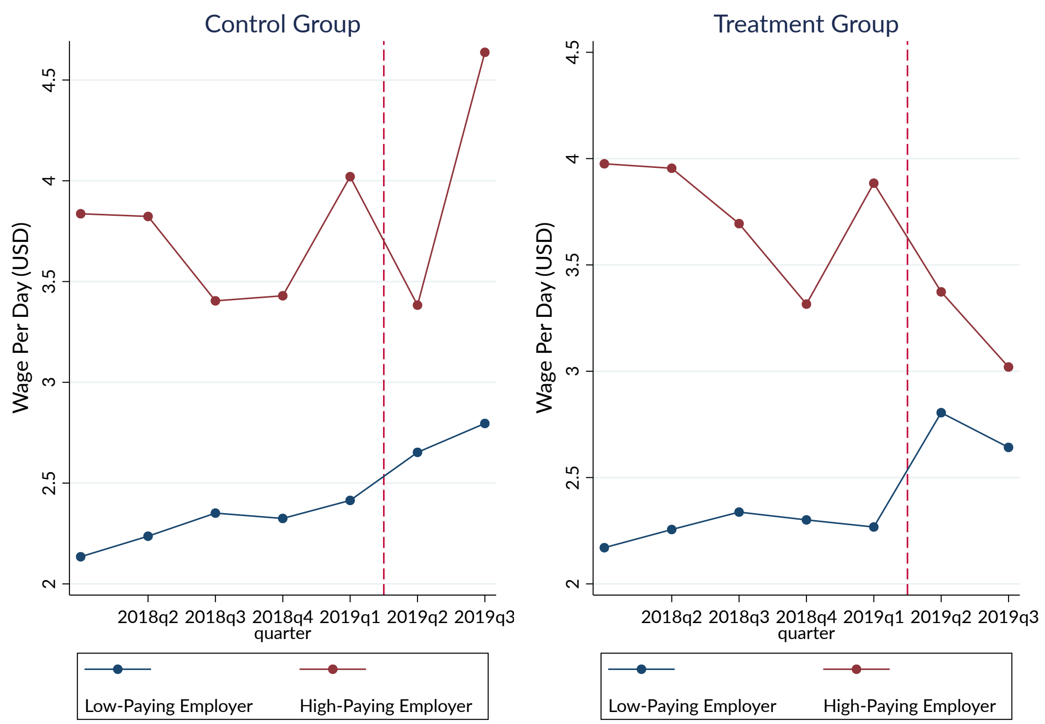
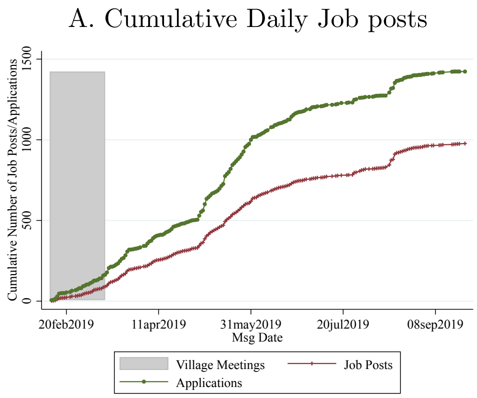


Figure A.12: Telerivet Usage



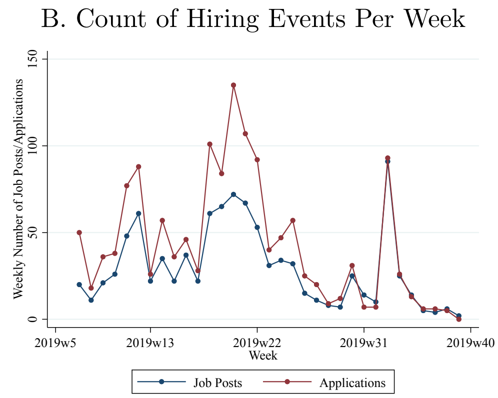
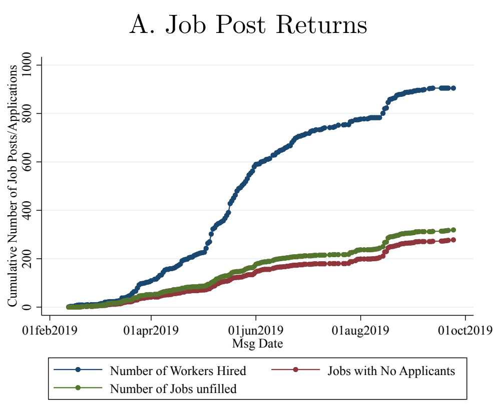


Figure A.13: Telerivet Job Post Performance



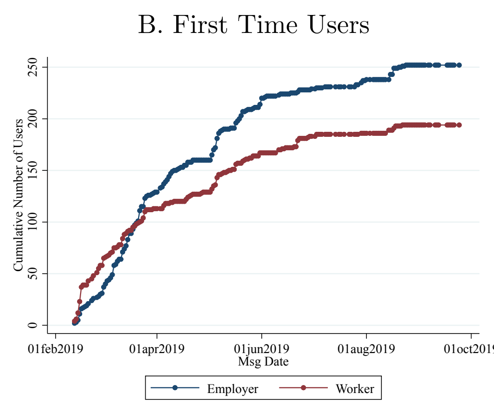


Table A.1: Within Treatment, Wage Display Is Cross-Randomized

	Show Wage	Message Example
Job AD	No	*Jennifer is looking for workers for weeding for beans, starting tomorrow in village MOSHI. To apply for the job, reply: Job 805*
	Yes	*Jennifer is looking for workers for weeding for beans, starting tomorrow in village MOSHI.* ***$4 per Day****. To apply for the job, reply: Job 805*
Worker Application	No	*Joseph from village MOSHI applied to your job post. Call 8312246278 to discuss details.*
	Yes	*Joseph from village MOSHI applied to your job post with a wage* ***$5 per Day****. Call 8312246278 to discuss details.*

Table A.2: Randomization Balance Check: Farmer Level

	(1) Control		(2) Treatment		T-test P-value
Variable	N/[Clusters]	Mean/SE	N/[Clusters]	Mean/SE	(1)-(2)
Abs. Percent Deviation From Village Avg Wage	136	1.989	201	-0.182	0.166
	[29]	(1.514)	[36]	(0.381)	
Standardized Wage	136	-0.054	201	0.037	0.592
	[29]	(0.093)	[36]	(0.142)	
Fraction of Households with Mobile Ownership	279	0.932	344	0.922	0.639
	[30]	(0.017)	[36]	(0.014)	
Respondent is Female	280	0.371	346	0.402	0.487
	[30]	(0.035)	[36]	(0.026)	
HH Hired Workers	280	0.689	346	0.671	0.705
	[30]	(0.038)	[36]	(0.032)	
HH Worked As A Casual Worker	280	0.421	344	0.430	0.872
	[30]	(0.042)	[36]	(0.035)	
Plot Size in Acreage	271	5.687	334	3.957	0.321
	[30]	(1.167)	[36]	(1.295)	
Total Labor Person Days	275	72.782	338	63.533	0.182
	[30]	(5.428)	[36]	(4.279)	
Labor Input Per Acre	271	31.478	334	36.926	0.182
	[30]	(3.048)	[36]	(2.690)	
Family Person Days	270	56.670	322	48.676	0.192
	[30]	(5.030)	[36]	(3.479)	
Hired Labor Person Days	275	24.015	338	20.556	0.578
	[30]	(4.505)	[36]	(4.303)	
Fraction of Hired Labor Person Days	274	0.222	334	0.232	0.818
	[30]	(0.031)	[36]	(0.029)	
Main Farming Season is Long Rainy Season	280	0.671	346	0.650	0.837
	[30]	(0.080)	[36]	(0.066)	
Used Chemical Fertilizer	271	0.247	334	0.266	0.813
	[30]	(0.057)	[36]	(0.058)	
Used Hybrid Seeds	270	0.578	334	0.614	0.639
	[30]	(0.059)	[36]	(0.050)	
Maize Harvest Quantity (Kg)	234	1209.718	293	900.638	0.221
	[30]	(215.943)	[36]	(129.626)	
Sold Maize	262	0.321	322	0.304	0.793
	[30]	(0.048)	[36]	(0.040)	

Notes: The balance test is shown for 650 study farmers based on the baseline survey collected *before* the intervention. The number of control villages is 30 and the number of treatment villages is 40. See Appendix Table A.3 for the balance table at village level using phone survey and endline survey data. Last column shows the p-value of the t-test for the equality of the two means.

Table A.3: Randomization Balance Check: Village Level

Variable	N	(1) Control Mean/SE	N	(2) Treatment Mean/SE	T-test P-value (1)-(2)
HH Worked as Casual Worker(s)	30	0.099	36	0.117	0.491
		(0.019)		(0.018)	
HH Hired Casual Workers	30	0.611	36	0.535	0.140
		(0.034)		(0.036)	
Average Wage	29	2.929	35	3.089	0.293
		(0.106)		(0.106)	
SD in Wage	29	1.130	35	1.101	0.813
		(0.080)		(0.089)	
CV in Wage	29	0.380	35	0.354	0.493
		(0.025)		(0.026)	
P50-p10 Wage Ratio	29	1.612	35	1.602	0.915
		(0.068)		(0.067)	
Mean-Min Wage Ratio	29	1.810	35	1.747	0.541
		(0.075)		(0.070)	
SD Log (Labor Per Acre)	30	0.831	36	0.777	0.527
		(0.065)		(0.054)	
CV Labor Per Acre	30	0.762	36	0.691	0.286
		(0.043)		(0.049)	
P50-p10 Labor Per Acre Ratio	30	3.529	36	3.587	0.942
		(0.662)		(0.491)	
Mean-Min Labor Per Acre Ratio	30	4.993	36	4.482	0.599
		(0.738)		(0.632)	
Number of Hired Workers	29	7.391	35	6.203	0.074*
		(0.501)		(0.427)	
Number of Prospective Workers Per Position	29	1.608	35	1.543	0.394
		(0.053)		(0.053)	
On-Farm Labor Days	30	82.135	36	76.368	0.539
		(8.326)		(4.956)	
On-Farm Family Days	30	48.442	36	46.323	0.736
		(5.235)		(3.677)	
On-Farm Hired Labor Days	30	24.612	36	18.567	0.137
		(3.435)		(2.287)	
On-Farm Exchange Labor Days	30	2.518	36	3.917	0.186
		(0.540)		(0.842)	
Used Fertilizer	30	0.226	36	0.254	0.698
		(0.051)		(0.049)	
Used Hybrid Seeds	30	0.524	36	0.545	0.748
		(0.047)		(0.042)	
Total Harvest Value (USD)	30	170.693	36	206.652	0.192
		(17.929)		(19.904)	

Notes: The balance test is done using the phone survey and endline survey. These surveys were implemented *after* the intervention, but the data for the pre-intervention period was also collected by recall as part of the surveys. See Appendix Table A.2 for the balance table using the baseline data implemented *before* the intervention. Last column shows the p-value of the t-test for the equality of the two means.

Table A.4: Attrition: Phone Survey

	Compliance: Phone Survey Completed			
	(1)	(2)	(3)	(4)
	Round1	Round2	Round3	All 3 Rounds
TREAT	-0.0318	-0.0182	-0.0254	-0.00921
	(0.0387)	(0.0378)	(0.0399)	(0.0396)
Constant	0.621***	0.611***	0.618***	0.579***
	(0.0281)	(0.0266)	(0.0301)	(0.0299)
Farmers	626	626	626	626
Villages	66	66	66	66

Notes: The attrition table includes those farmers for which we do not have a phone number. They are coded as non-compliance. Phone numbers of some treatment farmers were updated during the recent village meetings. In other words, we would have not been able to reach them if we did not hold village meetings. Since control villages did not have village meetings, I assume that those whose phone numbers got updated during the meetings are also non-compliant to ensure the balance between control and treatment villages. Round 2 and 3 asked farmers' all hiring/working events from the most recent survey date. For example, if a farmer participated in Round 1 but not in Round 2, then Round 3 asked their hiring/working events since the completed date of Round 1 survey. Standard errors are clustered at the village level.

Table A.5: Attrition

		Reasons of Non-Compliance					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Interviewed	Refused	Moved	Unidentified	Travelling	Work Away	Sick/Died
Treatment	-0.01	0.01	-0.00	-0.00	0.01	-0.00	0.01
	(0.03)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
Constant	0.91***	0.02**	0.02**	0.02**	0.01	0.01*	-0.00***
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
Farmers	626	626	626	626	626	626	626
Villages	66	66	66	66	66	66	66

Table A.6: Test for Equality of Two Variances of Wage Distributions in Figure 1.3.

Test	Outcome	Control SD	Treatment SD	P-value
Levene (1960)	Raw Wage (Pre)	1.377	1.371	0.966
Brown and Forsythe (1974)	Raw Wage (Pre)	1.377	1.371	0.835
Levene (1960)	Raw Wage (Post)	1.439	1.271	0.007
Brown and Forsythe (1974)	Raw Wage (Post)	1.439	1.271	0.008
Levene (1960)	Residualized Wage (Pre)	1.316	1.226	0.155
Brown and Forsythe (1974)	Residualized Wage (Pre)	1.316	1.226	0.169
Levene (1960)	Residualized Wage (Post)	1.292	1.141	0.006
Brown and Forsythe (1974)	Residualized Wage (Post)	1.292	1.141	0.007

Notes: Standard deviations of Control and Treatment group are reported in Columns 3 and 4. Column 5 reports the p-value from the test results with the null hypothesis of equal variances using the STATA command *robvar*. Brown and Forsythe (1974) replaces the mean in Levene (1960)'s formula with the median.

Table A.7: ANCOVA Estimation of the Main Regression Results

	p50-p10 Ratio		Mean-Min Ratio	
	(1)	(2)	(3)	(4)
	p5010	p5010	meanmin	meanmin
TREAT	-0.182*	-0.139	-0.256**	-0.183**
	(0.101)	(0.0888)	(0.0995)	(0.0847)
Baseline p50-p10 Ratio		0.546***		
		(0.118)		
Baseline Mean-Min Ratio				0.544***
				(0.102)
Observations	68	68	68	68
Villages	68	68	68	68
Control Mean	1.562	1.562	1.707	1.707

Notes: This regression is at a village-level and controls for baseline outcome measures. Two villages out of 70 villages are dropped because they do not have any hiring events data after the intervention.

Table A.8: Wage Dispersion Is Not Driven By Bidding Feature (Village Level)

	No Winsorization		Winsorized at p5 and p95	
	(1)	(2)	(3)	(4)
	Wage	Wage&B	Wage	Wage&B
TREAT	-0.0291	-0.0239	-0.0212	-0.0151
	(0.0302)	(0.0297)	(0.0259)	(0.0258)
TREAT × Post	-0.0596	-0.0595*	-0.0516	-0.0513
	(0.0379)	(0.0352)	(0.0349)	(0.0321)
TREAT_BID	0.0181	0.0199	0.0122	0.00940
	(0.0325)	(0.0339)	(0.0271)	(0.0279)
TREAT_BID × Post	-0.00708	-0.0159	-0.0153	-0.0160
	(0.0386)	(0.0361)	(0.0341)	(0.0308)
Stage FE	X	X	X	X
Task FE	X	X	X	X
Crop FE	X	X	X	X
Observations	1613	1613	1613	1613
Farmers	439	439	439	439
Villages	66	66	66	66
Control Mean	0.323	0.319	0.300	0.293

Notes: This regression is at a farmer-hiring event level data. Outcomes are individual percent deviation from the village mean wage and/or benefits in USD. The results are robust to using median wage as opposed to mean wage. Standard errors clustered at the village level.

Table A.9: Treatment Effect on Skipping Meals Controlling For Winning A Raffle

	Dep.Var: Skip A Meal (HH Level)			
	(1)	(2)	(3)	(4)
	Past6m	Past3m	Past6m	Past3m
TREAT	-0.035*	-0.032*	-0.034*	-0.032*
	(0.020)	(0.017)	(0.020)	(0.017)
Won A Raffle Prize of 10 USD	-0.043***	-0.020*		
	(0.016)	(0.011)		
Won A Raffle Prize of 10 USD Or A Feature Phone			-0.049***	-0.026**
			(0.014)	(0.010)
Observations	566	566	566	566
Households	566	566	566	566
Villages	66	66	66	66
HH Endowment	X	X	X	X
Control Mean	0.075	0.051	0.075	0.051

Notes: This table replicates columns 5 and 6 in Table 1.13 controlling for winning a raffle prize. During the village meetings, one randomly selected person was given a feature phone to motive the training session. Throughout the agricultural season, two person was randomly selected from each village to get 10 USD if they used the app. Note that the random selection includes all users (not just study farmers who were interviewed). Among the study sample, six farmers won the phone during the village meeting, and seven farmers won the 10 USD. This is the 2 percent of the study sample (13/584 = 0.02).

Table A.10: Feedback On The Search App (Treatment Villages Only)

	Mean	SD	N
The App Service Was Useful	0.93	0.25	44
Plan To Use The App in Future	0.50	0.50	274
I Am Willing To Contribute For The Service	0.74	0.44	129
Willingness To Pay Per Month	0.40	0.26	89
Willingness To Pay Per Ag. Season	1.49	2.84	86

B Appendix for Chapter 2

B.1 Remoteness and Market Access

Below, we show that a population-weighted average distance to hubs can be justified as an approximation for the market access measure in Donaldson and Hornbeck (2016). To see this, market access in Donaldson and Hornbeck is written as:

$$MA_v = \sum_h \tau_{hv}^{-\theta} N_h$$

where h indexes hubs, v indexes villages, τ is the iceberg trade cost, θ a trade elasticity to be estimated, and N_h is the share of population h in total population. Suppose that we can write the iceberg cost as $\tau_{hv} = f(d_{hv})$, where d_{hv} is distance. Then, market access becomes:

$$MA_v = \sum_h (f(d_{hv}))^{-\theta} N_h$$

A first-order approximation of this market access function, around the some point in space (with distance to each hub d_h), we have

$$MA_v \approx \sum_h (f(d_h))^{-\theta} N_h - \theta \sum_h (f(d_h))^{-\theta-1} N_h f'(d_h)(d_{hv} - d_h)$$

Collecting terms:

$$\begin{aligned} MA_v &\approx \underbrace{\sum_h (f(d_h))^{-\theta} N_h + \theta \sum_h (f(d_h))^{-\theta-1} N_h f'(d_h) d_h}_{\alpha_0} - \theta \sum_h (f(d_h))^{-\theta-1} N_h f'(d_h) d_{hv} \\ &\approx \alpha_0 - \theta \sum_h (f(d_h))^{-\theta-1} N_h f'(d_h) d_{hv} \end{aligned}$$

Assuming that the point in space that we choose is equidistant from all hubs ($d_h = d\ \forall h$), we can simplify market access as:

$$\begin{aligned} MA_v &\approx \alpha_0 - \theta (f(d))^{-\theta-1} f'(d) \sum_h N_h d_{hv} \\ &\approx \alpha_0 - \alpha_1 \sum_h N_h d_{hv} \end{aligned}$$

Standardizing this equation gives us:

$$MA_v^z \approx -\alpha_z \left(\sum_h N_h d_{hv} \right)^z$$

Thus, population weighted average distance can be justified as a first-order approximation to market access, after appropriate standarization.

B.2 Deriving farmer profits, revenues, and input expenditures

The production function under basic technology is:

$$Y_i = \widetilde{\theta}_{i0} K_i^{\alpha} L_i^{1-\alpha} \tag{28}$$

Here, $\widetilde{\theta}_{i0}$ is baseline productivity without technology for farmer i, K_i is land held by farmer i (which is assumed to be fixed), and L_i is labor hired/used by farmer i. Farmers who choose the baseline technology maximize the following profit function:

$$\Pi_{i0} = \max_{L_i} \left\{ p_i \widetilde{\theta}_{i0} K_i^{\alpha} L_i^{1-\alpha} - w_i L_i \right\} \tag{29}$$

where p_i is the output price and w_i is the local wage. The first-order condition with respect to labor is written as:

$$(1-\alpha) p_i \widetilde{\theta}_{i0} K_i^{\alpha} L_i^{-\alpha} = w_i \tag{30}$$

Multiplying both sides of the first order condition by L_i, it is straightforward to show that expenditures on labor are linked to revenues (R_{i0}) and profits (Π_{i0}) by

$$w_i L_i = (1-\alpha) p_i \widetilde{\theta}_{i0} K_i^{\alpha} L_i^{1-\alpha} = (1-\alpha) R_{i0} \tag{31}$$

and substituting into the profit function, we have:

$$\begin{aligned}\Pi_{i0} &= \alpha R_i \\ \Rightarrow w_i L_i &= \frac{1-\alpha}{\alpha}\Pi_{i0}\end{aligned}$$

Thus, labor expenditures are proportional to profits and revenues, a feature that will prove convenient when aggregating the model. Explicitly solving for labor in the first order condition, and substituting into the profit function, we have:

$$\begin{aligned}\Pi_{i0} &= \alpha_0(1-\alpha_0)^{\frac{1-\alpha_0}{\alpha_0}}\widetilde{\theta}_{i0}^{\frac{1}{\alpha_0}} p_i^{\frac{1}{\alpha_0}} w_i^{-\frac{1-\alpha_0}{\alpha_0}} K_i \\ &= \theta_{i0}\pi_{i0} \end{aligned} \tag{32}$$

Here, we have defined $\theta_{i0} = \alpha_0(1-\alpha_0)^{\frac{1-\alpha_0}{\alpha_0}}\widetilde{\theta}_{i0}^{\frac{1}{\alpha_0}}$ and $\pi_{i0} = p_i^{\frac{1}{\alpha_0}} w_i^{-\frac{1-\alpha_0}{\alpha_0}} K_i$. We return to these two terms momentarily when characterizing the adoption decision.

The production function *with* fertilizer splits variable inputs into labor and acquired fertilizer, X_{ijv}, and also provides a productivity shock, $\widetilde{\theta}_{ijv}$, which may vary by the agrovet j location v pair at which the fertilizer is purchased. Precisely, production is written as:

$$Y_i = \widetilde{\theta}_{ijv}\left(\theta_i K_i\right)^{\alpha} L_{ijv}^{(1-\alpha)\beta} X_{ijv}^{(1-\alpha)(1-\beta)} \tag{33}$$

The profit maximization problem when using fertilizer is written as:

$$\Pi_{i0} = \max_{L_i, X_{ijv}} p_i \widetilde{\theta}_{ijv}\left(\theta_i K_i\right)^{\alpha} L_{ijv}^{(1-\alpha)\beta} F_{ijv}^{(1-\alpha)(1-\beta)} - w_i L_{ijv} - r_{ijv} F_{ijv} \tag{34}$$

Since technology is Cobb-Douglas, including within variable inputs, similar results from above apply here. That is, writing expenditures on variable inputs as $c_{ijv}M_{ijv}$, where c_{ijv} is the unit cost of a bundle of variable inputs M_{ijv}, it is easily shown that

$$c_{ijv}M_{ijv} = (1-\alpha)p_i\widetilde{\theta}_{ijv}\left(\theta_i K_i\right)^{\alpha} L_{ijv}^{(1-\alpha)\beta} F_{ijv}^{(1-\alpha)(1-\beta)} = (1-\alpha)R_{ijv} \tag{35}$$

and

$$
\begin{aligned}
\Pi_{ijv} &= \alpha R_{ijv} \\
\Rightarrow c_{ijv} M_{ijv} &= \frac{1-\alpha}{\alpha} \Pi_{ijv}
\end{aligned}
$$

Further, since labor and fertilizer have β and $1-\beta$ share in variable inputs, respectively, expenditures on each input are written as:

$$
\begin{aligned}
w_i L_{ijv} &= \beta \frac{1-\alpha}{\alpha} \Pi_{ijv} \\
r_{ijv} F_{ijv} &= (1-\beta) \frac{1-\alpha}{\alpha} \Pi_{ijv}
\end{aligned}
$$

Thus, any results related to profits will apply to input expenditures as long as factor shares do not change.

Solving for the optimal labor and quantity of fertilizer from agrovet j and location v, profits of i from adopting at jv are written as:

$$
\Pi_i = \theta_{ijv} \pi_i r_{ijv}^{-\sigma} \tag{36}
$$

where $\sigma \equiv \frac{1-\alpha}{\alpha}(1-\beta)$, $\pi_i = p_i^{\frac{1}{\alpha}} w_i^{-\beta \frac{1-\alpha}{\alpha}} K_i$, and $\theta_{ijv} = \kappa_2 \widetilde{\theta}_{ijv}^{\kappa_1}$.[84] Here, the profitability of fertilizer at this location is a function of the productivity shock, θ_{ijv}, the (delivered) price of fertilizer itself, r_{ijv}, and profits based on local observables and technology π_i.

B.3 Distributions of Fertilizer Expenditures

Above, we used the following property to generate a market clearing condition that can be taken to the data:

$$
\mathbb{E}\left[rF_i \,|\, adopt \;\; at \;\; j \, in \, v\right] = \mathbb{E}\left[rF_i | adopt\right] \tag{37}
$$

That is, that the expected fertilizer expenditures, conditional on adopting at location j, is the same as the expected fertilizer expenditure, conditional on adopting anywhere. This is a similar result to Eaton and Kortum (2002), where the price distribution con-

84 κ_1 and κ_2 are constant functions of model parameters

ditional on being the lowest price supplier is the same as the unconditional price distribution at that destination. Here, we prove the similar result in the input adoption context.

In the model, fertilizer expenditures at a particular agrovet are a scalar function of ex-post profits when choosing that agrovet. Thus, we focus all proofs on the distribution of profits, and then the analogue to revenues and input expenditures follows directly. To begin, we first derive the distribution of profits for farmer i who buys from agrovet j in location v.

$$\begin{aligned}\Pr(\Pi_{ijv} > \pi) &= \Pr\left(\theta_{ijv}\pi_i r_{ijv}^{-\sigma} > \pi\right) && (38)\\ &= \Pr\left(\theta_{ijv} > \frac{\pi}{\pi_i} r_{ijv}^{\sigma}\right) && (39)\\ &= 1 - \exp\left(-T_{jv}\pi_i^{\varepsilon} r_{ijv}^{\varepsilon\sigma}\pi^{-\varepsilon}\right) && (40)\end{aligned}$$

Defining $\gamma_{ijv} \equiv \pi_i^{\varepsilon} r_{ijv}^{\varepsilon\sigma}$

$$\Pr(\Pi_{ijv} > \pi) = 1 - \exp\left(-T_{jv}\gamma_{ijv}\pi^{-\varepsilon}\right) \quad (41)$$

Similarly, the distribution of profits of the outside option of not purchasing fertilizer are written as:

$$\Pr(\Pi_{i0} > \pi) = 1 - \exp\left(-\Phi_{i0}\pi^{-\varepsilon}\right) \quad (42)$$

where $\tilde{\Phi}_{i0} = T_{i0}\gamma_{i0} \equiv \pi_i^{\varepsilon}$

Next, defining Π_i^{max} as the profits available from the best *agrovet* option for farmer i, we write the distribution of these profits as:

$$\begin{aligned}\Pr(\Pi_i^{max} > \pi) &= \Pr(\Pi_{ijv} > \pi \text{ for any } jv) && (43)\\ &= 1 - \Pr(\Pi_{ijv} < \pi \ \forall\ jv) && (44)\end{aligned}$$

Since θ's at each j, v pair are drawn from independent distributions, this probability is simplified as:

$$\Pr\left(\Pi_i^{max} > \pi\right) = 1 - \Pr\left(\Pi_{ijv} < \pi \ \forall \ jv\right) \tag{45}$$

$$= 1 - \prod_{v' \in \mathcal{V}} \prod_{j \in \mathcal{J}_v} \Pr\left(\Pi_{ijv} < \pi\right) \tag{46}$$

$$= 1 - \prod_{v' \in \mathcal{V}} \prod_{j \in \mathcal{J}_v} \exp\left(-\pi^{-\varepsilon}\right) \tag{47}$$

Defining $\tilde{\Phi}_i = \sum_{v' \in \mathcal{V}} \sum_{j \in \mathcal{J}_v} T_{jv}\gamma_{ijv}$, $\Pr\left(\Pi_i^{max} > \pi\right)$ can be simplified to:

$$\Pr\left(\Pi_i^{max} > \pi\right) = 1 - \exp\left(-\tilde{\Phi}_i \pi^{-\varepsilon}\right) \tag{48}$$

Thus, the CDF of max profits for village i is written as:

$$G_i^{max}(\pi) = \Pr\left(\Pi_i^{max} < \pi\right) = \exp\left(-\tilde{\Phi}_i \pi^{-\varepsilon}\right) \tag{49}$$

with pdf:

$$g_i^{max}(\pi) = \varepsilon \tilde{\Phi}_i \pi^{-\varepsilon - 1} \exp\left(-\tilde{\Phi}_i \pi^{-\varepsilon}\right) \tag{50}$$

Similarly, adding the option of not adopting, the distribution of profits considering all options, Π_i, is written as:

$$\Pr\left(\Pi_i > \pi\right) = \Pr\left(\Pi_{ijv} > \pi \text{ for any } jv \cup \Pi_{i0} > \pi\right) \tag{51}$$

$$= 1 - \Pr\left(\Pi_{ijv} < \pi \ \forall \ jv \ \cap \ \Pi_{i0} < \pi\right) \tag{52}$$

Since θ's at each j, v pair and for not adopting are drawn from independent distributions, this probability is simplified as:

$$\Pr\left(\Pi_i > \pi\right) = 1 - \Pr\left(\Pi_{ijv} < \pi \ \forall \ jv \ \cap \ \Pi_{i0} < \pi\right) \tag{53}$$

$$= 1 - \Pr\left(\Pi_{i0} < \pi\right) \prod_{v' \in \mathcal{V}} \prod_{j \in \mathcal{J}_v} \Pr\left(\Pi_{ijv} < \pi\right) \tag{54}$$

$$= 1 - \exp\left(-T_{i0}\gamma_{i0}\pi^{-\varepsilon}\right) \prod_{v' \in \mathcal{V}} \prod_{j \in \mathcal{J}_v} \exp\left(-T_{jv}\gamma_{ijv}\pi^{-\varepsilon}\right) \tag{55}$$

Using the definitions for $\tilde{\Phi}_{i0}$ and $\tilde{\Phi}_i$, this is simplified as:

$$\Pr\left(\Pi_i > \pi\right) = 1 - \exp\left(-\left(\tilde{\Phi}_{i0} + \tilde{\Phi}_i\right)\pi^{-\varepsilon}\right) \tag{56}$$

Thus, the CDF of max profits for village i is:

$$G_i\left(\pi\right) = \exp\left(-\left(\tilde{\Phi}_{i0} + \tilde{\Phi}_i\right)\pi^{-\varepsilon}\right) \tag{57}$$

with pdf:

$$g_i\left(\pi\right) = \varepsilon\left(\tilde{\Phi}_{i0} + \tilde{\Phi}_i\right)\pi^{-\varepsilon-1}\exp\left(-\left(\tilde{\Phi}_{i0} + \tilde{\Phi}_i\right)\pi^{-\varepsilon}\right) \tag{58}$$

B.3.1 Profits conditional on adoption

Using this pdf, we now derive the CDF of agrovet profits, conditional on adoption. To do this, we start from the conditional probability formula:

$$\Pr\left(\Pi_i^{max} < \pi \middle| adopt\right) = \frac{\Pr\left(\Pi_i^{max} < \pi \cap \Pi_i^{max} > \Pi_{i0}\right)}{\Pr\left(\Pi_i^{max} > \Pi_{i0}\right)} \tag{59}$$

This can be re-written as:

$$\begin{aligned}
\Pr\left(\Pi_i^{max} < \pi \middle| adopt\right) &= \frac{1}{\Pr\left(\Pi_i^{max} > \Pi_{i0}\right)}\int_0^{\pi}\Pr\left(s > \Pi_{i0}\right)g_i^{max}(s)ds \\
&= \frac{1}{\Pr\left(\Pi_i^{max} > \Pi_{i0}\right)}\int_0^{\pi}\exp\left(-\tilde{\Phi}_{i0}s^{-\varepsilon}\right)\varepsilon\tilde{\Phi}_i s^{-\varepsilon-1}\exp\left(-\tilde{\Phi}_i s^{-\varepsilon}\right)ds \\
&= \frac{1}{\Pr\left(\Pi_i^{max} > \Pi_{i0}\right)}\int_0^{\pi}\varepsilon\tilde{\Phi}_i s^{-\varepsilon-1}\exp\left(-\left(\tilde{\Phi}_{i0} + \tilde{\Phi}_i\right)s^{-\varepsilon}\right)ds \qquad (60)
\end{aligned}$$

Mulitplying by $\frac{\tilde{\Phi}_{i0}+\tilde{\Phi}_i}{\tilde{\Phi}_{i0}+\tilde{\Phi}_i}$, and then factoring out $\frac{\tilde{\Phi}_i}{\tilde{\Phi}_{i0}+\tilde{\Phi}_i}$, we have:

$$\Pr\left(\Pi_i^{max} < \pi \middle| adopt\right) = \frac{1}{\Pr\left(\Pi_i^{max} > \Pi_{i0}\right)}\frac{\tilde{\Phi}_i}{\tilde{\Phi}_{i0} + \tilde{\Phi}_i}\int_0^{\pi}\varepsilon\left(\tilde{\Phi}_{i0} + \tilde{\Phi}_i\right)s^{-\varepsilon-1}\exp\left(-\left(\tilde{\Phi}_{i0} + \tilde{\Phi}_i\right)s^{-\varepsilon}\right)ds$$

From standard derivations using Fréchet, $\Pr\left(\Pi_i^{max} > \Pi_{i0}\right) = \frac{\tilde{\Phi}_i}{\tilde{\Phi}_{i0}+\tilde{\Phi}_i}$, and thus:

$$\begin{aligned}
\Pr\left(\Pi_i^{max} < \pi \middle| adopt\right) &= \int_0^{\pi}\varepsilon\left(\tilde{\Phi}_{i0} + \tilde{\Phi}_i\right)s^{-\varepsilon-1}\exp\left(-\left(\tilde{\Phi}_{i0} + \tilde{\Phi}_i\right)s^{-\varepsilon}\right)ds \qquad (61) \\
&= \Pr\left(\Pi_i < \pi\right) \qquad (62)
\end{aligned}$$

B.3.2 Profits conditional on adoption from j

Next, we derive the expected profits, conditional on adopting fertilizer from location j. Precisely, we will derive:

$$\Pr\left(\Pi_{ijv} < \pi \middle| adopt\ from\ j\,in\,v\right) = \frac{\Pr\left(\Pi_{ijl} < \pi \cap \Pi_{ijv} > \Pi_{ij'l} \forall (j',l) \cap \Pi_{ijv} > \Pi_{i0}\right)}{\Pr\left(\Pi_{ijv} > \Pi_{ij'l} \forall (j',l) \cap \Pi_{ijv} > \Pi_{i0}\right)} \quad (63)$$

The denominator in this equation is simply λ_{ijv}, and thus, we factor it out of the probability. The numerator is written similar to the previous derivation, where

$$\Pr\left(\Pi_{ijv} < \pi \middle| adopt\ from\ j\,in\,v\right) = \frac{1}{\lambda_{ijv}} \int_0^{\pi} Pr\left(s > \Pi_{ij'l} \forall (j',l) \cap s > \Pi_{i0}\right) g_{ijv}(s)\, ds \quad (64)$$

Defining $\tilde{\tilde{\Phi}}_{ijv} = \left(\sum_{v' \in \mathcal{V}} \sum_{j \in \mathcal{J}_v} T_{jv}\gamma_{ijv}\right) - T_{jv}\gamma_{ijv}$, we can simplify $\Pr\left(s > \Pi_{ij'l} \forall (j',l) \cap s > \Pi_{i0}\right)$ as

$$\begin{aligned}
\Pr\left(s > \Pi_{ij'l} \forall (j',l) \cap s > \Pi_{i0}\right) &= \exp\left(-\tilde{\Phi}_{i0} s^{-\varepsilon}\right) \exp\left(-\tilde{\tilde{\Phi}}_{ijv} s^{-\varepsilon}\right) && (65)\\
&= \exp\left(-\left(\tilde{\Phi}_{i0} + \tilde{\tilde{\Phi}}_{ijv}\right) s^{-\varepsilon}\right) && (66)
\end{aligned}$$

Thus, $\Pr\left(\Pi_{ijv} < \pi \middle| adopt\ from\ j\right)$ is written as:

$$\Pr\left(\Pi_{ijv} < \pi \middle| adopt\ from\ j\right) = \frac{1}{\lambda_{ijv}} \int_0^{\pi} \exp\left(-\left(\tilde{\Phi}_{i0} + \tilde{\tilde{\Phi}}_{ijv}\right) s^{-\varepsilon}\right) \varepsilon T_{jv}\gamma_{ijv} \pi^{-\varepsilon-1} \exp\left(-T_{jv}\gamma_{ijv} s^{-\varepsilon}\right) ds$$

Factoring out $\frac{T_{jv}\gamma_{ijv}}{\tilde{\Phi}_{i0}+\tilde{\Phi}_i}$, and then noting that $\tilde{\Phi}_{i0} + \tilde{\Phi}_i = \tilde{\Phi}_{i0} + \tilde{\tilde{\Phi}}_{ijv} + T_{jv}\gamma_{ijv}$, we have:

$$\Pr\left(\Pi_{ijv} < \pi \middle| adopt\ from\ j\right) = \frac{1}{\lambda_{ijv}} \frac{T_{jv}\gamma_{ijv}}{\tilde{\Phi}_{i0} + \tilde{\Phi}_i} \int_0^{\pi} \varepsilon \left(\tilde{\Phi}_{i0} + \tilde{\Phi}_i\right) \pi^{-\varepsilon-1} \exp\left(-\left(\tilde{\Phi}_{i0} + \tilde{\Phi}_i\right) s^{-\varepsilon}\right) ds$$

Since $\lambda_{ijv} = \frac{T_{jv}\gamma_{ijv}}{\tilde{\Phi}_{i0}+\tilde{\Phi}_i}$, we land at the final result:

$$\begin{aligned}
\Pr\left(\Pi_{ijv} < \pi \middle| adopt\ from\ j\right) &= \int_0^{\pi} \varepsilon \left(\tilde{\Phi}_{i0} + \tilde{\Phi}_i\right) \pi^{-\varepsilon-1} \exp\left(-\left(\tilde{\Phi}_{i0} + \tilde{\Phi}_i\right) s^{-\varepsilon}\right) ds\\
&= \Pr\left(\Pi_i < \pi\right)
\end{aligned}$$

Thus, the distribution of profits adopting from j is the same as the distribution of profits adopting anywhere.

B.4 Production Function Estimation with and without Fertilizer

As our dataset is not equipped for panel production function estimation, we will be using the Tanzanian LSMS, which records output and input use by household-plot-time, and we exposit the estimation accordingly. That is, the production functions under different technologies should now be understood to be specific to a particular plot within a household. Simply manipulating the Cobb-Douglas production functions for plot p of household i in time t, we get the following representation for output per unit of land:

$$\begin{aligned} \log\left(\frac{Y_{ipt}}{K_{ipt}}\right) &= (1-\alpha_0)\log\left(\frac{L_{ipt}}{K_i}\right) \\ \log\left(\frac{Y_{ipt}}{K_{ipt}}\right) &= (1-\alpha)\beta\log\left(\frac{L_{ipt}}{K_{ipt}}\right)+(1-\alpha)(1-\beta)\log\left(\frac{M_{ipt}}{K_{ipt}}\right) \end{aligned}$$

To combine these equations into one specification, we need to eliminate $\log\left(\frac{M_{ipt}}{K_{ipt}}\right)$, which is not defined when fertilizer is not purchased. However, exploiting the fact that relative demand for fertilizer and labor is a constant function of local wages, delivered fertilizer prices and parameters, we can write:

$$\begin{aligned} \log\left(\frac{Y_{ipt}}{K_{ipt}}\right) &= (1-\alpha_0)\log\left(\frac{L_{ipt}}{K_i}\right) \\ \log\left(\frac{Y_{ipt}}{K_{ipt}}\right) &= (1-\alpha)\beta\log\left(\frac{L_{ipt}}{K_{ipt}}\right)+d_{it} \end{aligned}$$

where d_{it} is a dummy variable for household i, and year t (that is meant to absorb local wages and prices when using fertilizer). This motivates the following specification to test for differences in production parameters with and without fertilizer.

$$\log\left(\frac{Y_{ipt}}{K_{ipt}}\right)=(1-\alpha_0)\log\left(\frac{L_{ipt}}{K_{ipt}}\right)+(\alpha_0-\alpha)\log\left(\frac{L_{ipt}}{K_{ipt}}\right)\cdot\mathbf{I}(M_{ipt}>0)+DFT_{ipt}+Plot_{ip}+u_{ipt}$$

Here, DFT_{ipt} is a district-time variable, with and without fertilizer use, meant to absorb differences is local wages and prices, and other local and shocks, that may vary by time and whether fertilizer is used. While one could argue that local wages and prices should vary at a more granular level, this is about as far as we can push the data given the other sets of fixed effects that are utilized. $Plot_{ip}$ is a fixed effect to absorb plot-specific sources of productivity differences. Within these fixed effects, we estimate α_0 and α using labor per unit of land and an interaction with a dummy variable identifying fertilizer use. Appendix Table B.7 reports these estimates. In the preferred specification, we find that $\alpha_0 = 0.57$ and $\alpha = 0.421$.

The last production parameter to estimate is the expenditure share of labor compared relative to total expenditures on labor and fertlizer. For this measure, we also use the Tanzanian LSMS. We first average district-level, activity specific wages from all plots that hire labor, and then construct an implied labor cost on each plot by summing the product of labor hours on each activity and the average wage for that activity. Then, for those who adopt fertilizer, we divide the value of fertilizer used on that plot by the sum of this same value and implied labor expenditure. For the whole of Tanzania, the average of this fertilizer expenditure share is 0.25, and we use this value for our counterfactuals by imposing that $\beta = 0.75$. Of note, the mean and median values of fertilizer expenditure share for the subsample of regions in norther Tanzania (Arusha, Kilimanjaro, Manyara, Tanga) is slightly higher at 0.28.

B.5 Mark-ups

From above, we can write the expected fertilizer revenues for agrovet j in location v as:

$$\mathbb{E}[v_{jv}] = \sum_i \mu_i \lambda_{ijv|adopt} \mathbb{E}[F_i | adopt\ at\ jv]$$

Since fertilizer expenditures are proportional to profits, and profits are invariant to the choice that is made (in expectation)we have:

$$\mathbb{E}[v_{jv}] = (1-\beta)\frac{1-\alpha}{\alpha}\sum_i \lambda_{ijv}\mathbb{E}[\Pi_i] \tag{67}$$

Differentiating with respect to the fertilizer price, r_{jv}, the elasticity of expected revenues with respect to own price is:

$$\frac{d\mathbb{E}[v_{jv}]}{dr_{jv}}\frac{r_{jv}}{\mathbb{E}[v_{jv}]} = \sum s_{ijv}\left(\frac{d\lambda_{ijv}}{dr_{jv}}\frac{r_{jv}}{\lambda_{ijv}} + \frac{d\mathbb{E}[\Pi_i]}{dr_{jv}}\frac{r_{jv}}{\mathbb{E}[\Pi_i]}\right) \tag{68}$$

where $s_{ijv} = \frac{\lambda_{ijv|adopt}\mathbb{E}[rm_i]}{\sum_{i'}\lambda_{i'jv|adopt}\mathbb{E}[rm_{i'}]}$. As a function of model parameters, $\frac{d\lambda_{ijv}}{dr_{jv}}\frac{r_{jv}}{\lambda_{ijv}}$ is written as:

$$\frac{d\lambda_{ijv}}{dr_{jv}}\frac{r_{jv}}{\lambda_{ijv}} = -\varepsilon_a(1-\lambda_{ijv})$$

Given the assumption of the Frechet distribution, $\mathbb{E}[\Pi_i]$ can be written as:

$$\mathbb{E}[\Pi_i] = \kappa(\Phi_{i0}+\Phi_i)^{\frac{1}{\varepsilon}}$$

where κ is a function of distribution parameters. Log-differentiating, it is straightforward to show that:

$$\frac{d\mathbb{E}[\Pi_i]}{dr_{jv}}\frac{r_{jv}}{\mathbb{E}[\Pi_i]} = -\frac{\varepsilon_a}{\varepsilon}\lambda_{ijv}$$

Thus, the elasticity of expected revenues to price can be written as:

$$\frac{d\mathbb{E}[v_{jv}]}{dr_{jv}}\frac{r_{jv}}{\mathbb{E}[v_{jv}]} = -\varepsilon_a\sum_i s_{ijv}\left((1-\lambda_{ijv})+\frac{1}{\varepsilon}\lambda_{ijv}\right) \tag{69}$$

Since $\sum_i s_{ijv} = 1$ for each jv, the elasticity of expected revenues to price can be simplified as:

$$\varepsilon_v \equiv \frac{d\mathbb{E}[v_{jv}]}{dr_{jv}}\frac{r_{jv}}{\mathbb{E}[v_{jv}]} = -\varepsilon_a + \frac{\varepsilon-1}{\varepsilon}\varepsilon_a\sum_i s_{ijv}\lambda_{ijv} \tag{70}$$

B.6 External Validity

B.6.1 Price Dispersion

To compare price dispersion in our study region to other parts of Africa, we bring in evidence from five secondary datasets of prices in 1,512 markets in 56 African countries.[85]

[85] We include the following datasets: (1) prices of 6 staple crops in 41 major market centers in 8 East African countries from 1997-2015, collected by RATIN; (2) prices of 25 commodities from 276 markets in 53 countries in from 2013-2015, collected by Africafoodprices.io; (3) prices of 4 major varieties of fertilizer (Urea, DAP, CAN, and NPK complex 17-17-17) in 129 markets in 7 East African countries collected by AMITSA; (4) prices of 5 major varieties of fertilizer (Urea, CAN, DAP, and NPK 17 17 17) in 18 countries from 2010-16 in Africafertilizer.org; and (5) prices of a number of commodities in 38 countries from 1992-2016 collected by the WFP.

We compare this to a small dataset we assembled between March and April 2016 with 251 retailers of various sorts (shops, agrovets, and maize traders) in 82 markets in the Kilimanjaro region.[86] To quantify price dispersion, we decompose variation in spatial prices by running the following regression:

$$\log(p_{mcjt}) = \gamma_c + \gamma_j + \gamma_t + \varepsilon_{mjt} \tag{71}$$

where p_{mcjt} (log) prices in market m for product j at time t in country c, and the γ terms are country, product, and time fixed effects. We report the standard deviation of the resulting residual in Appendix Table B.9. In the secondary datasets, the standard deviation is 0.45 for all products, 0.34 for maize, and 0.12 for fertilizer; in our Tanzania data, the figures are 0.22, 0.14, and 0.09. While price dispersion is lower in our data (perhaps because of reduced measurement error, or that prices vary less within the geographic concentrated area of Kilimanjaro), this simple exercise sugggests substantial price dispersion in Northern Tanzania.

We also follow the literature,[87] to run dyadic regressions to look at price gaps, as follows:

$$\log(|\, p_{mjt} - p_{m'jt} \,|) = \theta \log(c_{mm'}) + \gamma_m + \gamma_{m'} + \gamma_j + \varepsilon_{mm'jt} \tag{72}$$

where $p_{mjt} - p_{m'jt}$ is the price gap between markets m and m' and $c_{mm'}$is the cost of transport between markets.[88]

Results are presented in Appendix Table B.10. For each dyad, we regress the absolute difference in log prices on two measures of distance: (1) kilometers between locations in Columns 1, 4, and 7, and (2) driving time between locations in Columns 2, 5, and

[86] To enroll participants, we visited each market and selected several types of retailers for project inclusion, including fertilizer retailers ("agrovets"), maize sellers, and retail shops. Each respondent was called once per month and asked about current retail and wholesale prices for each item in a pre-selected list of standardized goods (e.g., 200 ml box of Azam juice). Respondents were compensated for participation by mobile money transfer.

[87] See Engel and Rogers (1996). In addition, see papers on the effect of cell phones on price dispersion, for example Aker (2010), Aker and Fafchamps (2015), and Jensen (2007).

[88] These regressions are motivated by an assumption of free entry where an arbitrageur will enter if $|\,(p_m - p_{m'})\,| \geq c_{mm'}$.

8 (both calculated via Google Maps API), and we cluster standard errors by both the destination and origin market. In each of the secondary datasets, we find significant, positive coefficients, suggesting that price gaps are larger between more distant markets. The coefficients are economically meaningful: a doubling of travel costs would increase price gaps by about 1-3% in the secondary datasets. In Tanzania, we find that doubling distances would increase price gaps by a similar amount. We can also use this data to provide some descriptive evidence on road upgrading. We conjecture that price gaps should respond to the time it takes to travel from point to point, and not the geographic distance (since the time and other costs of traveling to sell items should be what is important). To examine this, we regress price gaps on both distance and duration in Columns 3, 6, and 9. Consistent with priors, we find that duration is significant, whereas distance is not – which suggests that improving road quality would reduce these gaps.

B.6.2 Fertilizer adoption

Finally, we use data assembled data from World Bank LSMS-ISA household panel surveys to study how remoteness affects fertilizer adoption in other African countries.[89] Using both measures of remoteness available in the dataset (distance to the main market, and distance to a population center), we find a negative association between remoteness and technology adoption (Appendix Table B.11).

[89] The countries included here are Ethiopia, Niger, Nigeria, Malawi, Tanzania, and Uganda.

Table B.1: Survey Compliance Rates

	(1) Survey Attempts	(2) Completed	(3) Compliance Rate
Farmer surveys 2016	583	573	0.98
Farmer surveys 2017	2535	2477	0.98
Agrovet surveys	585	532	0.91
Maize sellers at markets	445	438	0.98

Notes: See text of details of surveys.

Table B.2: Costs of Transporting Fertilizer and Transporting Farmer, by Distance

	(1)	(2)	(3)	(4)
	Cost of transporting fertilizer from agrovet in destination village (standardized to 50 kg)		Cost of farmer traveling himself to agrovet	
Google maps: kilometers to destination	0.04*		0.05***	
	(0.02)		(0.01)	
Google maps: hours to destination		1.28*		1.83***
		(0.68)		(0.26)
Number of villages	73	73	119	119
Number of observations	341	341	988	988

Notes: Data is constructed from Farmer Surveys, conditional on making input purchases and/or selling output. Clustered standard errors (by village) are reported in parentheses.
*, **, and *** indicate significance at 10%, 5%, and 1% respectively.

Table B.3: Locations of Input and Output Distributors

Panel A. Locations of Agro-Input Distributors

Locations	Share of Retailer Revenues	Cum. Share
Arusha Urban District	0.80	0.80
Kilimanjaro Moshi Urban District	0.14	0.94
Manyara Babati Urban District	0.02	0.96
Dar es Salaam Kinodoni District	0.01	0.97
Dar es Salaam Ilala District	0.01	0.98

Panel B. Locations of Output Distributors

B1. 2017 Maize Store Census

Locations	Share of Maize Purchase	Cum. Share
Arusha Urban District	0.61	0.61
Manyara Babati Rural District	0.35	0.97
Kilimanjaro Hai District	0.02	0.98
Manyara Babati Urban District	0.01	0.99
Arusha Rural District	0.01	1.00

B2. 2016 Maize Store Census

Locations	Share of Maize Purchase	Cum. Share
Kilimanjaro Moshi Rural District	0.74	0.74
Arusha Urban District	0.13	0.87
Manyara Babati Urban District	0.10	0.97
Manyara Babati Rural District	0.02	0.99

Notes: Locations of agro-input distributors are based on the surveys conducted on the universe of agro-input retailers. Locations of output distributors are based on the maize store censuses we conducted in both year 2016 and 2017.

Table B.4: Summary Statistics of Market Access Proxies

	(1) Remoteness measured by distance
Remoteness measured by elasticity-adjusted travel costs to hubs	0.84***
	(0.02)
Dependent variable mean before standardization	304.19
Dependent variable sd before standardization	31.56
Independent variable mean before standardization	-0.11
Independent variable sd before standardization	0.04
Observations	1,135

Notes: The regression is run at the village level. In all reduced-form regressions in the paper, the Donaldson-Hornbeck remoteness proxy is multiplied by -1 for consistent interpretation with the results from standardized distance remoteness. The regression coefficient is standardized. Standard errors in parentheses. *, **, and *** indicate significance at 10%, 5%, and 1%.

Table B.5: Remoteness and Fertilizer Retailer Sales, Prices, and Markups

	(1)	(2)	(3)
	Mean	(Standardized) coefficient from regression of dependent variable on remoteness measure based on (population-weighted):	
		Distance to hubs	Elasticity-adjusted travel costs to hubs
Panel A. Agrovet shop-level (N=509)			
Sells fertilizer	0.87	-0.03	-0.07***
	(0.34)	(0.02)	(0.02)
Number of varieties of fertilizer	1.67	-0.06	-0.11
	(1.54)	(0.09)	(0.09)
Quantity of fertilizer sold last year (kg)	5588	-250.26	-581.15
	(11642)	(707.66)	(755.20)
Sells seeds	0.72	0.03	0.07***
	(0.45)	(0.02)	(0.03)
Number of varieties of seeds	1.2	0.10	0.28***
	(1.26)	(0.07)	(0.07)
Quantity of seeds sold last year (kg)	2194	903.63	1,657.90***
	(8008)	(557.24)	(442.65)
Cost of transport from wholesaler (per 50 kg)	0.64	0.32***	0.34***
	(0.69)	(0.04)	(0.04)
Panel B. Prices and markups (Agrovet shop-variety level, N=938)			
Retail price for 50 kilograms	25.21	0.65***	0.54**
	(5.21)	(0.22)	(0.23)
Wholesale price for 50 kilograms	21.43	0.16*	0.20**
	(4.14)	(0.09)	(0.09)
Markup (percentage points)[1]	13.42	0.86	0.42
	(10.25)	(0.62)	(0.69)

Notes: In Column 1, standard deviations are in parentheses. Columns 2 and 3 show regression coefficients from separate regressions of the dependent variable on a measure of remoteness (equations 5 and 6 in the paper). See text for further discussion of these measures. Regressions in Panel B includes type and brand fixed effects.

*, **, and *** indicate significance at 10%, 5%, and 1%.

[1]Markup accounts for cost of transport to wholesaler.

Table B.6: Robustness of Travel-cost Adjusted Prices

	(1)	(2)	(3)
	Mean	(Standardized) coefficient from regression of dependent variable on remoteness measure based on (population-weighted):	
		Distance to hubs	Elasticity-adjusted travel costs to hubs
Panel A. Robustness to Dropping Villages Within 10km of Regional Borders			
A1. Input Side: Travel-cost adjusted fertilizer prices faced by farmers			
Minimum travel-cost adjusted price for 50 kg of Urea	23.98	2.70***	2.64***
	(4.44)	(0.14)	(0.13)
Decomposition of price between retail price and cost of transportation			
Retail price at the location with the lowest travel-cost adjusted price (USD)	19.91	1.32***	1.40***
	(2.67)	(0.09)	(0.08)
Cost of travel to obtain minimum travel-cost adjusted price (USD)	4.069	1.38***	1.24***
	(3.99)	(0.14)	(0.13)
A2. Output Side: Travel-cost adjusted maize prices if farmers were to sell in a local market			
Market survey: maximum travel-cost adjusted price immediately after 2017 harvest (USD)	30.07	-3.71***	-3.53***
	(7.18)	(0.23)	(0.22)
Decomposition of price between retail price and cost of transportation			
Retail price at the location with the highest travel-cost adjusted price (USD)	39.10	0.89***	0.23**
	(3.20)	(0.12)	(0.11)
Cost of travel to obtain the highest travel-cost adjusted price (USD)	9.03	4.60***	3.76***
	(7.18)	(0.21)	(0.21)
Panel B. Bounding regression coefficients by assigning prices to missing retailers[1]			
Input Side: Travel-cost adjusted fertilizer prices faced by farmers			
Minimum travel-cost adjusted price for 50 kg of Urea	24.09	2.26***	2.40***
	(4.69)	(0.13)	(0.12)
Decomposition of price between retail price and cost of transportation			
Retail price at the location with the lowest travel-cost adjusted price (USD)	19.84	1.10***	1.26***
	(2.58)	(0.07)	(0.07)
Cost of travel to obtain minimum travel-cost adjusted price (USD)	4.25	1.16***	1.14***
	(4.35)	(0.13)	(0.13)

Notes: Data is from the universe of villages in Kilimanjaro and Manyara region (N = 1,183). The unit of observation is the village. Travel costs imputed from transport surveys and Google maps. In Column 1, standard deviations are in parentheses. Columns 2 and 3 show regression coefficients from separate regressions of the dependent variable on a measure of remoteness (equations 5 and 6 in the paper). See text for further discussion of these measures.

*, **, and *** indicate significance at 10%, 5%, and 1%.

[1] In this calculation, we imputed prices to retailers with missing values. To do this, we estimated the distribution of prices within region. We then assigned high or low prices to the missing agrovet (defined as being at the 10th or 90th percentile of this price distribution) in a way that attenuated the regression coefficient. For example, a missing agrovet in a remote village was assigned a low price, causing a flattening of the regression.

Table B.7: Estimating Elasticity of Subsitution Across Agrovets Using IV

Panel A: Elasticity Estimation for Calibration

	(1)	(2)	(3)	(4)
	Dependent variable: log(Eta) (from Calibration)			
log(Price)	-5.566***	-7.516**	-4.731***	-8.121**
	(0.990)	(3.729)	(0.960)	(3.330)
log(Experience)			0.938***	0.870***
			(0.154)	(0.171)
	OLS	IV	OLS	IV
R-squared	0.37	0.35	0.43	0.4
Cragg-Donald Wald F statistic		33.93		32.34
Wu-Hausman		0.35		1.22
Observations	374	242	374	242

Notes: District fixed effects used in all regressions. Instrumental variable is 2011 population in the market catchment area from the Tanzanian census. ***, **, and * indicate significance at 1%, 5%, and 10%.

Panel B: Calibrated Market Access Terms and Remoteness

	(1)	(2)
	log(Market Access)	log(Outside Option)
Distance Remoteness	-0.777***	0.282
	(0.274)	(0.236)
R-squared	0.056	0.01
Observations	137	137

Notes: ***, **, and * indicate significance at 1%, 5%, and 10%.

Table B.8: Production Function Estimates with and Without Fertilizer

	(1)	(2)	(3)
	Dependent variable: log(Harvest/Acres)		
log(Labor/Acres)	0.42***	0.43***	0.43***
	(0.04)	(0.04)	(0.04)
log(Labor/Acres) x Used Fertilizer?	0.12*	0.12*	0.15*
	(0.08)	(0.07)	(0.08)
Used Fertilizer?	(0.33)	(0.33)	
	(0.30)	(0.30)	
District-Year fixed effects		X	
District-Year-Fertilizer Use fixed effects			X
Plot fixed effects	X	X	X
Observations	3,395	3,395	3,395
Plots	2,554	2,554	2,554

Notes: Regressions use World Bank LSMS-ISA household panel surveys from Tanzania, and Uganda. ***, **, and * indicate significance at 1%, 5%, and 10%.

Table B.9: Input and Output Market Price Dispersion Across Countries

	(1) Secondary Datasets[1]	(2) Tanzania Data[2]
Residual standard deviation in log prices for:[3]		
All products	0.45	0.15
Maize only	0.34	0.10
Fertilizer only	0.12	0.09

Notes: [1]Secondary datasets include RATIN (prices of major crops across 41 major markets in 5 countries - Kenya, Tanzania, Uganda, Burundi, and Rwanda - over the 1997-2015 time period), Africafoodprices.io (25 products over 276 markets in 53 countries), AMITSA (the Regional Agricultural Input Market Information and Transparency System for East and Southern Africa, which includes information on 9 fertilizer varieties in 95 markets in 8 countries), prices of 5 major varieties of fertilizer (Urea, CAN, DAP, and NPK 17 17 17) in 18 countries from 2010-16 in Africafertilizer.org; and prices of a number of commodities in 38 countries from 1992-2016 collected by the WFP.

[2]Maize prices are from a survey of market sellers in 98 markets conducted in October 2017. Fertilizer prices are from surveys of agro-input retailers in 2017.

[3]Calculated from a regression of log prices on product, country, and time fixed effects. See text for details.

Table B.10: Dyadic Price Dispersion

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Dependent variable: Absolute log price difference								
Panel A. Secondary									
Log (distance)	0.03***	0.000	0.000	0.03***	0.000	0.000	0.01***	0.000	0.010
	(0.002)	0.000	(0.010)	(0.002)	0.000	(0.015)	(0.002)	0.000	(0.014)
Log (travel time)		0.03***	0.03***		0.04***	0.04**		0.01***	0.000
		(0.002)	(0.011)		(0.003)	(0.017)		(0.002)	(0.016)
Products	All	All	All	Maize	Maize	Maize	Fertilizer	Fertilizer	Fertilizer
Dependent variable mean	0.21	0.21	0.21	0.20	0.20	0.20	0.11	0.11	0.11
Dependent variable sd	0.20	0.20	0.20	0.17	0.17	0.17	0.13	0.13	0.13
Observations	4,752,196	4,752,196	4,752,196	675,880	675,880	675,880	38,364	38,364	38,364
Number of locations	1335	1335	1335	1335	1335	1335	1335	1335	1335
Countries	49	49	49	43	43	43	18	18	18
Panel B. Northern									
Log (distance)	0.01***		-0.030	0.03***		-0.10**	0.003*		0.007
	(0.003)		(0.020)	(0.011)		(0.050)	(0.002)		(0.017)
Log (travel time)		0.01***	0.04*		0.04***	0.16**		0.004	-0.004
		(0.004)	(0.025)		(0.016)	(0.069)		(0.002)	(0.019)
Products	All	All	All	Maize	Maize	Maize	Fertilizer	Fertilizer	Fertilizer
Dependent variable mean	0.16	0.16	0.16	0.21	0.21	0.21	0.13	0.13	0.13
Dependent variable sd	0.14	0.14	0.14	0.18	0.18	0.18	0.10	0.10	0.10
Observations	22,386	22,376	22,376	6,873	6,873	6,873	15,064	15,056	15,056
Number of locations	82	82	82	65	65	65	60	60	60

Notes: Regressions include product, month and year fixed effects. All regressions are within country. Travel time and distances calculated from Google maps. See Web Appendix Table A3 and text for discussion of datasets.
Two-way clustered standard errors in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10%.

Table B.11: Adoption in LSMS-ISA Surveys

	(1)	(2)
	Dependent variable: used chemical fertilizer in last season	
Log of distance to nearest major market (km)	-0.027***	
	(0.005)	
Log of distance to nearest population center (km)		-0.019*
		(0.010)
Dependent variable mean	0.32	0.32
Independent variable mean	3.23	3.21
Independent variable sd	1.27	1.02
Observations	35,938	35,938
Individuals	26,653	26,653

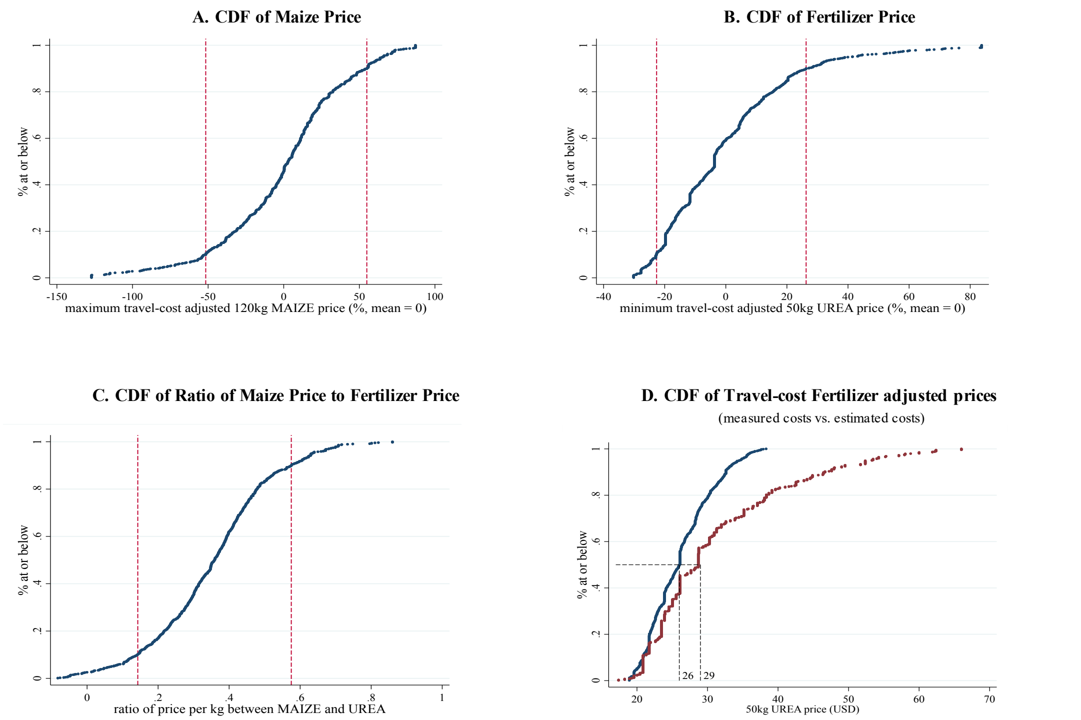
Notes: Regressions include World Bank LSMS-ISA household panel surveys in Ethiopia, Niger, Nigeria, Malawi, Tanzania, and Uganda. Standard errors clustered at the enumeration area level are in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10%.

Table B.12: Road Density in East Africa

	(1) Road Density (km per '00 sq km area)	(2) Percentage Roads Paved
Burundi	44.28	12.17%
Democratic Republic of Congo	6.54	1.82%
Djibouti	13.21	40.00%
Eritrea	3.41	21.80%
Ethiopia	10.00	13.00%
Kenya	27.82	8.93%
Madagascar	11.18	11.60%
Malawi	13.04	26.37%
Mozambique	3.88	23.70%
Rwanda	17.84	25.68%
Somalia	3.47	11.80%
South Sudan	1.13	2.74%
Tanzania	9.13	8.20%
Uganda	8.52	20.72%
Zambia	12.15	22.00%
Zimbabwe	24.90	19.00%
Sub-Saharan Africa Average	13.70	22.63%

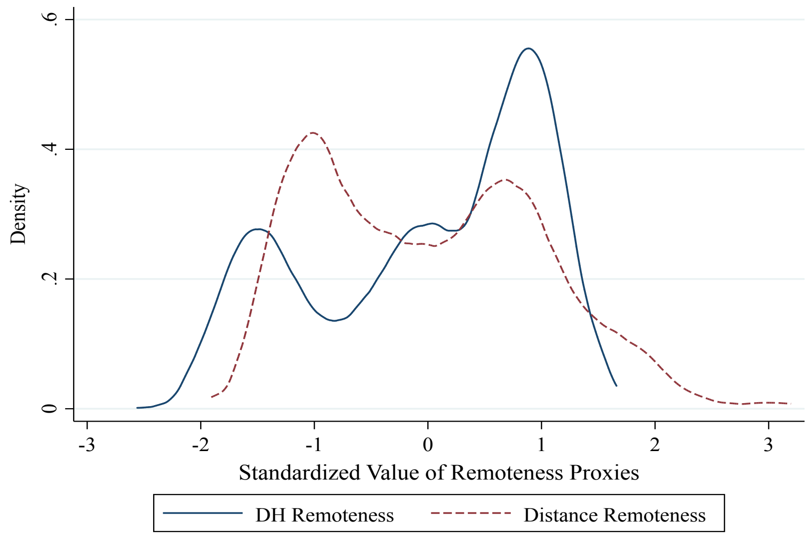
Notes: Data compiled from various World Bank and AfDB reports. Statistics correspond to years ranging between 2010 and 2016; DRC statistics are from 2001. We include all countries classified as Eastern African as per the United Nations Statistics Division scheme of geographic regions, except the island nations of Comoros, Mauritius and Seychelles, and the French Overseas Territories of Réunion and Mayotte. We also exclude Sudan because there is no data available for after it split from South Sudan.

Figure B.1: CDF of Travel-cost Adjusted Prices at the Nearest Locations



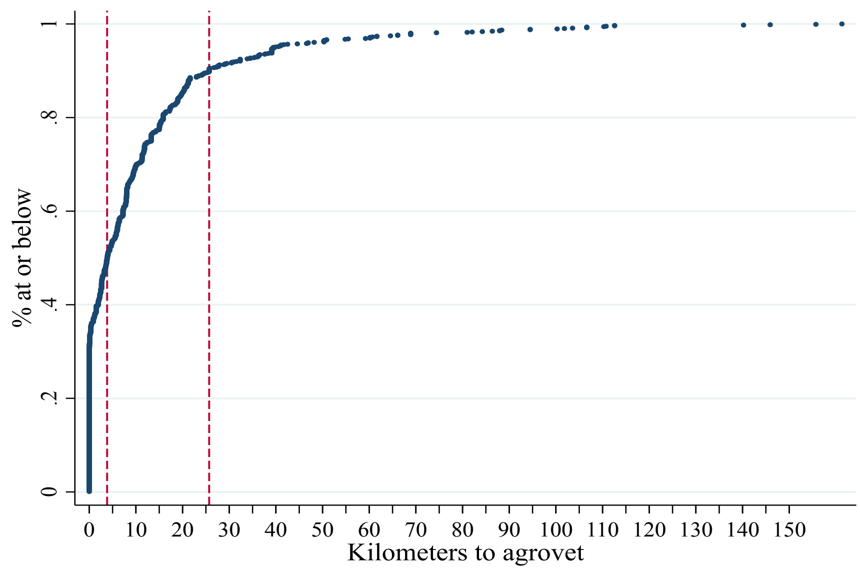
Notes: Each observation represents a village. Travel-cost adjusted prices are calculated through observed prices from an agrovet survey, a maize price survey at markets and transport cost information collected from interviews with transport operators. Vertical dotted lines represent a 10 percentile and a 90 percentile of the distribution.

Figure B.2: The Distribution of Remoteness Proxies



Notes: The distribution of remoteness proxies is depicted at the village level (N=1,135).

Figure B.3: CDF of Distance Farmers Travel to Purchase Inputs



Notes: Each point represents a farmer. Purchase events include any kinds of agricultural inputs. Vertical dotted lines indicate distances corresponding to the the 50th and 90th percentile.

C Appendix for Chapter 3

C.1 Additional Graphs and Tables

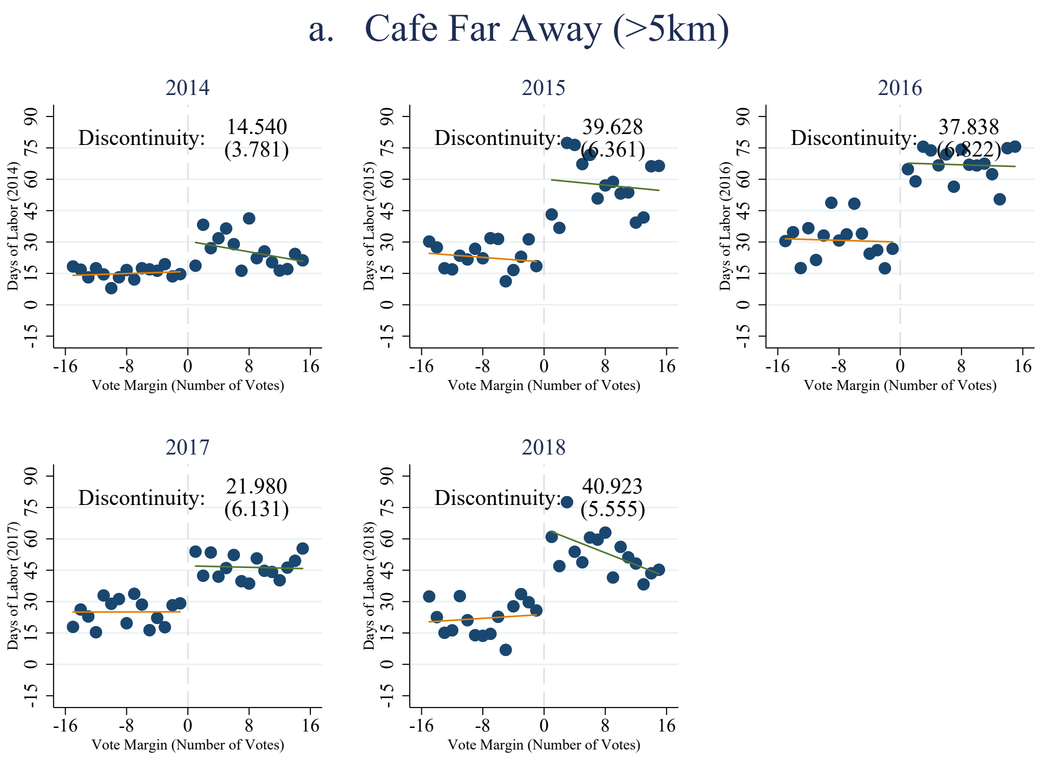
The results in this sub-section provide additional results or robustness checks for the results presented in the main paper.

Table C.1: Robustness of Internet Cafe Results

	(1)	(2)	(3)	(4)	(5)
			Days of Labor for President		
Near Cafe	-9.747**	-8.795**	-7.833*	-8.433*	-11.532**
	(4.203)	(4.270)	(4.312)	(4.401)	(4.657)
Avg. Days	1.199***	1.175***	1.188***	1.192***	1.457***
	(0.110)	(0.111)	(0.112)	(0.110)	(0.139)
2011 Night Lights		-0.672	-0.539	-0.540	-1.102*
		(0.409)	(0.429)	(0.467)	(0.642)
Distance to Border			-0.089	-0.099	-0.324
			(0.092)	(0.094)	(0.418)
Distance to Dehradun			0.104***	0.110***	0.278
			(0.039)	(0.041)	(0.377)
Distance to Sub-Dist HQ			0.108	0.103	0.199
			(0.097)	(0.097)	(0.129)
Distance to District HQ			-0.013	-0.021	-0.187*
			(0.050)	(0.052)	(0.109)
Literacy				5.642	11.268
				(29.014)	(35.173)
SCT Fraction				-12.179	-4.814
				(10.405)	(10.058)
Constant	25.218***	27.722***	17.598**	17.260	11.147
	(2.933)	(3.343)	(8.911)	(22.129)	(51.200)
Observations	340	340	340	340	340
Block FEs					X

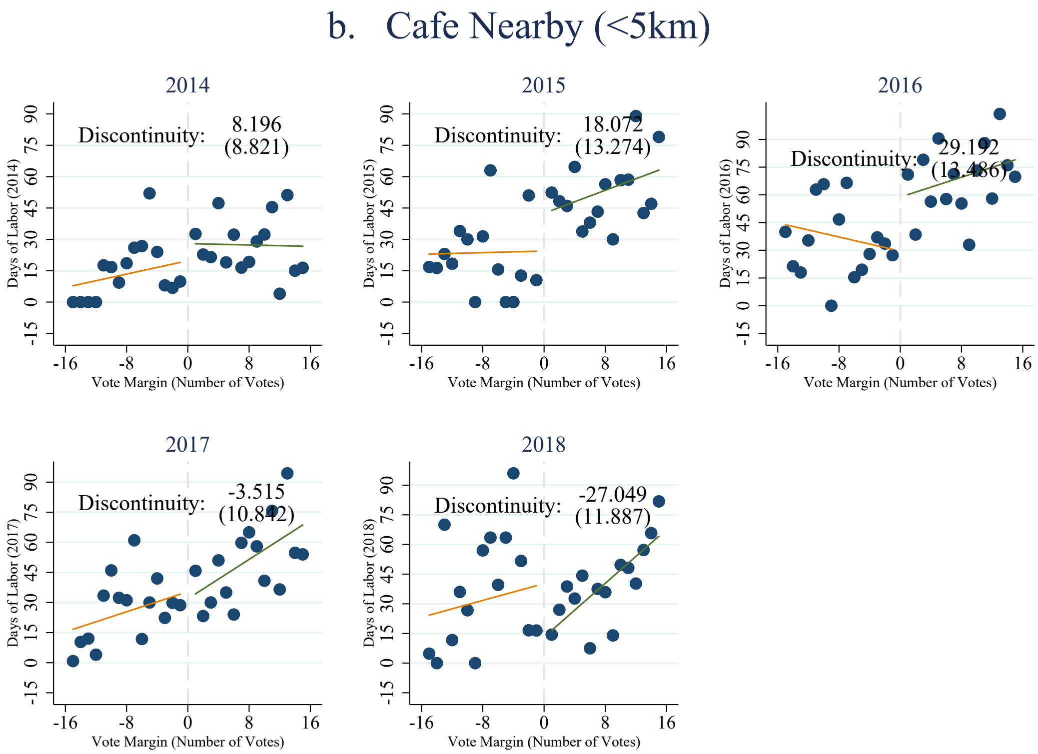
Note: Cafe refers to the nearest internet cafe according to Census data. Standard errors are robust to heteroskedasticity.
*p=0.10 **p=0.05 ***p=0.01

Figure C.1: RD Results for Long Internet Cafe Distance



Note: Figure shows the RD discontinuity graphs for each calendar year starting with the election year (2014). Standard errors are clustered by panchayat.

Figure C.2: RD Results for Short Internet Cafe Distance



Note: Figure shows the RD discontinuity graphs for each calendar year starting with the election year (2014). Standard errors are clustered by panchayat.

Table C.2: Dynamic Effects by Cafe Distance

	2012	2013	2014	2015	2016	2017	2018
Non-Candidate Average Days							
Main Effect	0.150	-0.104	0.529***	0.424***	0.320**	0.363**	0.522***
	(0.136)	(0.154)	(0.156)	(0.157)	(0.149)	(0.154)	(0.140)
× Near Cafe	0.012	0.326	-0.031	-0.182	0.106	0.099	-0.325**
	(0.322)	(0.259)	(0.204)	(0.186)	(0.140)	(0.183)	(0.158)
Observations	2863						
Panchayats	409						
Fixed-Effects	Block-Year						

Note: All coefficients in the table come from a single regression. Standard errors are clustered within panchayat.

Though Figure 3.7 suggests there is no link between self-dealing and the number of jobs generated for the village, it is possible that the average estimates hide a correlation like that found in Table 3.3. Table C.2 jointly estimates the correlation in each year and for villages that are close versus far from a cyber cafe. The upper half of the table shows the estimated coefficients on self-dealing from interacting each year between 2012 and 2018 with the average NREGS days created in the village. The lower half of the table estimates the triple interaction effect of calendar year and village-level NREGS benefits with an indicator variable for a distance of less than 5 kilometers to the nearest internet cafe.

If access to the internet allows voters to better observe both self-dealing and performance, we would expect the post-election coefficients (2014 to 2018) for the triple-interaction effects to be positive and significant, whereas the main effects should be small or close to zero. Instead, the results show the opposite. With one exception, the interaction effects of the post-election years (2014 to 2018) with closeness to a cafe are insignficant. 3 out of 5 coefficients, including the only statistically significant estimate in 2018, are also negative instead of positive. Instead, the main effects are typically large compared to the interaction effects and highly statistically significant. The overall pattern is therefore not consistent with internet access allowing voters to better observe both self-dealing and performance.

C.2 Extensions of the Analysis

C.2.1 Are Presidents Creating Phantom Projects?

Potentially the easiest way of self-dealing is to create phantom projects. For example, the president might assign himself to projects with very few other workers, or even to projects where he is the only worker. Such tricks would lower the probability that voters can detect self-dealing, at least at the physical worksites, since villagers would not be able to directly observe who gets jobs on these projects. If presidents are using this maneuver, we would expect the average number of workers on projects worked on by the president to be significantly lower than those worked on by the runner-up candidates in the election. We test this below using the NREGS project as the unit of analysis. Among the set of projects worked on by either winner or runner-up (but not both) we assign the running variable of the candidate. The results show that there is no significant discontinuity in the number of workers or in the fraction that are single-person projects (which in any case are very rare).

Table C.3: Presidents Do Not Work on Phantom Projects

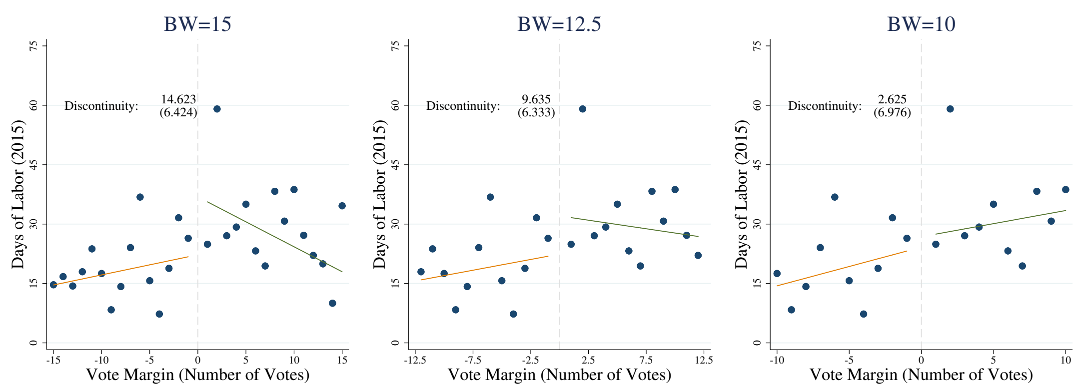
	# Workers	Solo Project
RD Estimate	3.637	-0.003
	(3.021)	(0.003)
Outcome at Disc.	19.900	0.003
Observations	1404	1404
Panchayats	607	607

Note: "Outcome at Disc." gives the estimate of the counterfactual outcome at the cutoff in the absence of treatment (that is, the left limit at the cutoff). Standard errors are clustered by panchayat. *p=0.10 **p=0.05 ***p=0.01

C.2.2 Greed or Patronage?

Since political factions in India often coalesce around powerful families (George and Ponattu, 2019), we test whether presidents reward their extended family. For each president's household, we know the name of the household head, who is almost always a man. The household head could be a male president himself, a female president's husband, the president's father, or the president's father-in-law. For all other households, we know the closest male relative of the household head, which we use as a proxy for extended family. We define a household as extended family to the president's household if it lists the president's household head as closest male relative. We assign these family members the vote margin of their contesting relative (excluding cases where the winner and runner-up are part of the same extended family).

Figure C.3: Extended Family Does Not Get Any Extra NREGS Labor



Note: Standard errors are clustered by panchayat.

Table C.4: Regression Specifications Shown in Figure C.3

	(1) BW=15	(2) BW=12.5	(3) BW=10
RD Estimate:			
–Candidate	35.607***	35.790***	33.999***
	(4.662)	(5.382)	(5.729)
–Family	14.623**	9.635	2.625
	(6.423)	(6.332)	(6.974)
Outcome at Disc.	22.257	22.430	24.180
Observations	2422	1914	1521
Panchayats	725	595	494

Note: "Outcome at Disc." gives the estimate of the counterfactual outcome at the cutoff in the absence of treatment (that is, the left limit at the cutoff). Standard errors are clustered by panchayat.
*p=0.10 **p=0.05 ***p=0.01

Figure C.3 is drawn analogously to Figure 3.5, but showing NREGS days allocated to extended family. These estimates are more sensitive to the choice of bandwidth than our estimates from Figure 3.5, so we present the same regression for three different choices of bandwidth. Though at the widest bandwidth (left panel) the estimate is positive, it is clearly an artifact of a bandwidth that is too wide. The estimate shrinks to insignificance at narrower choices of bandwidth (center and right panel), and the magnitude of the estimated discontinuity shrinks to almost zero.

Table C.4 shows the regression estimates of Figure C.3—estimates of excess labor for the extended family—alongside the estimates for the household of the candidates themselves (analogous to the estimates in Panel C of Table 3.2). We estimate both discontinuities simultaneously to correct for correlation in the coefficients. The estimates confirm that excess payments to family members shrink to insignificance as we shrink the bandwidth while those for the candidate remain unchanged. That suggests it is only the council president who receives extra NREGS labor, not her extended family.

Another form of patronage is to buy the complicity of officials who could otherwise check the president's power. Since the village council is in principle the most likely check, we asked each president in our survey to name the three most senior members of the council to test whether these members are disproportionately likely to receive large NREGS transfers. But we find no evidence that the council members are more likely to appear among the biggest NREGS recipients than would be expected by chance. In summary, there is no evidence that presidents reward supporters or form conspiracies with other politicians to self-deal NREGS jobs.

C.2.3 Does Reservation Affect the Size of Outside Payments?

Some prior work has proposed that between-group conflict can allow rent-seeking leaders to remain in power because their group fears that removing them will allow the other group to take power (Padró i Miquel, 2007). Conversely, some studies have found that reducing between-group conflict through caste reservation can induce better political selection (Munshi and Rosenzweig, 2008). Meanwhile, there is a body of work suggesting that female leaders in India govern better on some measures, but that the traditionally male-dominated system of politics in India effectively selects out these female leaders (Clots-Figueras, 2011; Chattopadhyay and Duflo, 2004).

Our data lets us test for whether villages selected for caste or gender reservation attract leaders who extract fewer excess days of NREGS labor. Table C.5 shows that although this interaction term is negative for both forms of reservation, it is small and statistically insignificant. Self-dealing is 35 days under Female Reservation versus 39 days in panchayats not reserved for women (Column 1), and 36 days under Caste Reservation versus 37.5 days in other panchayats (Column 2).

Table C.5: Reservations and Outside Payments

	(1) Female Reservation	(2) Caste Reservation
RD Estimate	39.298***	37.412***
	(6.320)	(4.910)
RD Estimate (Interaction)	-4.342	-1.580
	(8.883)	(11.230)
Outcome at Disc.	19.726	20.010
Observations	1105	1105
Panchayats	757	757

Note: "Outcome at Disc." gives the estimate of the counterfactual outcome at the cutoff in the absence of treatment (that is, the left limit at the cutoff). Standard errors are clustered by panchayat.
*p=0.10 **p=0.05 ***p=0.01

References

Abebe, G., S. Caria, M. Fafchamps, P. Falco, S. Franklin, and S. Quinn (2017). Job fair: Matching firms and workers in a field experiment in ethiopia. CSAE Working Paper.

Abebe, G., S. Caria, M. Fafchamps, P. Falco, S. Franklin, and S. Quinn (2019). Anonymity or distance? job search and labour market exclusion in a growing african city. *Review of Economic Studies*. Forthcoming.

Abel, M., R. Burger, and P. Piraino (2016). The value of reference letters-experimental evidence from south africa. *Harvard University. Processed*.

Adukia, A., S. Asher, and P. Novosad (2020). Educational investment responses to economic opportunity: evidence from indian road construction. *American Economic Journal: Applied Economics 12*(1), 348–76.

Afridi, F. and V. Iversen (2014). Social audits and mgnrega delivery: Lessons from andhra pradesh. *India Policy Forum 10*(1), 297–341.

Aggarwal, S. (2018). Do rural roads create pathways out of poverty? evidence from india. *Journal of Development Economics 133*, 375–395.

Aggarwal, S. (2019). The long road to health: Healthcare utilization impacts of a road pavement policy in rural india. Working Paper.

Aggarwal, S., E. Francis, and J. Robinson (2018). Grain today, gain tomorrow: Evidence from a storage experiment with savings clubs in kenya. *Journal of Development Economics 134*, 1–15.

Aggarwal, S., B. Giera, D. Jeong, J. Robinson, and A. Spearot (2019). Market access, trade costs, and technology adoption: Evidence from northern tanzania. *Working Paper*.

Aidt, T. S. (2003). Economic analysis of corruption: A survey. *Economic Jornal 113*, 632–652.

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

Aker, J. C. and M. Fafchamps (2015). Mobile phone coverage and producer markets: Evidence from west africa. *The World Bank Economic Review 29*(2), 262–292.

Albertus, M. (2019). The fate of former authoritarian elites under democracy. *Journal of Conflict Resolution 63*(3), 727–759.

Alder, S. (2019). Chinese roads in india: The effect of transport infrastructure on economic development. Working Paper.

Allen, T. (2014). Information frictions in trade. *Econometrica 82*(6), 2041–2083.

Allen, T. and D. Atkin (2016). Volatility and the gains from trade. *NBER Working Paper*.

Asher, S. and P. Novosad (2018). Rent-seeking and criminal politicians: Evidence from mining booms. Working Paper.

Asher, S. and P. Novosad (2020, March). Rural roads and local economic development. *American Economic Review 110*(3), 797–823.

Asturias, J., M. García-Santana, and R. Ramos (2019). Competition and the welfare gains from transportation infrastructure: Evidence from the golden quadrilateral of india. *Journal of the European Economic Association 17*(6), 1881–1940.

Atkin, D. and D. Donaldson (2015). Who's getting globalized? the size and implications of intra-national trade costs. Technical report, National Bureau of Economic Research.

Autor, D. H. (2001). Wiring the labor market. *The Journal of Economic Perspectives 15*(1), 25–40.

Avis, E., C. Ferraz, and F. Finan (2018). Do government audits reduce corruption? estimating the impacts of exposing corrupt politicians. *Journal of Political Economy 126*(5), 1912–1964.

Bagger, J. and R. Lentz (2018). An empirical model of wage dispersion with sorting. *The Review of Economic Studies 86*(1), 153–190.

Bardhan, P. (1997). Corruption and development: A review of issues. *Journal of Economic Literature 35*(3), 1320–1346.

Bassi, V. and A. Nansamba (2018). Screening and signaling non-cognitive skills: Experimental evidence from uganda. *USC-INET Research Paper* (18-19).

Baturo, A. (2017). Democracy, development, and career trajectories of former political leaders. *Comparative Political Studies 50*(8), 1023–1054.

Baye, M. R., J. Morgan, P. Scholten, et al. (2006). Information, search, and price dispersion. *Handbook on economics and information systems 1*, 323–375.

Beam, E. A. (2016). Do job fairs matter? experimental evidence on the impact of job-fair attendance. *Journal of Development Economics 120*, 32–40.

Beaman, L., D. Karlan, B. Thuysbaert, and C. Udry (2013). Profitability of fertilizer: Experimental evidence from female rice farmers in mali. *American Economic Review 103*(3), 381–86.

Benjamin, D. (1992). Household composition, labor markets, and labor demand: Testing for separation in agricultural household models. *Econometrica: Journal of the Econometric Society*, 287–322.

Berg, E., S. Bhattacharyya, R. Durgam, and M. Ramachandra (2012). Can rural public works affect agricultural wages? evidence from india.

Bergquist, L. F. and M. Dinerstein (2019). Competition and entry in agricultural markets: Experimental evidence from kenya. Working Paper.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, 841–890.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, 242–262.

Bertrand, M., S. Djankov, R. Hanna, and S. Mullainathan (2007). Obtaining a driver's license in india: An experimental approach to studying corruption. *The Quarterly Journal of Economics 122*(4), 1639–1676.

Bird, J. and S. Straub (2014). *The Brasilia experiment: road access and the spatial pattern of long-term local development in Brazil.* The World Bank.

Breza, E., S. Kaur, and N. Krishnaswamy (2019). Scabs: The social suppression of labor supply. *NBER Working Paper*.

Brooks, W., K. Donovan, et al. (2019). Eliminating uncertainty in market access: Evidence from new bridges in rural nicaragua. *Working Paper*.

Brown, M. B. and A. B. Forsythe (1974). Robust tests for the equality of variances. *Journal of the American Statistical Association 69*(346), 364–367.

Bryan, G., S. Chowdhury, and A. M. Mobarak (2014). Underinvestment in a profitable technology: The case of seasonal migration in bangladesh. *Econometrica 82*(5), 1671–1748.

Bryan, G. and M. Morten (2019). The aggregate productivity effects of internal migration: Evidence from indonesia. *Journal of Political Economy 127*(5), 2229–2268.

Burdett, K. and K. L. Judd (1983). Equilibrium price dispersion. *Econometrica: Journal of the Econometric Society*, 955–969.

Burdett, K. and D. T. Mortensen (1998). Wage differentials, employer size, and unemployment. *International Economic Review*, 257–273.

Burke, M., L. F. Bergquist, and E. Miguel (2019). Sell low and buy high: arbitrage and local price effects in kenyan markets. *The Quarterly Journal of Economics 134*(2), 785–842.

Butters, G. R. (1977, 10). Equilibrium Distributions of Sales and Advertising Prices. *The Review of Economic Studies 44*(3), 465–491.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica 82*(6), 2295–2326.

Campante, F. R. and Q.-A. Do (2014, AUG). Isolated Capital Cities, Accountability, and Corruption: Evidence from US States. *American Economic Review 104*(8), 2456–2481.

Casaburi, L., R. Glennerster, and T. Suri (2013). Rural roads and intermediated trade: Regression discontinuity evidence from sierra leone. *Available at SSRN 2161643*.

Casaburi, L. and T. Reed (2019). Interlinked transactions and competition: Experimental evidence from cocoa markets.

Chatterjee, S. (2019). Market power and spatial competition in rural india.

Chattopadhyay, R. and E. Duflo (2004). Women as policy makers: Evidence from a randomized policy experiment in india. *Econometrica 72*(5), 1409–1043.

Clots-Figueras, I. (2011). Women in politics: Evidence from the indian states. *Journal of Public Economics 95*, 664–690.

Combes, P.-P., G. Duranton, L. Gobillon, D. Puga, and S. Roux (2012). The productivity advantages of large cities: Distinguishing agglomeration from firm selection. *Econometrica 80*(6), 2543–2594.

Dey, N., J. Dreze, and R. Khera (2006). *Employment Guarantee Act: A Primer*. Delhi: National Book Trust, India.

Dhingra, S. and S. Tenreyro (2017). Piggy-back exporting, intermediation, and the distributional gains from trade in agricultural markets.

Di Tella, R. and E. Schargrodsky (2003). The role of wages and auditing during a crackdown on corruption in the city of buenos aires. *The Journal of Law and Economics 46*(1), 269–292.

Diermeier, D., M. Keane, and A. Merlo (2005). A political economy model of congressional careers. *American Economic Review 95*(1), 347—373.

Dillon, B., P. Brummund, and G. Mwabu (2019). Asymmetric non-separation and rural labor markets. *Journal of Development Economics 139*, 78–96.

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

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

Donaldson, D. and R. Hornbeck (2016). Railroads and american economic growth: A "market access" approach. *The Quarterly Journal of Economics 131*(2), 799–858.

Duflo, E., M. Kremer, and J. Robinson (2008). How high are rates of return to fertilizer? evidence from field experiments in kenya. *The American economic review (Papers and Proceedings Issue) 98*(2), 482–488.

Dutta, P., R. Murgai, M. Ravallion, and D. van de Walle (2012). Does india's employment guarantee scheme guarantee employment? World Bank Policy Research Working Paper 6003, http://documents.worldbank.org/curated/en/464991468051015810/pdf/WPS6003.pdf (accessed August 3, 2019).

Eaton, J. and S. Kortum (2002). Technology, geography, and trade. *Econometrica 70*(5), 1741–1779.

Eckert, F. (2019). Growing apart: Tradable services and the fragmentation of the us economy. Working Paper, Yale University.

Eggers, A. C. and J. Hainmueller (2009). Mps for sale? returns to office in postwar british politics. *American Political Science Review 103*(4), 513—533.

Engel, C. and J. Rogers (1996). How wide is the border. *American Economic Review 86*(1112), 25.

FAO (2016). Faostat. Food and Agriculture Organization of the United Nations, Rome, Italy.

Ferraz, C. and F. Finan (2008). Exposing corrupt politicians: The effects of brazil's publicly released audits on electoral outcomes. *The Quarterly Journal of Economics 123*(2), 703–745.

Fink, G., B. K. Jack, and F. Masiye (2018). Seasonal liquidity, rural labor markets and agricultural production. *NBER Working Paper*.

Fisman, R., F. Schulz, and V. Vig (2014). The private returns to public office. *Journal of Political Economy 122*(4), 806—-862.

Fisman, R. and J. Svensson (2007). Are corruption and taxation really harmful to growth? firm level evidence. *Journal of Development Economics 83*, 63–75.

Foltz, J. D. and K. A. Opoku-Agyemang (2015). Do higher salaries lower corruption? a policy experiment on west africa's highways. Working Paper.

Foster, A. D. and M. R. Rosenzweig (2017). Are there too many farms in the world? labor-market transaction costs, machine capacities and optimal farm size. *Working Paper*.

Foster, V. and C. M. Briceño-Garmendia (2009). *Africa's infrastructure: a time for transformation*. The World Bank.

Franklin, S. (2017). Location, search costs and youth unemployment: experimental evidence from transport subsidies. *The Economic Journal 128*(614), 2353–2379.

George, S. E. and D. Ponattu (2019). Like father, like son? the effect of political dynasties on economic development. Working Paper, https://scholar.harvard.edu/files/siddharthgeorge/files/sid_dynasties_draft_29jan2019.pdf (accessed February 16, 2019).

Gertler, P., T. Gracner, M. Gonzalez-Navarro, and A. Rothenberg (2019). Road quality and local economic activity: evidence from indonesia's highways. *Work. Pap., Univ. Calif., Berkeley*.

Ghani, E., A. G. Goswami, and W. R. Kerr (2016). Highway to success: The impact of the golden quadrilateral project for the location and performance of indian manufacturing. *The Economic Journal 126*(591), 317–357.

Gollin, D. and C. Udry (2019). Heterogeneity, measurement error and misallocation: Evidence from african agriculture. *NBER Working Paper* (w25440).

Government of India (2013). Mahatma gandhi national rural employment guarantee act, 2005 operational guidelines. Technical report, Ministry of Rural Development. https://nrega.nic.in/Circular_Archive/archive/Operational_guidelines_4thEdition_eng_2013.pdf (accessed July 8, 2018).

Government of India (2018). The mahatma gandhi national rural employment guarantee act, 2005. Technical report, Government of India. http://nrega.nic.in/amendments_2005_2018.pdf (accessed August 3, 2019).

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

Groh, M., D. McKenzie, N. Shammout, and T. Vishwanath (2015). Testing the importance of search frictions and matching through a randomized experiment in jordan. *IZA Journal of Labor Economics 4*(1), 7.

Hausman, J. A. (1996). Valuation of new goods under perfect and imperfect competition. In *The economics of new goods*, pp. 207–248. University of Chicago Press.

Head, K. and T. Mayer (2011). Gravity, market potential and economic development. *Journal of Economic Geography 11*(2), 281–294.

Hernandez, M. A., M. Torero, et al. (2011). *Fertilizer market situation: market structure, consumption and trade patterns, and pricing behavior*. Internat. Food Policy Research Inst.

Hornstein, A., P. Krusell, and G. L. Violante (2011). Frictional wage dispersion in search models: A quantitative assessment. *American Economic Review 101*(7), 2873–98.

Huntington, S. P. (1968). *Politicial Order in Changing Societies*. New Haven: Yale University Press.

Imbert, C. and J. Papp (2015). Labor market effects of social programs: Evidence of india's employment guarantee. *American Economic Journal: Applied Economics 7*(2), 233–263.

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

Kaur, S. (2019). Nominal wage rigidity in village labor markets. *American Economic Review*.

Khanna, G. (2019). Road oft taken: The route to spatial development. *Available at SSRN 2426835*.

Klašnja, M. (2015). Corruption and the incumbency disadvantage: Theory and evidence. *Journal of Politics 77*(4), 928–942.

LaFave, D. and D. Thomas (2016). Farms, families, and markets: New evidence on completeness of markets in agricultural settings. *Econometrica 84*(5), 1917–1960.

Leff, N. H. (1964). Economic development through bureaucratic corruption. *American Behavioral Scientist 8*(3), 8–14.

Levene, H. (1960). Robust tests for equality of variances in contribution to probability and statistics. *Olkin: Stanford University Press, Palo Alto*.

Lui, F. T. (1985). An equilibrium queuing model of bribery. *Journal of Political Economy 93*(4), 760–781.

MacMinn, R. D. (1980). Search and market equilibrium. *Journal of Political Economy 88*(2), 308–327.

McKenzie, D. (2012). Beyond baseline and follow-up: The case for more t in experiments. *Journal of development Economics 99*(2), 210–221.

McKenzie, D. (2017). How effective are active labor market policies in developing countries? a critical review of recent evidence. *The World Bank Research Observer*.

Melitz, M. J. and G. I. Ottaviano (2008). Market size, trade, and productivity. *The review of economic studies 75*(1), 295–316.

Méon, P.-G. and K. Sekkat (2005). Does corruption grease or sand the wheels of growth? *Public Choice 122*(1/2), 69–97.

Méon, P.-G. and L. Weill (2010). Is corruption an efficient grease? *World Development 38*(3), 244–259.

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

Mironov, M. and E. Zhuravskaya (2016, MAY). Corruption in Procurement and the Political Cycle in Tunneling: Evidence from Financial Transactions Data. *American Economic Journal: Economic Policy 8*(2), 287–321.

Morten, M. and J. Oliveira (2016). The effects of roads on trade and migration: Evidence from a planned capital city. *NBER Working Paper 22158*, 43.

Mortensen, D. (1988). *Equilibrium wage distributions: a synthesis*. Center for Mathematical Studies in Economics and Management Science, Northwestern University.

Mukhopadhyay, A., Himanshu, and M. R. Sharan (2015). The national rural employment guarantee scheme in rajasthan: Rational funds and their allocation across villages. *Economic and Political Weekly 50*(6), 1409—-1443.

Munshi, K. and M. Rosenzweig (2008). The efficacy of parochial politics: caste, commitment, and competence in indian local governments. NBER Working Paper 14335, https://www.nber.org/papers/w14335 (accessed July 20, 2019).

Muralidharan, K., P. Niehaus, and S. Sukhtankar (2016). Building state capacity: Evidence from biometric smartcards in india. *American Economic Review 106*(10), 2895–2929.

Niehaus, P. and S. Sukhtankar (2013a). Corruption Dynamics: The Golden Goose Effect. *American Economic Journal: Economic Policy 5*(4), 230–269.

Niehaus, P. and S. Sukhtankar (2013b). The Marginal Rate of Corruption in Public Programs. *Journal of Public Economics 104*, 52–64.

Olken, B. A. (2007). Monitoring corruption: Evidence from a field experiment in indonesia. *Journal of Political Economy 115*(2), 200–249.

Olken, B. A. and R. Pande (2012). Corruption in developing countries. *Annual Review of Economics 4*(1), 479—509.

Ortego-Marti, V. (2016). Unemployment history and frictional wage dispersion. *Journal of Monetary Economics 78*, 5–22.

Padró i Miquel, G. (2007). The control of politicians in divided societies: The politics of fear. *The Review of Economic Studies 74*(4), 1259–1274.

Pissarides, C. A. (2000). *Equilibrium unemployment theory.* MIT press.

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), 327–66.

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

Postel-Vinay, F. and J.-M. Robin (2002). Equilibrium wage dispersion with worker and employer heterogeneity. *Econometrica 70*(6), 2295–2350.

Querubín, P. and J. M. J. Snyder (2013). The control of politicians in normal times and times of crisis: Wealth accumulation by u.s. congressmen, 1850-1880. *Quarterly Journal of Political Science 8*, 409–450.

Ravallion, M. (2018). Guaranteed employment or guaranteed income? Center for Global Development Working Paper 482.

Redding, S. and A. J. Venables (2004). Economic geography and international inequality. *Journal of international Economics 62*(1), 53–82.

Redding, S. J. and D. M. Sturm (2008). The costs of remoteness: Evidence from german division and reunification. *American Economic Review 98*(5), 1766–97.

Reinikka, R. and J. Svensson (2004). Local capture: evidence from a central government transfer program in uganda. *The quarterly journal of economics 119*(2), 679–705.

Shamdasani, Y. (2016). Rural road infrastructure and agricultural production: Evidence from india. *Department of Economics, Columbia University, USA*.

Shamdasani, Y. (2020). Rural road infrastructure and agricultural production: Evidence from india. *Working Paper*.

Shiue, C. H. and W. Keller (2007). Markets in china and europe on the eve of the industrial revolution. *American Economic Review 97*(4), 1189–1216.

Sotelo, S. (2019). Domestic trade frictions and agriculture. *Journal of Political Economy*. Forthcoming.

Stewart, W., D. Dibb, A. Johnston, and T. Smyth (2005). The contribution of commercial fertilizer nutrients to food production. *Agronomy Journal 97*(1), 1–6.

Stigler, G. J. (1961). The economics of information. *Journal of political economy 69*(3), 213–225.

Storeygard, A. (2016). Farther on down the road: transport costs, trade and urban growth in sub-saharan africa. *The Review of economic studies 83*(3), 1263–1295.

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

Svensson, J. (2005). Eight questions about corruption. *Journal of Economic Perspectives 19*(3), 19–42.

Udry, C. and S. Anagol (2006). The return to capital in ghana. *American Economic Review 96*(2), 388–393.

Van den Berg, G. J. and A. Van Vuuren (2010). The effect of search frictions on wages. *Labour Economics 17*(6), 875–885.

Weaver, J. (2018). Jobs for sale: Corruption and misallocation in hiring. Working Paper.

Wei, S.-J. (2000). How taxing is corruption on international investors? *Review of Economics and Statistics 82*(1), 1–11.

World Bank Group (2007). World development report 2008 : Agriculture for development. World Bank, Washington, DC. https://openknowledge.worldbank.org/handle/10986/5990 License: CC BY 3.0 IGO.

World Bank Group (2017). Enabling the business of agriculture 2017. World Bank, Washington, DC. https://openknowledge.worldbank.org/handle/10986/25966 License: CC BY 3.0 IGO.

Zimmermann, L. (2018). Why guarantee employment? evidence from a large indian public-works program. Working Paper, https://drive.google.com/open?id=12JYprWjRRIkGW59h0zRdJiiCbSX8kEBl (accessed July 10, 2019).