UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title
Essays on the Role of Information in Development Economics, Trade and Political Economy

Permalink
https://escholarship.org/uc/item/12s8d5b6

Author
Wiseman, Eleanor

Publication Date
2023

Peer reviewed|Thesis/dissertation

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

Essays on the Role of Information in Development Economics, Trade and Political Economy

By

Eleanor Wiseman

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Agricultural and Resource Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Jeremy Magruder, Chair
Professor Elisabeth Sadoulet
Professor Edward Miguel

Summer 2023

Essays on the Role of Information in Development Economics, Trade and Political Economy

Copyright 2023
by
Eleanor Wiseman

Abstract

Essays on the Role of Information in Development Economics, Trade and Political Economy

by

Eleanor Wiseman

Doctor of Philosophy in Agricultural and Resource Economics

University of California, Berkeley

Professor Jeremy Magruder, Chair

This dissertation examines how information frictions play a role in fostering poverty, limiting firm growth and perpetuating weak institutions in low income settings. In contexts where markets and formal institutions are weak or not functioning optimally, informal economic systems play an important role. Networks and relationships are central to the existence of informal systems; therefore, exploring existing information constraints, understanding information diffusion and alleviating information frictions has the potential to significantly improve the livelihoods of these populations.

My dissertation looks at how formal and informal systems co-exist and more specifically whether reducing information frictions can improve welfare, taking both formal and informal channels into account. I study the interactions between formal and informal institutions and focus on information frictions in a variety of contexts, specifically in (1) trade and value chains in Kenya and Uganda (Chapter 1), (2) agriculture markets and technology adoption in India (Chapter 2) and (3) voting behaviors and government failures in low income populations of the USA (Chapter 3).

In Chapter 1, I focus on informal trade. In low- and middle-income countries, a large share of trade is conducted by small-scale informal traders – mostly women – and is missing from official trade statistics. Using the natural experiment of a border closing, a randomized controlled trial, and panel data collection, I study the role of information frictions in traders' choices of markets and border crossings at the Kenyan-Ugandan border and the consequences for livelihoods and prices in agriculture markets. First, I show that traders' choice of markets and routes is sticky. Second, some of this stickiness is driven by limited information about profitable arbitrage opportunities and true (tariff) costs of crossing the border. Third, I

build a model incorporating these frictions, which I test using an RCT. I find that giving information on tariff costs and local prices to traders (via a cellphone platform) increases switching across markets and routes, leading to large increases in traders' profits and significant formalization of trade. Consistent with the model, information provision has general equilibrium effects – specifically, a reduction in consumer prices in agricultural markets. Taken together, the results point to the centrality of information frictions in informal trade and highlight the promise of new information technology to ameliorate them.

In Chapter 2, I look at the role of information frictions in limiting adoption of new agricultural technologies in developing countries. Efforts to improve learning involve spreading information from government agents to farmers. We show that when compared to this government approach, informing *private input suppliers* in India about a new seed variety increases farmer-level adoption by over 50 percent. Suppliers become more proactive in informing potential customers and carrying the new variety. They induce increased adoption by those with higher returns from the technology. Being motivated by expanded sales offers the most likely motive for these results.

In Chapter 3, I explore how information affects voting behavior. We study changes in voting, new voter registration, and candidate choice in response to a criminal government failure which exposed one in twelve households in Flint, MI to lead in their tap water. We compare changes in outcomes for voters who received home lead test results just before versus just after an election. We find that the Flint water crisis, widely understood as the result of institutional racism, caused stark racial divergence in political participation between Black and White voters. Black voters increased turnout, accelerated registration, and rejected the incumbent, while White voters did not change their voting behavior.

To my family, who has always supported me through the highs and lows and to my grandmother who knew more than anyone the power of information.

Contents

List of Figures

List of Tables

Acknowledgments

As I was considering starting a PhD, I was warned that doing a PhD would present its set of challenges. There certainly were stressful moments but when I look back at these past few years, I do not regret at all my decision to embark on this journey. It has overall been a fantastic experience. I have learned and developed so much personally and professionally along the way and leave U.C. Berkeley with wonderful memories. I feel very grateful to have had the opportunity to be surrounded by talented, caring and passionate friends, colleagues and advisors who have encouraged me to care about the world, taught me the importance of good research and emphasized the value of putting our minds together.

I am very grateful to my advisors - Jeremy Magruder, Elisabeth Sadoulet, Alain de Janvry, Edward Miguel and Tavneet Suri. I would not be where I am without your invaluable support and help throughout all these years. You encouraged me to pursue my (sometimes crazy) ideas, were present when I needed it and I'll never forget your support and kindness when things got hard due to health issues. I cannot thank you enough for everything you have done for me.

I am also thankful for the wonderful people, colleagues and friends I was lucky to be surrounded by during my PhD:

My cohort who made everything better.

Carmen Karahalios and Diana Lazo, for your help throughout and your dedication to the department and students.

Co-authors of my dissertation: Elisabeth Sadoulet, Alain de Janvry, Kyle Emerick, Manzoor H. Dar (Chapter 2) and Kate Pennington (Chapter 3) for showing me the ins and outs of doing research.

Ben Wekesa, Peter LeFrancois and my survey team at Innovation for Poverty Action for your dedication to the project and incredible work, despite the challenges of the pandemic.

Sauti East Africa for your trust throughout the project.

J-Pal (specifically ATAI and the Government Initiative) and CEGA for funding my projects.

Teammates and the BBC/Anarchy crew for keeping me sane, making me laugh and pushing my limits physically.

Doctors Bellemans, Mourad and Jamar for saving my life.

Lastly, I am grateful to my family and long time friends who are always there for me, through the highs and the lows, even across continents. Thank you for believing in me and for your unconditional love.

Chapter 1

Border Trade and Information Frictions: Evidence from Informal Traders in Kenya

1.1 Introduction

In low- and middle-income countries, informal trade accounts for a substantial fraction of total trade. Such trade is dominated by women, with little education and often from rural communities. In the Southern African Development Community, informal cross-border trade accounts for an estimated USD 17.6 billion annually, or 30 to 40 % of total regional trade (Southern-Africa-Trust (2008)). Despite its importance, there is limited data regarding informal trade. An obvious reason is that informal trade flows are not recorded in official trade statistics. Moreover, because informal traders are frequently on the move, collecting reliable data is not straightforward. The sparse data currently available comes from surveyors stationed at border crossings, who estimate the volume of trade that passes. Finally, given that a significant amount of informal trade goes through unofficial crossings, some of the costs take the form of bribes; and data collection on corruption presents its own unique challenges (Hadley and Rowlatt (2019)).

Due to the scarcity of data on small-scale traders and the complexity of the environment, just a handful of studies provide empirical (descriptive) insights on this sector, e.g., Titeca and Kimanuka (2012), Hadley et al. (2018). In line with typical frameworks, traders maximize their profits by locating low-priced markets for buying and high-priced markets for selling, taking into account transportation, border, and other trade costs. Little is known, however, about the costs incurred by small-scale traders, who likely face a different cost structure than larger-scale traders. There are a number of plausible scale economies in trade, including in transport costs, tariff rates, corruption incidence, and — crucially for this paper — information acquisition. These traders tend to operate outside traditional business circles, complicating access to market information as well as information about policies or regulations related to crossing the border. In addition to the costs incurred to get accurate information, the lack of information motivates official border agents to extract money from small-scale traders[1], further increasing their trade costs (Klopp et al. (2022a)).

This paper examines the significance of information frictions in informal trade and how they distort consumer prices. To answer these questions, I collect high-frequency panel data on trade outcomes for a sample of 1,100 traders in Kenyan markets close to the Kenyan-Ugandan border, throughout 2020 and 2021. These informal traders are domestic traders or cross-border traders, and either use formal border crossings or cross the border via unofficial crossings in order to avoid taxes and smuggle goods[2]. This paper therefore refers to trade routes as (i) domestic trade, (ii) international trade through official crossings, (iii) international trade through informal crossings.

I first present descriptive findings on the informal trade sector, including how different types of routes co-exist and how stickiness reduces traders' opportunities to choose optimal markets and routes – in other words, missing out on arbitrage opportunities. I use a natural experiment[3] – the closure of the official border due to Covid-19 – to derive those findings. I hypothesize that the stickiness can be explained by information frictions. Then, I develop a model that incorporates information frictions. I test the model using a randomized controlled trial in which I provide traders with access to a platform that includes information on market prices and official border costs. Through the experiment, I examine whether informational frictions restrict access to markets and trade route utilization. As predicted by the model, I demonstrate that an intervention that decreases information frictions has general equilibrium impacts on market prices. Lastly, I quantify the scale of the intervention's impacts and the role of information and informal trade, using calculations of welfare and cost-effectiveness, and counterfactual simulations of the model.

The trader level approach in this paper allows me to meaningfully contribute to the existing literature focusing on the role of information technologies (such as mobile phones) on agriculture market integration[4]. The existing literature, which mostly uses market level

[1]Small-scale traders are more likely to face coercive corruption (pay more than the official tax rates), whereas larger-scale traders enter in collusive agreements with officials to split rents.

[2]The literature and governments are moving away from defining those traders as smugglers and instead referring to informal cross-border traders. Conceptually, traders using unofficial crossings can be associated with smuggling.

[3]Despite calling it a "natural experiment", it is clear that the event does not generate perfect exogenous variation. This event is only used to generate findings that motivate the rest of the paper.

[4]I describe my contributions to the literature in more details below.

data, shows that increased access to mobile phone networks reduces price dispersion across markets. It rationalizes this (without direct evidence) through a framework in which technology increases information transmission and reduces search costs for traders; leading to a shift of trade across space and potential welfare improvements. However, I am able to directly explore the mechanisms at play. First, I directly measure who has information and the type of information received. Second, I observe trader level choice of markets and trade routes; implying I can explain market price effects through changes in traders' choice of markets but also routes (which includes internationalization). Third, having detailed data on traders' revenues and costs (in addition to markets and routes used) allows me to differentiate market effects stemming from changes in marginal costs (e.g., traders are buying from cheaper markets) and changes in marginal revenue (e.g., traders are selling in better markets). Thanks to my data, I can therefore look at how information frictions impact competition in markets and how much they contribute to the existence of market power. Lastly, my paper shows that information frictions play a role in traders' stickiness across markets and trade routes, which is novel in this literature.

Understanding informal trade and its inherent frictions is important for policy in my study area (the Kenya-Uganda border) and in sub-Saharan Africa generally. First, in addition to being a major component of trade, informal traders contribute to livelihoods and food security by connecting farmers with consumers (Ackello-Ogutu (1997); Little (2005)). They frequently serve as the main point of contact in agricultural markets: only 3% of traders in my sample who were found in Kenyan markets, are registered businesses. Second, these informal traders are small firms and their owners are often women heading low-income households, at a time when women's economic empowerment is a major development concern (Klopp et al. (2022b)). Third, informal trade results in missed opportunities for tax revenue collection. In order to boost tax revenues, governments and other organizations have made substantial investments (e.g., One Stop Border Posts[5]) and implemented policies (e.g., Simplified Trade Regime in East Africa[6]) to reduce costs and formalize trade. Fourth, due to their size and gender, these small-scale traders are disproportionately impacted by non-tariff barriers such as corruption and harassment at the border, and governments' efforts such as the implementation of One Stop Border Posts are also intended to address harassment and corruption (DMT Final report (2017); Brenton et al. (2013)).

In the first part of the paper, I use my high frequency panel data to establish descriptive facts. I document that traders face costs that account for over 80% of their sales. Expenses include purchase and transport costs as well as border costs that depend on which border

[5]Institutional framework, facility and associated procedures that enable goods, people and vehicles to stop in a single facility to undergo necessary controls to exit one State and enter the adjoining State.

[6]The Simplified Trade Regime was created to make it easier and faster for small-scale cross-border traders with products that are grown or manufactured in the EAC to clear customs.

crossing they choose[7]. Second, informal traders specialize in types of routes (domestic, formal border crossing and informal border crossing) and markets, despite the fact that large price variations between markets creates arbitrage possibilities. Nonetheless, market and route stickiness can be overcome with a sufficiently severe shock: the closing of the official border crossing. This compelled traders who used to travel through the official border to use either informal routes or domestic supply chains. I find that switching is sticky: when the border reopened a few months later, just 25% of traders returned to their original markets and routes. Lastly, traders do not operate with full information; only 38% of informal traders report knowing the market price for their primary commodity in other markets, only 51% report informing themselves about market prices, and over 50% report not always knowing the taxes they should pay at the border.

Building on these facts, I formulate a model in which risk-averse traders select markets and routes. I present a micro-foundation of market and route stickiness through information frictions, including uncertainty regarding purchasing prices, demand in selling markets, and official border fees. The model predicts that reducing information frictions will increase the number of markets traded in (Prediction 1), traders' sales and profits (Prediction 2), and cross-border trade and formalization (Prediction 3). The model finds ambiguous effects on bribes (Prediction 4), and predicts general equilibrium effects such as a decrease in consumer prices in selling markets (Prediction 5) and ambiguous effects in buying markets (Prediction 6).

I test each of these hypotheses using a randomized controlled trial designed to test the theory and find empirical evidence for the model's main predictions. A random 50% of traders in my sample received access to market information through a phone platform developed by Sauti East Africa that reports market price information, official taxes and tariffs, and exchange rates. I additionally randomize the treatment intensity[8] at the market level to account for spillovers and to evaluate price impacts in general equilibrium. The treatment increases the likelihood of trading by 5 percentage points and the likelihood that trading is their primary source of income by 4 percentage points. Treatment increases the number of types of goods traded by each trader by 8%. Turning to supply chains, the number of selling and sourcing markets increases by 7-11% for treated traders, and the number of trips increases by 0.4 trips (Prediction 1). Access to information increases treated traders' sales and profits by 6 and 7%, respectively (Prediction 2). I see no indication of treatment impacts on markups, indicating perfect passthrough. Treatment increases the number of cross-border traders (rather than using domestic routes) by 20% and leads to the formalization of trade;

[7]In another paper I characterizes border costs based on whether a formal or informal border crossing is utilized. These costs estimates are obtained as part of an "audit study" experiment in which trained traders crossed both borders while I exogenously altered important factors.

[8]I therefore have variation in the share of treated traders across markets.

i.e., treatment boosts trade through the official crossing point (Prediction 3). The model predicts ambiguous effects on bribes; I find no substantial effects of treatment on levels of bribery or instances of corruption and harassment (Prediction 4).

Beyond these direct treatment effects, there are broader market level effects that I am able to estimate. Focusing on general equilibrium results, I show large market price effects: marketplaces with a higher share of treated traders see a reduction of 7% in consumer prices (Prediction 5). The effect on market prices in buying markets is less straightforward; still, I estimate a 7% fall in buying prices in initial buying markets (Prediction 6). These significant effects on market prices demonstrate that information frictions heavily distort markets.

In the last part of the paper, I quantify the magnitude of the effects. First, relating my results to the literature on more general gravity models, it is natural to compare information friction costs to transport costs. I show that the reduction of information frictions induced by the intervention is large and equivalent to transport costs incurred by being 150 kms away from the border (when markets in my sample are all within 40 kms of the border). Second, I provide welfare calculations and cost-effectiveness estimates of information technologies such as the one used in this intervention. I find meaningful welfare gains, equivalent to USD 683 per trader per year; 88% stemming from consumer surplus (from reduced market prices) and 12% from government surplus (from increased tax revenues). At an approximate cost of USD 7 per user, this ranks the intervention as highly cost-effective. Third, through counterfactual simulations of the model, I can predict the effect of closing specific routes on prices and welfare. I first look at the effect of closing the official route. I compare the results from the model to out-of-sample reduced form estimates using the official border closure shock. The model predicts an increase in price in markets close to the border, which approximately matches the reduced form results from the border closure shock (8%). Lastly, using the model, I assess consequences of shutting down informal crossings (in addition to formal crossings) and show that informal trade does play an important role in smoothing prices, especially during border shocks. Therefore potential welfare gains from trade formalization need to be measured against the role informal crossings play in smoothing prices during these shocks.

Contributions to the literature

This paper contributes to the literature on information frictions in agricultural markets – specifically, papers looking at the use of cell phones to alleviate information frictions, e.g., Aker (2010), Jensen (2007) and Allen (2014). In line with the literature, my paper highlights the importance of information frictions in trade in developing countries. These papers use the expansion of mobile phone networks to look at how a reduction in information frictions affects market price dispersion. The research is centered around the (plausible) assumption that the expansion of phone networks led to traders collecting more market information. In contrast, my paper directly gives traders access to information through their phones, allowing me to understand the mechanisms at work, both in terms of access to information and use of information. In terms of access, I know which traders actually received information (rather than which markets have access to mobile phone networks) and therefore can directly link the reduction of information frictions to the use of mobile phone technology. I also explore mechanisms in terms of usage to understand how traders search for information and how that translates into actual trade behaviors, i.e., choice of markets and trade routes. For a more general overview of the role of mobile phone technologies in development, Aker and Mbiti (2010) provide a good summary. In addition, the models used in the literature frame information frictions as search costs (Stigler (1961)); and therefore use sequential search models (Stiglitz (1989)). In my paper, I suggest that information frictions can also be framed as uncertainty about market conditions for risk-averse traders.

My paper also speaks to the literature about information frictions in *domestic* agriculture trade. Fafchamps and Minten (2012) and Mitra et al. (2018) explore whether informing farmers about market prices can help grow their business by allowing them to sell their goods to traders in better-priced markets, rather than at the farm gate. Generally, the literature finds little evidence that access to information affects farmers or market conditions. My paper focuses on traders rather than farmers, which seems to be the right segment of the value chain to target, as transporting goods is a main feature of their business.

Third, this paper contributes to the literature on frictions in *international* or *cross-border* trade and the role of informality. More generally, the role of information frictions in market access and contracts for small firms has been studied; e.g.,Atkin and Donaldson (2015), Atkin et al. (2017), Hjort et al. (2020). Startz (2021) explores search and contracting frictions between traders and international suppliers. My paper directly adds to such research on search frictions by testing the existence of search costs – taking the form of information frictions about market prices – through a RCT. Turning to the role informality plays in international trade, the literature on informal trade is relatively small and predominantly qualitative. Using national aggregates of trade flows by product, a few papers quantitatively explore determinants of informal versus official trade flows – mostly observable costs such as tariffs (Bensassi et al. (2019)) or trade facilitation policies (Siu (2020)). My paper contributes by proposing a micro-economics approach through trader-level data and border crossing choice. I also use a representative sample of traders, collected in markets at the border, rather than relying on observational data from traders crossing the border. Beyond this official-unofficial choice of border crossing, my paper also considers the role of information frictions in the choice between domestic and international trade. Models generally assume that more productive firms are international traders, while less productive firms remain domestic traders. I provide evidence that information frictions about market prices and taxes are barriers to international trade and that relaxing those frictions allows small-scale firms to engage in international trade, while helping consumers with lower prices. To my knowledge, this is the first paper to conduct an experiment on cross-border traders that explores the interdependence between domestic and international trade, including official and unofficial crossings.

My paper considers corruption and bribes as a direct cost to international traders, and explores whether information frictions play a role in corruption and bribe levels. In line with the framework in Sequeira and Djankov (2014), the type of corruption observed at the official crossings varies by trader size[9]; however, small-scale traders face both collusive and coercive corruption. As in Croke et al. (2021), I explore whether coercive corruption can be due to lack of information about true taxes. As for collusive corruption (such as Reid and Weigel (2019)), small traders often use informal crossings to avoid taxes and coercive corruption, prevalent at official crossings. I find suggestive evidence that traders face lower taxes and bribes when they have information about official taxes and market prices. However, this small effect masks both a likely increase in border costs from increased demand for cross-border routes (compared to domestic routes), and a potential decrease in bribes from traders' improved bargaining power.

Lastly, my paper contributes to a growing literature on intermediaries in trade and the role of traders in value/supply chains. Grant and Startz (2022) and Iacovone and McKenzie (2022) highlight the role of intermediaries in a value chain. Similarly to Bergquist and Dinerstein (2020) and Startz (2021), I show that traders contribute to high selling prices in agricultural markets through market structure. My paper extends this literature by highlighting the role of information frictions in inflating those high consumer prices.

The remainder of the paper is organised as follows. In Section 1.2, I provide some background about the project location and trade in East Africa. I describe my experimental design and data in Section 1.3. In Section 1.4, I lay out key motivating stylized facts and in section 1.5 a theoretical model based on these. I present my empirical framework in section 1.6 and main findings in Section 1.7. Finally, in Section 1.8, I give welfare estimates and use the model to run counterfactual analysis, and Section 1.9 concludes.

1.2 Background and Context

1.2.1 Location

This project is located in markets around Busia, a town situated at the border between Kenya and Uganda. Busia is one of the main border crossings between the two countries. In

[9]Anecdotal evidence suggests that large-scale traders -not the focus of this paper- tend to engage in collusive corruption with border agents.

2018, Busia's border posts were replaced by a One Stop Border Post (OSBP), which is now regarded as one of the flagship OSBPs. Malaba, a smaller official border crossing between Kenya and Uganda, is situated 35 kilometers from Busia and is also used by traders who operate in the area. Busia, like many other border towns, relies on commerce. This has shaped the town in many ways; Busia (and by extension most of the county) counts many markets that attract suppliers, traders and buyers from all over the country, as well as from other neighboring countries. Agricultural products and food found in markets come from diverse sources, as some traders source from domestic suppliers while others cross the border to reach foreign suppliers. The area's economy – including employment – is centered around trade, and Busia's border crossing is considered the main focal point in the town's urban planning and mobility infrastructure.

1.2.2 Domestic and Cross-Border Trade: Formal and Informal Crossings

This paper focuses on small-scale informal traders who are found in markets in the areas surrounding Busia's border crossing. In this paper, informal traders are defined as businesses that operate without being officially registered. These small-scale traders can either be domestic traders (use domestic supply chains) or be cross-border traders (international traders). Due to the proximity of the border crossing, a disproportionate number of traders are cross-border traders (at least prior to the pandemic). Busia is not only one of the main official border crossings in East Africa, it also accounts for 74% of total informal agricultural trade flows between Kenya and Uganda. Informal cross-border trade (often referred to as ICBT in the literature) – trade activities which are not recorded in official trade statistics – is pervasive in developing countries. Much of this trade is conducted by small-scale traders who cross the border multiple times a week to source or market agricultural goods. A phone survey carried out in 2017 by Sauti East Africa showed that 80% of the traders are women, and an average trader trades 1.8 types of goods sourced from 1.7 markets and sold in 2.2 markets (Sauti-East-Africa (2017)). Traders in my sample who are cross-border traders can either trade through an official crossing, i.e., an OSBP (at Busia or Malaba) or through informal crossings. In Busia, people refer to the border as "porous"; there are ways to cross it informally without having to go through the formal checkpoints. Those routes, which are located on either side of the official checkpoints, are called informal routes (or "panya" routes in Swahili). The best-known ones in Busia are Sofia and Marachi (referred to as *main* informal border crossings in this paper). Although official border posts are manned by Kenya/Uganda Revenue Authority officials, the informal routes are manned by the police who are known to extract bribes in exchange for silence.

1.2.3 Costs and Information

Besides the costs incurred to purchase their goods, there is little evidence on the other costs faced by informal traders. It is, however, likely that they face cost structures that are very different to larger official traders. For one, they do not benefit from likely returns to scale in purchasing goods, transport, and, importantly for this paper, access to information. They do not operate in standard business circles and therefore have to rely on informal networks to get information. Informal traders lack reliable, accessible and accurate trade and market information. In addition to costs associated with finding reliable market information, lack of information and informality exposes traders to corruption and harassment.

At the official border crossings, informal traders — mostly women — often face challenging conditions and high barriers to trade, such as the prevalence of corruption among border officials, frequent harassment, and other personal safety risks. In their 2017 final report, Uganda Women Entrepreneurs Association (UWEAL) and TradeMark East Africa (TMEA) surveyed female cross-border traders across Uganda to identify key Non-Tariff Barriers (NTB) to trade. The four most frequently occurring NTBs were customs clearance issues (67% of respondents), payment of bribes (57%), immigration document requirements (30%) and roadblocks (17%). They also identify Uganda, Tanzania and Kenya as the countries that present the most NTBs. The results from their study are clear: 65% of respondents are able to clear their goods through customs in less than two hours, but the speed of clearance is bribe-driven. Forty-one percent of respondents pay a bribe every time they cross the border. Through survey evidence, they identify information asymmetry as being the main reason that female cross-border traders face unofficial charges and harassment in their attempts to conduct trade. Many women cross-border traders are not aware of the tariffs they should be paying, or of their rights or trade procedures. Clearing agents and border officials exploit this information gap to extract money for personal gain. A survey carried out between November 2016 and January 2017 in Busia shows that over 75% of the traders surveyed have encountered incidents of corruption at the border. Moreover, 80% of the respondents report that corruption at the border happens daily or weekly. High rates of harassment, coercive corruption at the official crossings and confusing procedures have been highlighted as reasons for traders to rely on informal crossings (in addition to wanting to avoid taxes).

At informal border crossings, informal traders pay bribes to the police against passage[10]. Safety concerns at informal crossings have also been reported.

[10] I have another paper that compares costs (including bribes) in both crossings through an audit study.

1.2.4 Insights on Intermediaries

High trade barriers and trade costs faced by domestic and cross-border traders – both at the official and informal border crossings – may be exacerbated by the fact that trading involves many types of actors as well as intermediaries. Legal actors include the Kenya Revenue Authority, the Uganda Revenue Authority, the police and the municipal tax collectors (the first three are only relevant to cross-border traders, municipal tax collectors to all traders). Revenue authorities are assigned to official border posts across the country for a specific amount of time. Anecdotally, revenue officials prefer some border posts to others because there is variation in terms of how much bribe money they can extract across different border posts. Twenty-four percent of traders surveyed report that revenue agents collect bribes at the border.

Municipal collectors collect municipal-level taxes. They are usually seen on either side of the border, located strategically to ensure they can stop everyone who imports or trade goods. They also locate themselves in marketplaces to levy sales-related taxes. Taxes depend on the type and value of the goods. Unlike the Revenue Authority officials, the municipal collectors are also found at informal border crossings.

Although the police do not have a mandate to collect taxes, 69% of the survey respondents report that the police collect bribes from traders, along informal border crossings and at roadblocks. Moreover, 58% report that the police are responsible for harassment at the border.

The other intermediaries include brokers, transporters and clearing agents. Payments to intermediaries are usually set through bargaining. There are also different types of transporters. Most relevant are those who help transport the goods across the border, usually via bike, because no motor vehicles are allowed to cross the border with shipments. Clearing agents help traders clear their goods and get the correct approvals and documents before crossing the border. There are established clearing agent companies that usually deal with large traders, but also individual clearing agents (or at least people who call themselves clearing agents) who patrol the border and offer services to smaller traders. Through the Simplified Trade Regime and Simplified Certificate of Origin, clearing goods should be an easy and quick task for traders with small consignments – such as traders in my sample – and should not require clearing agents. In practice, it seems that clearing agents take advantage of traders' lack of information about trade procedures.

1.2.5 Trade Policies in the Area

The Simplified Trade Regime and Simplified Certificate of Origin stem from regional integration efforts that acknowledge the role of informal traders as contributors to development, such as a supplementary source of family income to under-employed people. Therefore, to facilitate informal trade, member states from the East African Community (EAC) and the Common Market for East and Southern Africa (COMESA) have adopted Simplified Trade Regimes tailored to small-scale cross-border traders. One of these is the East African Community Certificate of Origin, a trade facilitation document which is used for clearance of goods that have been grown or produced in the EAC partner states and whose value is less than USD 2000. The simplified procedures were introduced in 2007 in an effort to reduce smuggling. In the EAC, 370 products currently qualify for clearance through the Simplified Certificate of Origin. The ease of use of the Simplified Certificate of Origin has allowed cross-border traders to clear their consignments quickly and with less hassle. In line with this definition, I consider "small-scale traders" to be traders who trade goods valued at less than USD 2000 (per trip).

1.3 Sample, Data and Descriptive Analysis

1.3.1 Sample and Data Collection

In January 2020, a census of traders who trade either agricultural goods or shoes and clothing was conducted in Kenyan markets located within a 40-km radius of the Kenya-Uganda border in Busia. One thousand six hundred fifty traders were censused in 30 markets, all located on the Kenyan side of the border. In February 2020, I carried out a round of baseline data collection for 1,100 randomly selected traders. All are small-size traders who transport their goods by foot, bike and motorbikes. Twenty percent are men and 80% are women. About 55% are cross-border traders and 45% are domestic traders. Thirty-seven percent mainly cross the official border while 63% prefer using the non-official border crossings. Eighty percent trade mostly in agricultural products while 20% trade in shoes and clothing. I should note that this sampling strategy captures a representative sample of traders located in Kenyan markets – therefore, they are mostly either Kenyans who trade domestically in Kenya or Kenyan cross-border traders who buy goods in Uganda and sell them in Kenya. This is a different sample than would have been selected if I had sampled traders crossing at the border itself – a sampling strategy often used by governments in an attempt to estimate informal trade. Table 1.1 Panel A presents the main socioeconomic characteristics of the traders in the sample.

Throughout 2020, a high-frequency phone survey was carried out at intervals of roughly two weeks to a month. Each phone survey round asked traders about their experiences "in the past two weeks". In February 2021, a second baseline (referred to as the updated baseline) was conducted to ensure that I had up-to-date data about traders before the intervention part of the experiment. The intervention was launched in May 2021 and three rounds of follow-up data were collected in June, July and August/September. The final endline was conducted in October and November 2021. My panel therefore counts 18 rounds of data collection, including a baseline, an updated baseline and an endline. Phone surveys collected outcomes on traders' businesses, including the health of their business, the type of goods they trade, and their supply chain. I also collect data on what trade route they choose as well as reports of corruption and harassment. In addition to collecting outcomes on traders' businesses, the phone surveys have also served as a way to collect details about shocks such as market closures, product bans, and market prices. Figure C.1.1 presents the timeline of events and Figure B.1.2 shows the attrition across the high-frequency rounds.

1.3.2 Other Data Sources

In addition to the survey data, this paper also uses

- Market-level price data from a phone platform generated by the intervention partner. The information experiment, which will be described in more detail below, includes providing price information to traders. I therefore have access to the same market price data, which I use to estimate market-level outcomes. The market data is not collected by the implementation partner; they bring together data from over 10 different sources, standardize it across products and markets, and continuously update the database by the most updated price point.

- Usage data from the phone platform: data on each trader's usage of the platform at the interaction level. This includes details of what the trader requested and what was sent back by the platform, at the interaction level.

1.4 New Insights on Informal Trade Sector

1.4.1 Closure of Official Border

The official border between Kenya and Uganda was closed between April 2020 and October 2020 due to the Covid-19 pandemic. The official border was closed to people, including small- and medium-size traders, but trade vehicles were allowed to go through as an attempt to encourage movement of goods and minimize trade disruptions.

The key dates and events related to Covid-19 are as follows: On the 12th of March 2020, the first case of Covid-19 was reported in Kenya and the official borders between Uganda and Kenya were closed. At the end of September 2020, the government announced that the borders were re-opening. In October 2020, the borders re-opened.

I use the closure of the official border as a shock and highlight key insights on the informal trade sector.

1.4.2 Key Descriptive Patterns

Informal traders can trade across different types of routes

Table 1.1 Panel A demonstrates that the majority of dealers in my sample are Kenyan importers. They can conduct business across three distinct sorts of routes. Forty-five percent are domestic traders who purchase and sell within Kenya, 19% are cross-border traders who import their goods through official border stations, and 36% use informal crossings. For 95% of them, trading is their primary source of income. Small-scale traders are typically opportunistic and deal in a variety of items, whereas larger traders prefer to specialize in a specific good.

Informal traders face large costs

Table 1.1 Panel B shows that over 80% of traders' sales are comprised of costs, leaving traders with low profits. Traders' costs include large purchasing costs as well as transport and border costs. Table A.3.2 characterizes border costs, depending on whether traders use the informal or formal crossings. Table A.3.2 does not suffer from selection bias because the data was collected through an "audit study", i.e., trained traders crossed both types of border and reported costs and experiences[11] . Table A.3.2 shows that bribes are extracted at both types of crossings. Bribes are larger at the informal border crossings, while waiting times are longer at the official border crossings.

Large price dispersion and lack of market integration

Prices for agricultural products vary across time, markets and products. Figure 1.2 plots prices across time for different markets in Uganda and Kenya for markets situated within 100 km of the study site. In a standard trade framework and assuming no friction, with full information and prices taken as given, traders would optimize their business by buying products in markets that have the lowest price and selling in markets that have the highest price, taking travel costs into consideration. Figure 1.2 shows that market prices vary significantly across markets, and patterns are similar for other goods. This should lead to arbitrage opportunities.

[11] I have another paper that looks at determinants of border costs, especially bribes, at both types of border crossings

Informal traders specialize

Despite large variation in market price, traders specialize in markets and routes. Figure 1.3 shows that over 70% of traders use only one type of route to trade their goods, i.e., they only use domestic supply chains, formal crossings or informal crossings. Thirty percent use a combination of those routes – however, barely any traders use both the formal and informal border crossings.

Similarly to the route specialization, Figure 1.3 shows that the specialization also happens at the market level. Eighty-five percent of traders always sell in the same market. The probability of buying in the same market is, however, lower – about 45% of traders always buy in the same market.

Market and route stickiness can be overcome with a large enough shock

Can a large enough shock make traders switch markets and routes? I use the closure of the official border as a shock to look at how traders' choices of routes and market change. The closure of the official border implies that traders who were using official routes faced a large disruption in their supply chain, as they no longer could reach suppliers.

Figure 1.4 shows that traders who were initially cross-border traders, and therefore most impacted by the border closure shock, did switch routes, to either domestic routes or informal routes. A small proportion exited the activity. Domestic traders, on the other hand, were disproportionately more likely to exit, as they were crowded out by cross-border traders switching to domestic supply chains (Figures A.4.3 and A.4.4). This implies that traders can trade profitably by switching routes and markets, assuming that they are rational profit maximizers.

Switching is sticky

What happens when the border re-opens? Figure 1.5 shows that only 25% of traders who had switched due to the border closure returned to their initial type of route after the border re-opened. Assuming that traders maximized their utility before the closure of the border, we would expect them to return to their initial optimal choice when the border re-opens. Instead, I find stickiness.

Lack of information

Table 1.1 Panel C shows that only 38% of traders know market prices for their main good in other markets. Only 51% report informing themselves about market prices and those who do get information rely on word of mouth.

1.4.3 Rationalizing Stylized Facts

There are several explanations that could explain the stickiness results highlighted above.

1. The world has changed and traders' optimal solutions are therefore different, e.g., uncertainty about future border closings, actual increased border costs (e.g., taxes, Covid requirements).

2. It is costly to get up-to-date information. The border shock pushed them to discover better solutions and there is no reason to switch back to sub-optimal outcomes.

3. Switching costs, e.g., relationships with suppliers, path dependency, route-specific capital.

There may be other explanations. Given the context and the apparent lack of information, this paper focuses on the possible role played by information frictions.

1.5 Theoretical Framework

Based on these stylized facts, I derive a model that provides the micro-foundations of the stickiness through information frictions about buying prices (or demand in buying markets), demand in selling markets and official border costs.

In this theoretical framework, I am modeling trader i's utility from profits gained from buying and selling a certain quantity of a good at time t. Trader i maximizes her utility by choosing quantities to sell in selling markets as well as which buying market and trade route to use. Trader i faces uncertainty about prices (or equivalently, demand) in buying markets as well as uncertainty about demand in selling markets (except her own selling market) and uncertainty about border costs. As noted above, I define a trade route as a choice between official cross-border trade, informal cross-border trade, or domestic trade (or exit). This theoretical framework also allows me to estimate market prices.

I solve the problem by backward induction. At time t and for a given route k, trader i decides which quantity q_{ihtk} to sell in her home market h and which quantity q_{ijtk} to sell in an alternative market j. Based on the optimal quantities and expected revenues for each of the four market routes, trader i then chooses which buying market-route k to use, in time t, i.e., which one yields higher utility. This is a one-period model; stocking decisions do not play a role.

1.5.1 Assumptions

Trader i is risk averse, trades in 1 good and sells in Kenya. She can sell in at most two markets: she always sells in her home Kenyan market h, where there is no randomness or uncertainty about demand, and can also sell in an alternative Kenyan market j, where there is randomness and uncertainty about demand. On the buying side, she chooses to buy her goods from one of four possible buying market-routes k: Uganda/Formal, Uganda/Informal, Kenya/Domestic or not at all (Exit). Traders behave as monopolies in selling markets in the sense that they sell differentiated products and have an upward-sloping supply curve with the elasticity of buying market price with respect to quantity $\epsilon^B \geq 0$. In addition to the costs associated with purchasing goods, trader i faces heterogeneous border costs. Note that this model presents traders as monopolies to simplify and provide clarity on the different moving pieces; however, the predictions of the model remain the same if traders are in Cournot competition equilibrium or are price takers with enough treated traders affecting aggregate demand and supply (see Appendix).

Each selling market has a simple downward-sloping demand curve[12] with constant elasticity[13]. Elasticities are assumed to be the same across all selling markets and across time (v does not vary by trader or time).

The inverse market demand function for selling market m is:

$$p_{mt} = \alpha \omega_{mt} q_{mt}^{\frac{1}{v}} \tag{1.1}$$

where ω_{mt} is some randomness in demand due to market-specific high-frequency demand shocks. The randomness in selling market ω_{mt} is normally distributed $(1, (\sigma^{\omega}_{mt})^2)$. Note that the price elasticity of demand is v ($v \leq -1$) and the inverse price elasticity of demand is $1/v$.

The supply curve function for buying market k is:

$$p_{kt} = \zeta b_{kt} Q^B_{kt}(q_{it}, q_{-i,t}) \tag{1.2}$$

where b_{kt} is some randomness in supply quantity in buying markets due to market-specific high-frequency supply shocks and Q^B_{kt} total quantity in market k. The randomness in buying market b_{kt} is normally distributed $(1, (\sigma^{\omega}_{kt})^2)$. Traders take into account the effect of their own demand on buying market prices but do not take into account how that may affect other traders' buying decisions [14].

[12]See appendix for the model without imposing structure on the demand curve.

[13]Its constant elasticity property makes it tractable in this context, but my conclusions do not depend on this assumption.

[14]Imposing a positive relationship between trader individual demand and market prices in buying markets is not necessary for the results to hold, i.e., traders could be price takers. I include this positive relationship to remain more conservative about the possible effects.

The randomness in this model appears in four places : (1) the selling market demand shocks ω_{mt} following a normal distribution $(1, (\sigma^{\omega}_{mt})^2)$, (2) the buying market level demand shocks b_{kt} informing price p^B_{kt} following a normal distribution (1, $(\sigma^B_{kt})^2$), (3) the border shocks bc_{kt} , following a normal distribution (1, $(\sigma^{BC}_{kt})^2$) and (4) λ_{ik} an unobserved/random preference term for route k following an extreme value distribution.

Figure 1.6 gives a representation of the model.

1.5.2 Trader's Maximization Problem

1.5.2.1 Trader's Utility

Trader i maximizes her (risk-averse) utility by maximizing expected revenues and minimizing expected costs. Her utility is a standard profit function including a quadratic term in price gaps between selling and buying markets and in border costs. The utility for trader i, using market route k at time t, is as follows. Note that trader i's utility for market route k at time t includes quantities sold in home market h and alternative market j, conditional on using market route k (which is why quantities vary by market route k).

$$\begin{aligned} &MaxV_{ikt} = \\ &E[\sum_{m=h,j} \{[p^S_{mtk}(q_{imtk}) - p^B_{kt}(q_{ijtk} + q_{ihtk})] * q_{imtk} + \delta[p^S_{mtk}(q_{imtk}) - p^B_{kt}(q_{ijtk} + q_{ihtk})]^2 * q_{imtk}\} - \\ &[\delta_3 BC_{ikt}(1 + \gamma_{ikt}) + \delta_4 BC^2_{ikt} + \mu d_{ikt}] * (q_{ihtk} + q_{ijtk}) + c_{iht}q_{ihtk} + c_{ijt}q_{ijtk} + \lambda_{ik} + u_{ikt}] \end{aligned} \tag{1.3}$$

Trader i has full information about random demand shocks ω_{ht} in her home market h but faces uncertainty about random demand shocks ω_{jt} in the other selling market j. Passing through the expectations, expected prices become $E[p_{jtk}] = \alpha E[\omega_{jt}] q^{\frac{1}{v}}_{ijtk}$ and $p_{htk} = \alpha \omega_{ht} q^{\frac{1}{v}}_{ihtk}$ and trader i's utility simplifies to the following:

$$\begin{aligned} &MaxV_{ikt} = \\ &[\alpha E[\omega_{jt}] q^{\frac{1}{v}}_{ijtk} - \delta_2 (\sigma^{\omega}_{ijt})^2] * q_{ijtk} + \alpha \omega_{ht} q^{\frac{1}{v}}_{ihtk} * q_{ihtk} - [E[p^B_{kt}(q_{ijtk} + q_{ihtk})] + \delta_1 (\sigma^B_{ikt})^2] * (q_{ihtk} + \\ &q_{ijtk}) - [\delta_3 E[BC_{ikt}](1 + \gamma_{ikt}) + \delta_4 (\sigma^{BC}_{ikt})^2 + \mu d_{ikt}] * (q_{ihtk} + q_{ijtk}) + c_{iht}q_{ihtk} + c_{ijt}q_{ijtk} + \lambda_{ik} + u_{ikt} \end{aligned} \tag{1.4}$$

with

- Expected revenues from selling in market h and j, conditional on using market route k: $[\alpha E[\omega_{jt}]q_{ijtk}^{\frac{1}{v}} - \delta_2(\sigma_{ijt}^{\omega})^2] * q_{ijtk} + \alpha\omega_{ht}q_{ihtk}^{\frac{1}{v}} * q_{ihtk}$
- Expected costs split between (i) purchasing costs $E[p_{kt}^B]$, (ii) border costs $\delta_3 E[BC_{ikt}] + \delta_4(\sigma_{ikt}^{BC})^2$ (Tariffs if k = Uganda/Formal and Bribes if k = Uganda/Informal), (iii) bargaining power $1 + \gamma_{ikt}$ (with $\gamma_{ikt} \geq 0$) (iv) distance μd_{ikt} and (v) selling market-specific marginal cost for home market c_{ijt} and alternative market c_{iht}
- λ_{ik} utility associated with using buying market-route k. λ_{ik} includes supplier relationship, experience/comparative advantage, access to information or fixed costs
- $\delta_1 \geq 0$, $\delta_2 \geq 0$, $\delta_3 \geq 0$, $\delta_4 \geq 0$

1.5.2.2 Order of Maximization

Order of maximization (backwards induction):
Step 1: For each possible market-route k, trader chooses optimal quantities q_{ihtk}^* and q_{ijtk}^* to sell in home market h and alternative selling market j, conditional on using market route k
Step 2: Taking optimal quantity for market route k as given $q_{ikt} = q_{ihtk} + q_{ijtk}$, trader i chooses which market route k^* to use (Uganda/Formal, Uganda/Informal, Kenya/Domestic, Exit) to maximize utility V

1.5.3 Step 1: Solving for Prices and Quantities in Selling Markets

1.5.3.1 Solving for Prices

Trader i chooses q_{ihtk} and q_{ijtk} by maximizing V_{ikt}. Trader i therefore computes an optimal pair of quantities sold in home and alternative market q_{ihtk} and q_{ijtk} for each of the four alternative market routes k. The derivations are included in Appendix. I maximize trader's utility by taking first order conditions with respect to quantities in home and alternative selling markets.

Following the standard monopoly optimal pricing strategy, setting the mark-up over marginal costs as a function of the price elasticity of demand in the selling market, I solve for price[15] as a function of the price elasticity of demand in selling market v and price elasticity of supply (elasticity of marginal cost)[16] ϵ_{kt}^{Buy}

[15]See Appendix for derivations of the model without structure

[16]Again, I am assuming the partial effect on expectation of price is the same as the partial effect on price

$$E[p^S_{jtk}] = \frac{1}{1+1/v} * [\zeta E[b_{kt}]Q^B_{kt}(1+\frac{2}{\epsilon^{Buy}_{kt}}) + \delta_3 E[BC_{ikt}](1+\gamma_{ikt}) + \delta_4(\sigma^{BC}_{ikt})^2 + \mu d_{ikt} + c_{ijt} + \delta_1(\sigma^B_{ikt})^2 + \delta_2(\sigma^\omega_{ijt})^2]$$
$$p^S_{htk} = \frac{1}{1+1/v} * [\zeta E[b_{kt}]Q^B_{kt}(1+\frac{2}{\epsilon^{Buy}_{kt}}) + \delta_3 E[BC_{ikt}](1+\gamma_{ikt}) + \delta_4(\sigma^{BC}_{ikt})^2 + \mu d_{ikt} + c_{iht} + \delta_1(\sigma^B_{ikt})^2] \tag{1.5}$$

with $1/v = 1/\epsilon^{Sell}_{htk} = \frac{\partial p^S_{htk}(q_{ihtk})}{\partial q_{ihtk}} * \frac{q_{ihtk}}{p^S_{htk}} = 1/\ \epsilon^{Sell}_{jtk} = \frac{\partial p^S_{jtk}(q_{ijtk})}{\partial q_{ijtk}} * \frac{q_{ijtk}}{p^S_{jtk}}$ as I am assuming partial effect on expectation of price is the same as partial effect on price $\frac{\partial E[p^S_{jtk}(q_{ijtk})]}{\partial q_{ijtk}} * \frac{q_{ijtk}}{E[p^S_{jtk}]} = \frac{\partial p^S_{jtk}(q_{ijtk})}{\partial q_{ijtk}} * \frac{q_{ijtk}}{p^S_{jtk}}$; and $1/\epsilon^{Buy}_{kt} = \frac{\partial p^B_{kt}(q_{ijtk}+q_{ihtk})}{\partial q_{ihtk}} * \frac{q_{ihtk}}{p^B_{kt}} = \frac{\partial p^B_{kt}(q_{ijtk}+q_{ihtk})}{\partial q_{ihtk}} * \frac{q_{ijtk}}{p^B_{kt}}$

1.5.3.2 Selling Market Entry Conditions

Traders enter selling market m if their expected profits from selling in market m are positive. The entry condition for home market and for alternative market are such that expected profits from selling in markets h and j exceed cost of entry (see Appendix for derivations).

1.5.3.3 Solving for Quantities

Using the price function (1.1), (1.2) and the optimal price expressions from the optimization (A.4), I solve for quantities, including market entry conditions (A.6):

$$q_{ihtk} = [\frac{1}{\alpha\omega_{htk}} * \frac{1}{1+1/v} * [\zeta E[b_{kt}]Q^B_{kt}](1+\frac{2}{\epsilon^{Buy}_{kt}}) + \delta_3 E[BC_{ikt}](1+\gamma_{ikt}) + \delta_4(\sigma^{BC}_{ikt})^2 + d_{ikt} + c_{iht} + \delta_1(\sigma^B_{ikt})^2]^v \tag{1.6}$$

with $\alpha\omega_{ht} q^{1/v}_{htk} - \zeta E[b_{kt}]Q^B_{kt} - \delta_1(\sigma^B_{ikt})^2 - \delta_3 E[BC_{ikt}](1+\gamma_{ikt}) - \delta_4(\sigma^{BC}_{ikt})^2 - \mu d_{ikt} \geq c_{iht}$

$$q_{ijtk} = \begin{cases} 0 \text{ if } E[\omega_{jt}]\alpha_{jt}q_{ijtk}^{1/v} - [\zeta E[b_{kt}]Q_{kt}^B] - \delta_1(\sigma_{ikt}^B)^2] - \delta_3 E[BC_{ikt}](1+\gamma_{ikt}) - \\ \delta_4(\sigma_{ikt}^{BC})^2 - \mu d_{ikt} - \delta_2(\sigma_{ijt}^{\omega})^2 < c_{ijt} \\ \\ [\frac{1}{\alpha E[\omega_{jt}]} * \frac{1}{1+1/v} * [\zeta E[b_{kt}]Q_{kt}^B](1+\frac{2}{\epsilon_{kt}^{Buy}}) + \delta_3 E[BC_{ikt}](1+\gamma_{ikt}) + \delta_4(\sigma_{ikt}^{BC})^2 + \\ \mu d_{ikt} + c_{ijt} + \delta_1(\sigma_k^B)^2 + \delta_2(\sigma_{ijt}^{\omega})^2]]^v \quad \text{ if } E[\omega_{jt}]\alpha_{jt}q_{ijt}^{1/v} - [\zeta E[b_{kt}]Q_{kt}^B] - \\ \delta_1(\sigma_{ikt}^B)^2 - \delta_3 E[BC_{ikt}](1+\gamma_{ikt}) - \delta_4(\sigma_{ikt}^{BC})^2 - \mu d_{ikt} - \delta_2(\sigma_{ijt}^{\omega})^2 \geq c_{ijt} \end{cases} \tag{1.7}$$

1.5.4 Step 2: Choosing Buying Market and Route

1.5.4.1 Choice model

Trader i will compare her utility across each market route, taking the optimal quantity for each route as given.

Trader i will pick buying market route k' iff $V_{ik't} \geq V_{ikt}$ (See Appendix for derivations). The intuition is that increased profits from lower marginal costs in a new buying market route need to be larger than the lost utility from switching market routes $\lambda_{ik} - \lambda_{ik'}$.

There is a $\bar{\lambda}_i$, at which $V_{ik't} = V_{ikt}$. And if $\lambda_{ik} - \lambda_{ik'} \leq \bar{\lambda}_i$, trader i switches to k'. $\lambda_{ik} = \hat{\lambda_k} + \lambda'_{ik}$ with λ'_{ik} being an unobserved/random term that follows an extreme value distribution.

1.5.4.2 Choice Probabilities

Using a Mixed Logit Model, the probability of choosing buying market route k is :

$Prob(Y_{it=k}) = \int \frac{exp(V_{ikt}(\beta))}{\Sigma exp(V_{ikt}(\beta))} * f(\beta|\theta) * d\beta$ with β coefficients in V and θ parameters for the mixing distribution, estimated through simulations.

1.5.5 Predictions and Comparative Statics

The model will be estimated in the last section of the paper. In this section, I derive how quantities, prices and choice of market-trade route vary with marginal changes in costs. I specifically look at the effect of the following costs (as they are directly related to information frictions):

- $E[p_{kt}^B(q_{ihtk} + q_{ijtk})]$ as variable purchasing cost for route k

- $(\sigma_{ikt}^{B})^2$ as purchasing cost uncertainty for route k
- $VC_{ikt} = \delta_3 E[BC_{ikt}](1+\gamma_{ikt}) + \delta_4(\sigma_{ikt}^{BC})^2 + \mu d_{ikt} + c_{ijt}$ as other variable cost for route k (including mean and variance of border costs, distance and selling market specific marginal costs)
- $(\sigma_{ijt}^{\omega})^2$ as selling price uncertainty

Table 1.2, Panels (A) and (B) show the marginal effect of a change in purchasing costs, purchasing costs uncertainty, other variable costs and selling price uncertainty on quantity and prices. Note that those marginal changes are conditional on using route k and assume entry conditions are met. Since $\upsilon \leq -1$ and $\epsilon_{kt}^{Buy} \geq 0$; all marginal price changes are ≥ 0 and marginal quantity changes ≤ 0. I will describe in the next section that reducing information frictions (decreasing the four types of costs outlined above) theoretically leads to a reduction in equilibrium prices and an increase in quantities traded.

Choice of supplier market route:

- $\frac{\partial V_{ikt}}{\partial E[p_{kt}^{b}]} \leq 0 \implies \frac{\partial ProbY_{it}=k}{\partial E[p_{kt}]} \leq 0$
- $\frac{\partial V_{ikt}}{\partial (\sigma_{ikt}^{B})^2} \leq 0 \implies \frac{\partial ProbY_{it}=k}{\partial (\sigma_{kt}^{B})^2} \leq 0$
- $\frac{\partial V_{ikt}}{\partial E[BC_{ikt}]} \leq 0 \implies \frac{\partial ProbY_{it}=k}{\partial E[BC_{ikt}]} \leq 0$
- $\frac{\partial V_{ikt}}{\partial (\sigma_{ikt}^{BC})^2} \leq 0 \implies \frac{\partial ProbY_{it}=k}{\partial (\sigma_{ikt}^{BC})^2} \leq 0$

An increase in market route price (mean or variance), border costs (mean or variance) reduces the probability of choosing that market route.

1.5.6 Effect of Reducing Information Frictions

A reduction in information frictions about market prices and official border costs (tariffs) implies:
E$[p_{kt}^{B}] = p_{kt}^{B}$ & reduces variance $(\sigma_{ikt}^{B})^2$ for buying markets [Marginal Cost Effect] (a)
E$[\omega] = \bar{\omega}$ & reduces variance $(\sigma_{ijt}^{\omega})^2$ for selling markets [Marginal Revenue Effect] (b)
E$[BC_{ikt}] = BC_{ikt}$ & reduces $(\sigma_{ijt}^{BC})^2$ for k = Formal [Border Cost Effect] (c)
A reduction in γ_{ikt} for k = Formal [Bargaining Effect] (d)

I describe below how the treatment's four effects impact equilibrium market prices and trader's quantities in home selling market, alternative selling market and buying markets (according to the model). As an extreme case, let's assume treatment removes all uncertainty.

(a) MARGINAL COST EFFECT: $E[p^B_{kt}] + \delta_1(\sigma^B_{ikt})^2 = p^B_{kt} \implies MC_{ikt} \downarrow$

- Home selling market: $\Delta p_{htk} < 0$ and $\Delta q_{htk} > 0$
- Other selling market: $\Delta q_{jtk} \geq 0$ and $\Delta p_{jtk} \leq 0$ as
  - If entry condition already satisfied pre-treatment $E[p^S_{jtk}(q_{ijtk}) - \delta_2(\sigma^\omega_{ijt})^2] - E[p^B_{kt}(q_{ihtk} + q_{ijtk}) + \delta_1(\sigma^B_{ikt})^2] - BC_{ikt} * (1 + \gamma_{ikt}) + d_{ikt} \geq c_{ijt} \implies \Delta E[p_{jtk}] < 0$ and $\Delta q_{jtk} > 0$
  - If entry condition not satisfied pre-treatment and still not satisfied post-treatment: $q_{ijtk} = 0 \implies \Delta q_{jtk} = 0$ and $\Delta P_{jtk} = 0$
  - If entry condition not satisfied pre-treatment and becomes satisfied post-treatment: $\Delta q_{jtk} > 0$
- Buying market: Probability of switching buying market $\uparrow$
  - If same buying market [based on λ] ($k* = k_0$): $\Delta p_{ik*t} > 0$ as $\Delta q_{ik*t} > 0$
  - If new buying market [based on λ] ($k* \neq k_0$): $\Delta p_{ik*t} > 0$ as $q_{ik*t} > 0$ and $\Delta p_{ik_0t} < 0$ as $\Delta q_{ik_0t} < 0$

(b) MARGINAL REVENUE EFFECT: $\mathrm{E}[\omega_{ijt}] + \delta_2(\sigma^\omega_{ijt})^2 = \bar{\omega}_{jt} \implies MR_{ijt} \uparrow$

- Other selling market: $\Delta q_{jtk} \geq 0$ and $\Delta p_{jtk} \leq 0$ as
  - If entry condition already satisfied pre-treatment: $\Delta p_{jtk} < 0$ and $\Delta q_{jtk} > 0$
  - If entry condition not satisfied pre-treatment and still not satisfied post-treatment: $q_{ijtk} = 0 \implies \Delta q_{jtk} = 0$ and $\Delta p_{jtk} = 0$
  - If entry condition not satisfied pre-treatment and becomes satisfied post-treatment: $\Delta q_{jtk} > 0$
- Buying market: $\Delta q_{itk*} \geq 0$

(c) BORDER COSTS EFFECT $E[BC_{ikt}] + \delta_4(\sigma^{BC}_{ikt})^2 = BC_{ikt}$

- Same conclusions as $E[p^B_{kt}] + \delta_2(\sigma^B_{ikt})^2 = p^B_{kt} \implies MC_{ikt} \downarrow$ but only for $k = F$, i.e., for market routes that are formal

(d) BARGAINING EFFECT γ_{ikt} is reduced

- Same conclusions as $E[p^B_{kt}] + \delta_2(\sigma^B_{ikt})^2 = p^B_{kt} \implies MC_{ikt} \downarrow$ but only for $k = F$, i.e., for market routes that are formal

Overall model predictions about reducing information frictions about market prices and official border costs

(1) *Reducing information frictions leads to more markets connected by trade.*
A reduction in information frictions leads to traders buying in new markets[17]. They also sell in new markets (in addition to their home market), thereby increasing the number of markets in which they sell.

(2) *Reducing information frictions leads to higher trade volumes, higher sales and profits for traders.*

(3) *Reducing information frictions increases cross-border trade and formalization (both incidence and trade volumes).*

(4) *Reducing information frictions has ambiguous effects on bribes.*
Demand for cross-border trade increases (higher likelihood of corruption), but demand for informal crossing decreases. Moreover, bargaining at the border increases (by assumption, lower likelihood of corruption).
(5) *Reducing information frictions leads to cheaper prices for the consumer and more supply.*
Traders decrease prices and increase quantity in their home market to account for the reduction in marginal cost. Passthrough is non-zero.

(6) *Reducing information frictions leads to higher buying prices in new buying markets and may lead to cheaper buying prices in initial buying markets.*
Overall quantity purchased and sold increases, leading to an increase in price in new buying markets. Prices in initial buying markets are ambiguous, as overall quantity bought increases (putting upward pressure on the price) but the probability of switching to a new buying market increases too (putting downward pressure on the price). Prices decrease for markets that face significant switching-out, while markets with little switching-out face an increase in prices.

[17] The number of buying markets also increases if the model is expanded across multiple time periods.

1.6 Empirical Strategy

In May 2022, I carried out a Randomized Controlled Trial based on an information intervention. Treatment traders received access to trade and market information through a phone platform, while Control traders did not.

1.6.1 Intervention: Access to Information

Treated traders received free access to a mobile-based trade and market information platform developed by Sauti East Africa[18]. Sauti East Africa empowers women-led small and medium-sized enterprises to trade legally, safely and profitably across East Africa's borders. The platform is accessible on any phone through USSD[19] and offers multiple features (Figure A.4.5). First, it provides information about market prices for each product in main markets across East Africa, encouraging traders to seek out profitable opportunities in terms of goods and markets. Second, it gives traders information about exchange rates, helping them decide when to trade cross-border and to better negotiate exchange rates. Third, traders can request information on taxes, tariffs and procedures applicable to traded products, informing them of cross-border procedures and increasing their bargaining power at the border to reduce corruption and harassment. Fourth, traders can get access to weather forecast for the next day in locations of their choice. Examples of weather forecast include "Partly Cloudy", "Very Cloudy", "Rainy", "Sunny". Traders are interested in weather forecast for the business and it both gives them information about the market conditions in potential markets on specific days (e.g., fewer customers on a rainy day) as well as transport costs (e.g., longer public transport waiting time).

"Now, at the comfort of my couch or kitchen, I can get all the business and customs information I need right in my cheap old phone. I'm now more confident to pass through the gazetted route and not scared of personally clearing my goods. It is like a secret partner in my business. Before I even leave my house I know the price of groundnuts in Gulu and Lira, the current exchange rate and the amount of tax I will pay." Interview collected by Sauti East Africa in Busia Uganda, 2016.

The platform can be accessed by all traders through an access code (i.e., traders would, for example, text #1234*). Usage was not restricted to my sample; however, my implementation partner had not targeted Busia or carried out marketing campaigns in the area, implying that the initial adoption rate at baseline was relatively low (as confirmed by screening questions

[18]https://sautiafrica.org

[19]Similar technology to text messages (SMS).

in the baseline survey). Treated traders received an invitation to a workshop where they were given the access code and shown how to use the platform. Workshops were held in 3 locations, over the period of 10 days.

1.6.2 Randomization

The randomization is at the trader level: 50% of traders received access to the platform. In addition, the treatment intensity was varied at the market x industry level to control for spillovers and estimate general equilibrium results at the market level. Industry is defined as agriculture or shoes and clothing; I therefore have approximately 60 market x industry clusters. The intensity of treatment at the market x industry level ranged from 0 to 75% of the market.

Trader-level randomization was stratified by market, gender and trader type (domestic, informal crossings, official crossings); while market-level randomization was stratified by market size and location area.

Table A.3.4 shows that treatment and control traders are balanced across covariates before implementation. Table A.3.5 shows that survey attrition post-implementation was balanced across treatment and control groups.

1.6.3 Take Up

Take-up is characterized in two steps: first, attendance to the workshop, and second, usage of the platform. The analysis in the paper focuses on "Intent to Treat" effects, and therefore includes traders who did not attend the workshops.

Eighty-five percent of treatment traders attended the workshops and take-up of the platform averaged 70%, i.e., 59% of treatment traders (irrespective of whether they attended the workshops) accessed the platform at least once.

The workshops were organized in April 2021 and the endline survey was carried out in November 2021. In this paper, I look at usage data from May 2021 to November 2021, i.e., 7 months of usage. Traders access Sauti East Africa's platform through sessions. In each session they can request as many features (market prices, weather, trade procedures or exchange rates) as they want and multiple alternatives for each features, e.g., multiple markets or goods if they request market prices, multiple locations if they request weather forecasts, multiple products and values if they request tax procedures and multiple amounts if they request exchange rates. I highlight below 3 main insights on traders' search process using the platform.

Traders continue to use the platform after implementation. Figure 1.7 shows the distribution of number of months of usage per trader. 40% of traders never use the platform.

Conditional on using the platform at least once, the mean number of months of usage is 2.8 months. Indeed, 67.5% of users use the platform up to 3 months. This however does not mean that for these 67.5% of users, the 3 months of usage necessarily are the first 3 and that traders stop using the platform after the first 3 months. Figure A.4.7 shows that more than 50% of traders use the platform both within the first 3 months and within the last 4 months. Figure 1.8 shows the distribution of users (Panel A), sessions (Panel B) and sessions by user (Panel C) across the 7 months. Conditional on using the platform in a given month, traders use the platform on average for 3 sessions a month.

Traders are interested in using the 4 features, query different features at different times and prioritize features that inform them of high frequency shocks to market conditions, i.e., market prices, weather and exchange rates. Figure 1.9 (left column) shows that amongst users, traders tend to look up a single feature per session but over 50% of users are interested in 3 or 4 features across the 7 months. Figure 1.9's right column shows that weather forecasts, market prices and exchange rates are the most demanded features, reaffirming that consistent with the model, information frictions seem to come from a lack of information on high frequency shocks that affect buying and selling markets i.e., weather (affecting demand and supply), market prices and exchange rates.

Over time, traders look up different markets and alternatives. Table A.3.6 shows the number of alternative market prices, exchange rate amounts, trade policies (tariffs) and weather locations requested by session, day, month and overall. Table A.3.6 shows that traders look up 3 markets over time and those markets are different from each other (see alternative per month versus unique alternatives per month). On the other extreme, weather forecasts are asked for multiple locations repeatedly.

The patterns highlighted here show that the workshops and the platform did not simply play the role of nudging traders to look elsewhere. Instead traders repeatedly request different markets, locations or alternatives to help them make informed decisions for their business.

1.6.4 Spillovers

The platform is accessible by all traders, conditional on having the access code. Spillovers in this context could be of two types: (1) control traders could have access to the code and use the platform and (2) treatment traders could tell control traders about the information included in the platform, without traders in the control group having to access the platform. In the first case, I can control for non-compliers as I have access to the usage data. However, this is not the case for the second type of spillovers. This is why I varied the intensity of treatment at the market x industry level to test for spillovers. Only 1.5% (9 traders) of control group traders accessed the platform.

1.6.5 Empirical Strategy

I run 3 types of regressions to assess the effect of treatment on outcomes for trader i, in round t and market m. $Treat_i$ is a dummy variable for whether the trader is in the treatment group, $TreatxPost_{it}$ and $IntensityTreatxPost_{mt}$ the interaction between the treatment variable and a dummy $Post$ relating to the outcome being measured after the intervention. T_t are rounds fixed effects and X_i are trader level characteristics/strata. I run 3 types of regressions to assess the effect of treatment on outcomes for trader i, in round t and market m. $Treat_i$ is a dummy variable for whether the trader is in the treatment group and $IntensityTreat_m$ the market intensity of treatment defined by the ratio of the number of treated traders over all traders, either buying or selling (depending on the specification) in market m at baseline. $TreatxPost_{it}$ and $IntensityTreatxPost_{mt}$ the interaction between the treatment variable and a dummy $Post$ relating to the outcome being measured after the intervention. T_t are rounds fixed effects and X_i are trader level characteristics/strata.

$Outcome_i = \alpha + \beta_2 Treat_i + X_i + \epsilon_i$

for a specific round t, with robust standard errors.

$Outcome_{it} = \alpha + \beta_1 Treat_i + \beta_2 TreatxPost_{it} + T_t + X_i + \epsilon_{it}$

for all rounds t including baseline, with standard errors clustered at the trader level.

$Outcome_{mt} = \alpha + \beta_1 IntensityTreat_m + \beta_2 IntensityTreatxPost_{mt} + T_t + \epsilon_{mt}$

for all rounds/month t including baseline, with standard errors robust or clustered at the market level.

The coefficient of interest is β_2.

1.7 Results

1.7.1 Main Results

Table 1.3 (Panel A) shows that information has a direct effect on trading. Treatment increases the likelihood of being in business by 3.5 to 4.6 percentage points and increases the probability of relying on trading as the main source of income by 3.5 percentage points. Moreover, treated traders also diversify by increasing the number of goods in which they trade by nearly 10% .

Table 1.3 (Panel B) shows that information led to an increase in the number of buying markets by 7% and an 11% increase in the number of selling markets (Model Prediction 1). This means traders expanded their set of markets and potentially switched away from their initial markets. Table A.3.7 shows that, after the intervention, an average control trader sold in 82% of the markets he/she used to sell in pre-intervention and 86% of control traders sold in at least one of their initial markets. However, treatment traders were significantly less likely to sell in the same market. This points to the fact that traders tend to stick to their selling markets but that the intervention both increased the number of markets sold in and induced a fraction of traders to switch out of their initial markets for new markets. On the other hand, only 40% of control traders report buying in the same markets pre- and post-intervention, pointing to the fact that switching costs are higher for selling markets than buying markets (this is not surprising, as traders have to pay fees to have a selling spot in a market). Information did increase the number of markets from which traders buy, but did not have an effect on whether or not traders switched out of their initial buying markets for other markets.

Relaxing information frictions improved treated traders' business by increasing both sales and profits by 17-18% at endline (Table 1.4 Panel A) (Model Prediction 2). As per the model, the positive effect on profits can stem from (i) buying cheaper quality-adjusted goods, (ii) selling goods at a higher price (increasing markups), (iii) reducing transport and border costs and/or (iv) increasing quantity. I am unable to assess the effect on quality; however, Table 1.4 (Panel B) shows that treated traders purchase higher quantities, leading to increased purchasing costs (Column 1). However, once I control for the increased quantity purchased, there is no effect on markups[20] (Columns 3 and 4). Either treatment isn't allowing traders to purchase goods at a lower price or they buy better quality-adjusted goods and have a near 100% passthrough. Equally I do not find treatment effects when I include other costs (in addition to purchasing costs) to the calculation of markups or significant treatment effects on profit margins[21]. Treatment does, however, seem to reduce transport and border costs by 11-13% (Columns 4 and 5 of Table 1.4 Panel C), meaning that information allowed traders to negotiate better. Note that, despite finding no effect on bribes paid, bribes are often lumped with transport and taxes, as traders are not aware of how much they should be paying for transport and taxes across the border. The negative impact on transport and border costs points in the direction of reduced corruption and bribes paid at the border.

Table 1.3 (Panel C) focuses on trade routes and shows that treatment pushed traders to become cross-border traders, resulting in a 20% increase in the incidence of cross-border traders (Model Prediction 3). This is a meaningful result as (i) domestic traders becoming cross-border traders implies both buying from new markets and navigating importing procedures; (ii) cross-border trade was relatively low at the time of the intervention due to the trade restrictions imposed during the pandemic (only 24% of control traders in my sample were cross-border traders post-implementation of Covid restrictions, a stark reduction from

[20]Here I am looking at an approximation of reversed markups: purchasing price over selling price. I do not have marginal costs and am therefore assuming here marginal cost equals average costs.

[21]Note that another explanation for finding no treatment effect on markups and profit margins is that measures of profits and costs tend to be noisy and I may not have the power to detect any effect

the 55% at the beginning of 2020). Moreover, the treatment helped formalize trade, increasing formal trade by 25%, which implies that domestic traders switching to cross-border trade opted to use the formal border crossing. Interestingly, the results on cross-border trade are concentrated in the first few months after implementation. Which traders became cross-border traders? Table 1.5 (Panel A) shows that traders who become cross-border traders due to the experiment are not traders who switched from cross-border trading to domestic during the border closure and did not switch back (called "Sticky" in the table).

Table 1.6 shows that there are no effects on reports of corruption or harassment (Model Prediction 4).

1.7.2 General Equilibrium Effects on Market Prices

Reducing information frictions has general equilibrium effects on market prices. Reducing information frictions has general equilibrium effects on market prices. I use the variation in treatment intensity to look at how market prices changed due to the intervention. I use two sets of data: (i) the market price data from the platform and (ii) reported buying/selling price data from surveys of traders in my sample. They both have advantages and drawbacks. The data in the platform is more complete and does not rely on sample traders actually buying or selling in markets; however, markets on the platform do not perfectly match those in my sample (geographically) and are more numerous. I therefore assign to each market on the platform the treatment intensity of the closest treated sample market. The markets included are located within 25 km of the study site.

Table 1.7 (Panel A) shows that reducing information frictions reduces aggregate consumer market prices. Indeed, markets that were more intensively treated have lower retail consumer prices (Model Prediction 5). In Columns 1 and 2, the regression is at the market price level and I control for product fixed effects. Goods included are agriculture goods and goods sold by traders in my sample. In Column 3 (my preferred specification) I build a consumer price index over all targeted goods, at the market level. Markets where more traders were treated experienced a relative decrease in consumer prices. Related to the model predictions, traders now buy goods at cheaper prices and pass the cost reduction through to consumers. Table 1.7 (Panel B) shows that reducing information frictions also affects market prices on the buying side. Treatment reduces market prices in markets in which traders used to buy (Model Prediction 6). Again Columns 3 and 6 show that markets that were more intensively experienced a larger reduction in prices, both for retail and wholesale prices. Related to the model, traders now buy in new markets, lowering demand in initial buying markets.

Table A.3.12 and A.3.13 find similar results when I use reported prices from traders' surveys, although the result is less obvious for the buying markets. Note that the results are robust to different ways of constructing the CPI variable.

As robustness checks, I first include markets that are farther away from the study site and control for distance to the border and the interaction of distance and treatment intensity. Table in Online Appendix shows that the treatment effect declines as distance from the study site increases. This implies that there is no effect of treatment on markets outside the study site. Second, instead of only looking at the effect on prices for goods and industries in which traders traded, I look at the effect of treatment on goods which the project does not focus on. I find no effect of treatment (Table in Online Appendix).

1.8 Magnitudes, Welfare Analysis and Counterfactuals

1.8.1 Magnitude of Treatment Effect and Information Frictions

Table 1.8 shows that the closure of the border increased market prices but that the increased costs differentially affected markets closer to the border, who are more likely to rely on cross-border trade. More generally, the coefficient on the interaction between Closure and Distance can be interpreted as a clean estimate of the cost of being a km farther away from the border.

Along those lines, comparing the treatment effect on prices (-21.55) in Table 1.7 to the cost of distance, treatment was equivalent to pushing markets closer to the border by about 150 km.

1.8.2 Cost Effectiveness

As a reminder, the key results from reducing information frictions are:

- Reduction in consumer prices in selling markets (initial and new selling markets)
- Increased quantity sold
- Reduction in purchasing prices in initial buying markets; increase in purchasing prices in new buying markets
- Increase in profits for treated traders
- Increase in official cross-border trade flows

Consumers' Welfare Gains Consumer prices decreased by 6.6% due to the intervention. Based on average sales of USD 9148.3 per year at baseline, this is equivalent to an increase in USD 604 in consumer surplus. In this welfare calculation, I focus on the first-order components and do not include the benefit for consumers of increased demand. This means that my analysis will be a lower bound.

Suppliers' Welfare Gains: I assume the decrease in prices in treated markets is compensated for by the increase in prices in new purchasing markets[22], leading to no change in welfare for suppliers. Again, this is underestimating the welfare gains, as it does not include the increase in purchased quantity (treated traders increase purchasing costs by 6-17%)

Traders' Welfare Gains: I assume the increased profits for treated traders are compensated for by a reduction in profits for other traders.

Governments' Welfare Gains: Treated traders' probability of using the official border increases by 20%. This is equivalent to an increase in USD 246 of trade flows over three months (based on USD 1229 per month of purchasing costs). Due to seasonality, and taking the fact that monthly sales during those three months needs to be 16x to get to yearly sales (USD 1634 for 3 months in February 2022, USD 9148 for the year), I assume official trade flow increases by USD 3936 by year, leading to an increase in USD 79 in tax revenues. Note that I do not include any potential effect on bribes in this analysis.

Intervention Costs: Sauti East Africa estimates their usage cost to be USD 7 per user. I do not include costs related to price data collection.

Welfare and Cost Effectiveness: Reducing information frictions leads to an increase in welfare of USD 683 per trader per year (USD 604 from consumer surplus, USD 79 from government surplus). At a cost of USD 7 per user, this is equivalent to a cost-benefit ratio of 1%.

1.8.3 Estimating the Model

Following the model described above, I estimate the parameters α and v in the model. I estimate the parameters first by simply matching means (which I refer to Simple Mean Matching method) and then by a two-step Generalized Method of Moment (GMM) estimation with Gauss-Newton optimization. I rely only on the updated baseline (February 2021) to (i) avoid contamination from the treatment after baseline and (ii) allow myself to do a out of sample test of the model for the 2020 data (see next sections). For the GMM method I either use markups (referred to as GMM with markups) or profit margins (referred to as GMM with profit margins). Table 1.9 shows the results.

[22]The model assumes iso-elastic supply curves, with the same elasticity across markets; which goes in the direction of my argument.

I use the following equilibrium relationship[23]:

$$p_m = \alpha \omega_m q_m^{\frac{1}{v}} \tag{1.8}$$

I add the following equilibrium results from the model :

$$p_m = \frac{1}{1 + 1/v} * M_i \tag{1.9}$$

with M_i being marginal costs for trader i[24].
From (1.8) and (1.9)

$$Markup_i = \frac{p_m * q_i}{M_i * q_i} = \frac{1}{1 + 1/v} \tag{1.10}$$

$$ProfitMargin_i = \frac{Profits_i}{Sales_i} = -\frac{1}{v} \tag{1.11}$$

The general idea (for either method) is that equations 1.10 or 1.11 can be used to estimate v and equation 1.8 can then be used to estimate α, using estimated v.

1.8.3.1 Simple Mean Matching

In this very simple method, I simply match means. v is computed from using the mean of trader level markups in the data (using purchasing costs only for costs) and assigning estimated v to the mean. Panel A of Table 1.9 shows the means from the data used for each method. For the Simple Mean Matching, I match average $Sales_i$, $nmarkets_i$, $Markups_i$ and p_m to the data. Only one α is computed by estimating the average quantity per market q_m taking traders' average total sales (corrected for the average number of markets sold in) and dividing this average by the average price index across markets (see equation 1.12 and 1.13). Note that I use Sauti East Africa's back-end data to get an average price per market. Panel B of Table 1.9 shows the estimated α and v.

[23]The time subscript is removed as I only rely on one round of data.

[24]Note that is assumes that each trader has the same markups and profit margins across the different markets they sell in.

$$Sales_m = \frac{Sales_i}{nmarkets_i} = \alpha \omega_m q_m^{\frac{v+1}{v}} \tag{1.12}$$

$$q_m = \frac{Sales_i}{p_m} * \frac{1}{nmarkets_i} \tag{1.13}$$

1.8.3.2 Generalized Method of Moment

I also estimate parameters using a two-step generalized method of moment estimation (GMM), with Gauss-Newton optimization. For v, I continue to use trader level markups but use them as instruments in a GMM estimation. For α, I estimate a specific α per good g, using the data at the transaction level. For each trader, I know the quantity of good sold in what market and at what price. The coefficient and SE in Panel B of Table 1.9 are therefore the average of all 30 estimated α and the respective standard error of the average. Each good level demand curve is estimated using prices per kg and quantities in kg. For this reason, in the GMM estimations (both GMM using markups[25] or profit margins) restrict the sample to agriculture goods for more consistent prices per kg.

Lastly, in the GMM estimation with profit margins, instead of using markups that only relies on purchasing costs for costs, I use trader level profit margins (dividing trader's profits by sales).

1.8.3.3 Results of Estimated Parameters

Panel B of Table 1.9 shows the estimated parameters and corresponding standard errors, using the three different methods. All methods end up estimating relatively similar parameters. v is estimated to be between -5.76 and -4.4 (depending on the methods) and is relatively precisely estimated. The average α across all goods varies between 273 and 353. Note that the large standard errors here do not mean α's are not precisely estimated, rather that there are (unsurprisingly) a large variation in α's across each good. Indeed, standard errors around each estimated α per good are on average 31.63 (not shown in Table).

1.8.4 Counterfactual Simulations

Using the estimated model, I now run a few counterfactuals. First, I use the model to predict prices in a scenario where formal crossings are closed. I compare my model predictions to

[25] Again, I use average costs rather than marginal costs.

the reduced form effect of the official border closure on prices. The data used to estimate my model (updated baseline) is different than the rounds of data used to estimate the reduced form effect of the official border closure. This implies that the comparison between my model predictions and reduced form results is valid. Then, I run a second counterfactual analysis and look at what would happen to market prices if the informal crossings were also shut down.

1.8.4.1 Out-of-Sample Prediction: No Formal Route

Using a Panel Mixed Logit model with route-specific random intercept (correlated), I estimate the mean and variance of the normal distribution of route-specific preferences (or fixed costs) as well as their covariance. That allows me to estimate the predicted choice probabilities, using control traders. Table 1.10 (Panel A), Row 1 shows choice probabilities averaged across the study period.

Table 1.10 (Panel A), Row 2 shows how, based on estimated route-specific preferences and covariance, the predicted choice probabilities change when a formal route is no longer an option. When the formal route closes, 81% of traders who used to use the official crossing are predicted to switch to domestic trade, 8% to informal border crossings, and 10% to exit[26]. Table 1.10 (Panel B) then highlights the model's predictions in terms of market prices. The model predicts that closing the formal border leads to a 7.5% increase in prices, which is comparable to the 8.7% increase estimated in the reduced form analysis of the effect of the closure of the official border (Table 1.8).

1.8.4.2 No Informal Route

A similar exercise can be carried out, now assuming the informal crossings are also closed. Table 1.10 (Panel A), Row 3 shows that 86% of informal traders who can no longer cross the border become domestic traders, while 14% exit. This means that, without the existence of informal crossings, 14% more traders would have stopped trading. Table 1.10 (Panel B) shows that, according to the model, having informal crossings prevented prices from going up by another 11.5%. This points to the importance of informal crossings in smoothing shocks. Without the existence of informal crossings, the closure of the border (and the consequences from Covid-19-related restrictions) would have led to significantly higher consumer prices.

[26]Note that the large share of domestic traders comes from the fact that I am using the updated baseline to estimate the model.

1.8.5 Discussion about Formalization

As described above, reducing information frictions leads to welfare improvements stemming from trade formalization (governments' welfare increase from more tax revenues). However, the majority of the welfare gains come from consumers being able to buy cheaper goods. My results, taken together, do not advocate for a complete formalization of trade (i.e., closing informal crossings) as a way to maximize welfare. To the contrary, the different counterfactual analysis presented in this section show that closing informal crossings would lead to a reduction in welfare, especially during border shocks. Indeed, we would see gains from formalization but a larger loss from increased consumer prices. Reducing information frictions leads to increased welfare, however policies that push for a complete formalization of trade may not.

1.9 Conclusion

This paper shows that information frictions play a significant role in informal trade, which is an under-studied but important segment of trade. Using the closure of the official border as a shock, the paper first documents key insights about informal trade that point toward the existence of informal frictions. I develop and estimate a model that embeds informal frictions taking the form of uncertainty about and randomness in market conditions in other buying and selling markets. I test the model using a Randomized Controlled Trial that gave traders access to information about market prices and formal border costs. I provide evidence that information frictions play a large role: reducing information frictions improves traders' profits, increases formalization of trade, and reduces equilibrium market prices for consumers, leading to a large increase in welfare. It is important to keep in mind that, despite the potential welfare improvements of switching to formal routes (due to increased tax revenues), informal routes significantly help smooth prices in the event of shocks at the formal border.

In future work, I hope to disentangle the role played by traders' increased bargaining power and reduction in uncertainty in the effects observed due to reducing information frictions. In addition, insights on whether the results highlighted by this paper continue to hold long-term seems crucial. More rigorous and experimental work should also be done on understanding market structures and how information frictions affect those. Lastly, while this paper focuses on small-scale informal traders, more research is needed on understanding the role of scale such as the inter-dependencies between informal small-scale traders and formal larger traders and whether there is a transition from one to the other.

Tables

Table 1.1: Traders' characteristics

	mean
Panel A. Socio-economic characteristics - Baseline	
CB-Official Crossing	0.19
CB-Informal Crossing	0.36
Domestic trader	0.45
Ag	0.79
Men	0.19
Age	40.81
Kenyan	0.94
Trade is main income	0.95
Has other source of income	0.40
N goods sold in past 3 months	2.51
Panel B. Traders' Costs - Updated baseline	
Total sales (3M, 00 Kshs)	1633.73
Total purchase costs (3M, 00 Kshs)	1228.83
Total costs (3M, 00 Kshs)	103.98
Total profits (3M, 00 Kshs)	288.20
Panel C. Information Environment - Baseline	
Market Prices	
Knows market price for main goods in other markets	0.38
Informs themselves about market prices	0.51
Echange Rates	
Informs themselves about exchange rates	0.18
Types of info shared	
Traders share info about market prices	0.77
Traders share info about exchange rates	0.29
Traders share info about taxes/tariffs	0.20
Observations	1166

Panel A shows means of traders' socio-economic characteristics. Panel B shows means of traders' profitability measures "in the past 3 months" (sales, costs and profits) measured in hundreds of Kenyan Shillings (Kshs) and over all goods traded. Traders who did not trade were assigned values of 0. Panel C describes means of variables related to traders' information environment and are all ratios or shares (i.e., total is 1). Panels A and C come from the baseline data (February 2020) and panel B comes from the updated baseline in February 2021.

Table 1.2: Model's Comparative Statics of Treatment Effect on Quantities and Prices

Panel A. Comp. Statics on Q	q_{ijtk}	q_{ihtk}
Purchasing Costs ($E[p^B_{kt}]$)	$(1+\frac{2}{\epsilon^{Buy}_{kt}})\frac{1}{\alpha E[\omega_{jt}]}\frac{1}{1+1/v}v[\frac{1}{\alpha E[\omega_{jt}]}\frac{1}{1+1/v}*Z_j]^{v-1}$	$(1+\frac{2}{\epsilon^{Buy}_{kt}})\frac{1}{\alpha\omega_{jt}}\frac{1}{1+1/v}v[\frac{1}{\alpha\omega_{jt}}\frac{1}{1+1/v}*Z_h]^{v-1}$
Purch Costs Uncert ($(\sigma^B_{ikt})^2$)	$\frac{\delta_1}{\alpha E[\omega_{jt}]}\frac{1}{1+1/v}v[\frac{1}{\alpha E[\omega_{jt}]}\frac{1}{1+1/v}*Z_j]^{v-1}$	$\frac{\delta_1}{\alpha\omega_{ht}}\frac{1}{1+1/v}v[\frac{1}{\alpha\omega_{ht}}\frac{1}{1+1/v}*Z_h]^{v-1}$
Other Variable Costs	$\frac{1}{\alpha E[\omega_{jt}]}\frac{1}{1+1/v}v[\frac{1}{\alpha E[\omega_{jt}]}\frac{1}{1+1/v}*Z_j]^{v-1}$	$\frac{1}{\alpha\omega_{ht}}\frac{1}{1+1/v}v[\frac{1}{\alpha\omega_{ht}}\frac{1}{1+1/v}*Z_h]^{v-1}$
Selling P Uncert ($(\sigma^\omega_{ijt})^2$)	$\frac{\delta_2}{\alpha E[\omega_{jt}]}\frac{1}{1+1/v}v[\frac{1}{\alpha E[\omega_{jt}]}\frac{1}{1+1/v}*Z_j]^{v-1}$	0

Panel B. Comp. Statics on P	p_{jtk}	p_{htk}
Purchasing Costs ($E[p^B_{kt}]$)	$\frac{1}{1+1/v}(1+\frac{2}{\epsilon^{Buy}_{kt}})$	$\frac{1}{1+1/v}(1+\frac{2}{\epsilon^{Buy}_{kt}})$
Purch Costs Uncert ($(\sigma^B_k)^2$)	$\frac{1}{1+1/v}\delta_1$	$\frac{1}{1+1/v}\delta_1$
Other Variable Costs	$\frac{1}{1+1/v}$	$\frac{1}{1+1/v}$
Selling Price Uncert ($(\sigma^S_\omega)^2$)	$\frac{1}{1+1/v}\delta_2$	0

Panel A shows how traders' quantities derived in the model change with a one unit change in the following: (i) Expected Purchasing Costs (line 1), (ii) Purchasing Costs Uncertainty (line 2), Other variable Costs which include Border Costs (means and variance), distance and market specific marginal cost (line 3) and (iv) Selling Price Uncertainty (line 4). The comparative statics are done for quantities in alternative market (Column 2) and home market (Column 3). In Panel B, the same structure is repeated for prices, i.e., Panel B shows comparative statics for the equilibrium market prices derived in the model. Note that $Z_j = E[p^B_{kt}(q_{iht}+q_{ijt})](1+\frac{2}{\epsilon^{Buy}_{kt}}) + \delta_3 E[BC_{ikt}](1+\gamma_{ikt}) + \delta_4(\sigma^{BC}_{ikt})^2 + \mu d_{ikt} + c_{ijt} + \delta_1(\sigma^B_{ikt})^2 + \delta_2(\sigma^\omega_{ijt})^2$ and $Z_h = E[p^B_{jt}(q_{iht}+q_{ijt})](1+\frac{2}{\epsilon^{Buy}_{kt}}) + \delta_3 E[BC_{ikt}](1+\gamma_{ikt}) + \delta_4(\sigma^{BC}_{ikt})^2 + \mu d_{ikt} + c_{iht} + \delta_1(\sigma^B_{ikt})^2$.

Table 1.3: RCT Results: Trade, Supply Chain and Trade Route Choice

Panel A: RCT Results: Trade Outcomes

	Traded		N Goods		Trade Main Income
	(1) Rounds 1-3	(2) Endline	(3) Rounds 1-3	(4) Endline	(5) Endline
Treatment	-0.007 [0.008]	0.035* [0.021]	-0.061 [0.140]	0.007 [0.188]	0.035** [0.014]
Post x Treatment	0.046*** [0.016]		0.230** [0.103]		
Dep Var Mean (Control)	0.964	0.873	2.784	2.690	0.936
R-Squared	.037	.003	.013	0	.007
Pre-Period	X		X		
Observations	4952	915	4951	914	894

Panel B: RCT Results: Supply Chain Outcomes

	N Supp. Markts		N Selling Markets		N Trips	
	(1) Rounds 1-3	(2) Endline	(3) Rounds 1-3	(4) Endline	(5) Rounds 1-3	(6) Endline
Treatment	-0.029 [0.042]	0.085** [0.041]	-0.105 [0.072]	0.105 [0.076]	0.410** [0.178]	1.863 [1.236]
Post x Treatment	0.078* [0.044]		0.181*** [0.060]			
Dep Var Mean (Control)	1.172	1.017	1.593	1.424	2.610	20.435
R-Squared	.033	.005	.01	.002	.005	.003
Pre-Period	X		X			
Observations	3653	912	3766	914	2832	913

Panel C: RCT Results: Choice of Trade Routes

	Cross Border			Formal		Informal	
	(1) Rounds 1-3	(2) Endline	(3) Rounds 1-2	(4) Rounds 1-3	(5) Endline	(6) Rounds 1-3	(7) Endline
Treatment	0.020 [0.024]	0.024 [0.030]	0.047* [0.026]	-0.026 [0.016]	-0.002 [0.018]	0.042** [0.021]	0.018 [0.025]
Post x Treatment	0.015 [0.022]			0.035** [0.017]		-0.008 [0.021]	
Dep Var Mean (Control)	0.399	0.279	0.247	0.150	0.083	0.246	0.162
R-Squared	.072	.001	.003	.02	0	.052	.001
Pre-Period	X			X		X	
Observations	4947	914	1886	4947	914	4947	914

This table shows the treatment effects of the Randomized Controlled Trial on trade outcomes (Panel A), supply chain outcomes (Panel B) and route choice outcomes (Panel C). In Panel A, Traded is measured by a binary variable switching to 1 if traders traded in the past two weeks (Column 1) and in the past month (Column 2). Columns labeled "Rounds 1-3" include follow-up surveys 1, 2 and 3; Columns labeled "Rounds 1-2" include follow-up surveys 1 and 2; while Columns labeled "Endline" focus on the endline. Columns that include "Pre-period" means that the specification included baseline and updated baseline. When that's the case, the variable of interest is "Post x Treatment". Columns labeled "Rounds 1-3" include rounds fixed effects. Standard errors (reported in brackets) are clustered at the trader level for specifications labeled "Rounds 1-3" and robust otherwise. * p<0.1, ** p<0.05, *** p<0.01.

Table 1.4: RCT Results: Sales, Profits and Costs

Panel A: RCT Results: Sales and Profits

	Sales		Profits		Stock	
	(1) Rounds 1-3	(2) Endline	(3) Rounds 1-3	(4) Endline	(5) Rounds 1-3	(6) Endline
Treatment	0.556***	0.455*	0.483***	0.353*	0.175	0.138
	[0.192]	[0.250]	[0.172]	[0.214]	[0.230]	[0.292]
Dep Var Mean (Control)	8.751	10.028	7.059	8.456	6.421	6.576
R-Squared	.009	.004	.008	.003	.001	0
RoundFE	X		X		X	
Observations	2790	898	2792	895	2806	906

Panel B: RCT Results: Main Costs

	Total		Per Sales		Per Trip	
	(1) Purch. Costs	(2) Oth. Costs	(3) Purch. Costs	(4) Oth. Costs	(5) Purch. Costs	(6) Oth. Costs
Treatment	0.292**	0.039	0.009	-0.010*	0.133	-0.034
	[0.123]	[0.081]	[0.018]	[0.006]	[0.085]	[0.087]
Dep Var Mean (Control)	9.738	7.701	0.827	0.117	9.265	6.801
R-Squared	.007	.008	.001	.003	.006	.005
Observations	2430	2434	2402	2398	2076	2072

Panel C: RCT Results: Deep Dive in Other Costs

	Total			Per Sales			Per Trip		
	(1) Formal Taxes	(2) Transport	(3) Bribes	(4) Formal Taxes	(5) Transport	(6) Bribes (E)	(7) Formal Taxes	(8) Transport	(9) Bribes
Treatment	30.944	-315.911	107.014**	-0.002*	-0.011**	0.000	-5.563	-16.940	4.001
	[20.320]	[367.111]	[52.939]	[0.001]	[0.005]	[0.000]	[8.789]	[197.730]	[2.888]
Dep Var Mean (Control)	219.849	2238.009	78.128	0.018	0.081	0.001	111.202	860.783	5.826
R-Squared	.005	.001	.005	.002	.004	.001	.002	.001	.003
Observations	2832	2463	894	2416	2412	809	2102	2098	783

This table shows the treatment effects of the RCT on sales, profits and stock value (Panel A), costs (Panel B) and specific costs (Panel C). Values are in Kshs. In Panel A and B, they are transformed to inverted hyperbolic sine. B and in Panel C, they are in levels and in Kshs (transformations into IHS do not change the results). In Panel A, Columns labeled "Rounds 1-3" include follow-up surveys 1, 2 and 3, while columns labeled "Endline" focus on the endline. None of the specifications include baseline or updated baseline controls. Columns labeled "Rounds 1-3" include rounds fixed effects. Standard errors (reported in brackets) are clustered at the trader level for specifications labeled "Rounds 1-3" and robust otherwise. In Panel B and C, all specifications include rounds 1-3, except for Columns 3, 6 and 9 of Panel C, which includes the endline. * p<0.1, ** p<0.05, *** p<0.01.

Table 1.5: Results: Heterogeneity by Trader Type

Panel A: Treatment Effect Heterogeneity on Probability of Crossing the Border

	Prob of being Cross-border Trader
Treatment × Sticky (CB-Dom-Dom)	0.002
	[0.059]
Treatment × Adaptors (CB-Dom-CB)	0.126*
	[0.066]
Treatment × Domestic (Dom-Dom-Dom)	0.022
	[0.044]
Treatment × CB (CB-CB-CB)	0.230*
	[0.125]
Treat	-0.006
	[0.041]
Sticky (CB-Dom-Dom)	0.116***
	[0.042]
Adaptors (CB-Dom-CB)	0.436***
	[0.048]
Domestic (Dom-Dom-Dom)	-0.163***
	[0.031]
CB (CB-CB-CB)	0.527***
	[0.110]
Dep Var Mean (Control)	0.312
R-Squared	.295
Round FE	X
Observations	1845

Panel B: Treatment Effect Heterogeneity on Route Choice

	(1) Formalization-Init.Formal	(2) Formalization-Init.Inf	(3) Formalization-Init.Domestic
Treatment	-0.082	-0.077*	-0.033**
	[0.060]	[0.044]	[0.015]
Post x Treatment	0.034	0.053	0.044**
	[0.097]	[0.061]	[0.019]
Dep Var Mean (Control)	0.799	0.781	0.297
R-Squared	.181	.035	.021
Pre-Period	X	X	X
Round FE	X	X	X
Observations	319	675	4661

This table looks at heterogeneity of treatment on the probability of crossing the border to trade (Panel A) and on probability of choosing the official route (Panel B). In Panel A, heterogeneity groups are defined by status at baseline-updated baseline-endline. Rounds included in the analysis are follow-up 1 and 2. Standard errors (reported in brackets) are robust. Panel B runs separate specifications for each type of traders: those who were initially cross border traders crossing officially (Column 1), initial cross border traders crossing unofficially (Column 2) and initial domestic traders (Column 3). The specifications focus on baseline, updated baseline, follow-up surveys 1, 2 and 3 and endline and include rounds fixed effects. Standard errors (reported in brackets) are clustered at the trader level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 1.6: RCT Results: Non Tariff Barriers

	Corruption		Harassment		Corruption (CB Sample)		Harassment (CB Sample)	
	(1) Rounds 1-3	(2) Endline	(3) Rounds 1-3	(4) Endline	(5) Rounds 1-3	(6) Endline	(7) Rounds 1-3	(8) Endline
Treatment	0.009	0.023	0.000	-0.001	0.020	0.068	-0.001	-0.008
	[0.010]	[0.018]	[0.005]	[0.006]	[0.023]	[0.062]	[0.013]	[0.023]
Post x Treatment	0.002		0.004		0.003		0.011	
	[0.012]		[0.007]		[0.037]		[0.023]	
Dep Var Mean (Control)	0.046	0.067	0.013	0.008	0.137	0.339	0.036	0.036
R-Squared	.01	.002	.003	0	.085	.005	.012	.001
Pre-Period	X		X		X		X	
Observations	4905	915	4905	915	1465	226	1465	226

This table looks at the effect of the RCT on non tariff barriers such as corruption and harassment. The outcome variables are incidence of corruption and harassment, defined as the probability of traders reporting facing corruption (Columns 1-2 and 5-6) or harassment (Columns 3-4 and 7-8). Columns 1-4 look at the whole sample, while Columns 5-8 restrict the sample to traders who cross the border (note these are traders who cross the border at the time of the survey, not those who initially crossed the border at baseline; which therefore implies selection issues). In Columns 1, 3, 5 and 7, rounds included in the analysis are baseline, updated baseline and follow-ups 1, 2 and 3. The specifications include rounds fixed effects. The coefficient of interest is Post x Treatment. Standard errors (reported in brackets) are clustered at the trader level. In Columns 2, 4, 6 and 8, the only round included in the analysis is the endline. The coefficient of interest is Treatment. Standard errors (reported in brackets) are robust.
* $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 1.7: GE Effects of RCT: Market prices

Panel A: Selling Market Prices ("App Data")

	Retail (Sell)		
	(1)	(2)	(3)
	Price Levels	Price Logs	CPI (Log)
Intensity Treat x Post	-21.558***	-0.038*	-0.282**
	[3.369]	[0.023]	[0.116]
Intensity Treat	14.819***	-0.013	0.174
	[3.333]	[0.022]	[0.114]
Post	22.448***	0.106***	0.226**
	[2.713]	[0.016]	[0.094]
Dependent Variable Control Mean	117.252	4.471	4.271
R-Squared	.728	.744	.067
Product FE	X	X	
Observations	21965	21965	323

Panel B: Buying Market Prices ("App Data")

	Retail (Buy)			Wholesale (Buy)		
	(1)	(2)	(3)	(4)	(5)	(6)
	Price Levels	Price Logs	CPI (Log)	Price Levels	Price Logs	CPI (Log)
Intensity Treat x Post	-6.489	-0.144***	-0.246***	-15.643***	-0.128***	-0.291***
	[5.145]	[0.012]	[0.063]	[2.041]	[0.012]	[0.063]
Intensity Treat	2.754	0.036***	0.137***	8.091***	0.018*	0.142***
	[2.496]	[0.010]	[0.048]	[2.026]	[0.011]	[0.052]
Post	6.299**	0.123***	0.089**	-13.957***	-0.011	-0.147***
	[2.709]	[0.008]	[0.038]	[1.294]	[0.008]	[0.040]
Dependent Variable Control Mean	90.846	4.247	4.243	83.367	4.117	4.122
R-Squared	.157	.607	.078	.375	.622	.224
Product FE	X	X		X	X	
Observations	62184	62184	1064	56339	56339	1013

This table looks at the effect of treatment intensity on aggregate market prices, using back-end data from the phone platform. Goods included are agriculture goods and the types of goods sold by traders in my sample. Panel A describes selling prices and Panel B buying prices (Columns 1-3 are retail, Columns 4-6 are wholesale). Variable Intensity of Treatment is constructed using the intensity of treatment of a buying or selling market based on baseline randomization. As the platform markets do not perfectly match the randomization, the randomization value of the closest sample market is assigned to the platform market. I only include markets that are less than or equal to 25kms to a sample market. Data ranges from 2019 to 2021. Post in a binary variable taking the value of 1 if the data is after implementation. Specifications control for countries and currency fixed effects. Columns 1 and 4 run a specification that includes all data and controls for product fixed effects. Columns 2 and 5 run the same specification in logs. Columns 3 and 6 run a market level specification after creating a standard consumer price index (average of log prices). The CPI here is flatly weighted (see Appendix for weighted CPIs). Standard errors (reported in brackets) are clustered at the market level.
* p<0.1, ** p<0.05, *** p<0.01.

Table 1.8: Comparing treatment effect to cost of distance

	Retail		Wholesale	
	(1) Price Levels	(2) CPI (Log)	(3) Price Levels	(4) CPI (Log)
Closure x Dist to Border	-0.146*	-0.002**	-0.110	-0.002**
	[0.085]	[0.001]	[0.069]	[0.001]
Distance to border (Kms)	0.143*	0.001*	0.103	0.002***
	[0.085]	[0.001]	[0.072]	[0.001]
Closure of Off Border	6.979	0.093*	4.124	0.072
	[7.264]	[0.051]	[6.368]	[0.053]
Dep Var Mean (Control)	79.273	4.208	64.857	3.872
R-Squared	.401	.238	.718	.18
Year FE	X	X	X	X
SE Clustered	X		X	
Selected Goods	X	X	X	X
Dist from Border	100	100	100	100
Observations	28534	457	25335	441

This table regresses market prices from the platform's back-end data on distance to the border, a dummy variable for whether the official border was closed and the interaction of both of these variables. The specification also includes country fixed effects, currency fixed effects and year fixed effects. Distance is calculated as a straight line from the border and is in kms. Market prices used are prices from markets located within 100 kms from the border in Busia. Goods included are agriculture goods and types of goods sold by traders in my sample. I only include goods that were easily convertible to a price per kg. Columns 1-2 report retail prices and Columns 3-4 report wholesale prices. Columns 1 and 3 use all prices in Kenyan Shillings per kg (Ugandan prices were transformed to Kshs using the average exchange rates) and Columns 2 and 4 use a standard consumer price index (average of log prices). The CPI here is flatly weighted. Standard errors (in bracket) are clustered at the market level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 1.9: Estimating Model's Parameters

Panel A: Sample data used for moments	Mean	SE
Simple Mean Matching		
$Sales_i$ (Kshs)	24616	
$nmarkets_i$	1.6	
$Markup_i$	1.21	
p_m (index)	117	
Computed $Sales_m$ (Kshs)	15385	
Computed q_m (index bundle)	131.5	
GMM		
$Markup_i$	1.53	2.02
$ProfitMargin_i$	0.22	0.23
q_{igm} (Kg)	4587.73	21955.89
p_{igm} (Kshs per kg)	81.42	86.45

Panel B: Parameter Estimates	Coefficient	SE
Simple Mean Matching		
v	-5.76	
α	273	
GMM with markups		
v	-4.40	0.2
α_g	353.02	243.58
GMM with profit margins		
v	-4.86	0.12
α_g	306.73	204.94

This table shows the estimates of the model parameters. Panel A shows the summary statistics taken from the data and used to estimate the parameters. Panel B shows the resulting parameter estimates and Standard Errors when relevant. I use 3 methods to estimate the parameters. First, the Simple Mean Matching simply matches means. v is computed from trader level markups (using purchasing costs only). Only one α is computed using traders' average sales (corrected for the average number of markets sold in) and the average price index across markets. Second I use a generalized method of moment estimation, using sample data variables as instruments and a Gauss-Newton optimization in a 2 step estimation. Here, a α per good g is estimated. The coefficient and SE in Panel B are therefore the average of all 30 estimated α and the respective SE of the average. v is computed from trader level markups (using purchasing costs only) or from trader level profit margins (using profit data). The data used is the updated baseline and Sauti's price data (for the Simple Mean Matching method) while the GMM estimations restrict the sample to agriculture goods for more consistent prices per kg.

Table 1.10: Counterfactual Model Simulations: Closure of trade routes

Panel A: Choice Probabilities: Substitution Patterns

	Domestic	**Exit**	**Formal**	**Informal**
Initial Model (4 routes)	0.769	0.129	0.038	0.063
	(0.012)	(0.010)	(0.041)	(0.008)
No Formal	0.8	0.133	0	0.066
	(0.013)	(0.010)	(0.000)	(0.008)
No Formal, No Informal	0.857	0.142	0	0
	(0.011)	(0.011)	(0.000)	(0.000)

Panel B: Counterfactual Simulations: Model Predictions on Prices

	Mean
Initial quantity by trader (Kg)	131
Initial $p_{mtk}(Kshs)$	114
No Formal Route	
Total Quantity exited (Kg)	554
Quantity exited by market (Kg)	41
New quantity by trader (Kg)	90
New p_{mtk} (Kshs)	122
Δp_{mtk}	7.5%
No Informal Route, No Formal Route	
Total Quantity exited (Kg)	1183
Quantity exited by market (Kg)	82
New quantity by trader (Kg)	49
New p_{mtk} (Kshs)	136
Δp_{mtk}	19%

This table shows the results of the counterfacutal analysis. Panel A shows how the shares in route choice vary in each scenario. Taking these shares into consideration, Panel B shows how aggregate market prices would change in each scenario.

Figures

Figure 1.1: Data Collection and Intervention Timeline

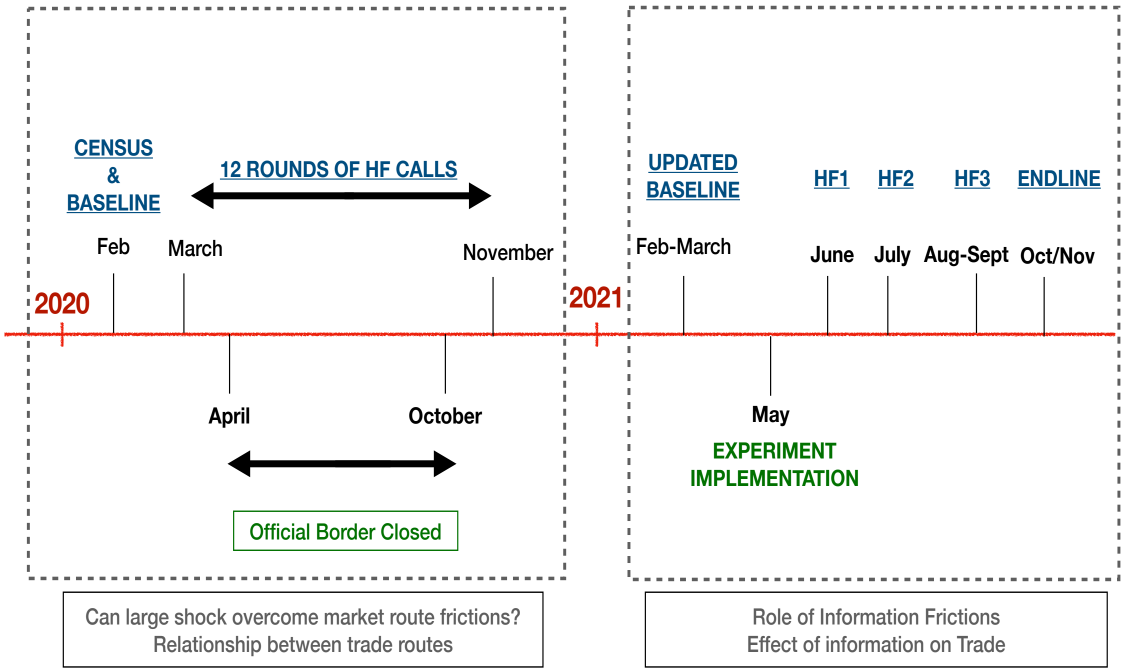


Figure 1.2: Market prices across time

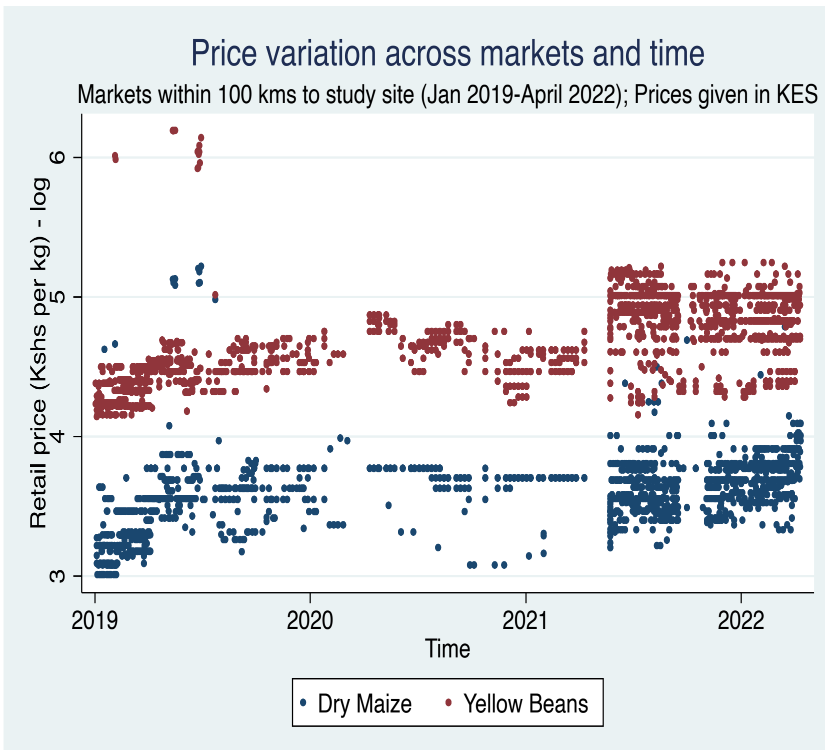
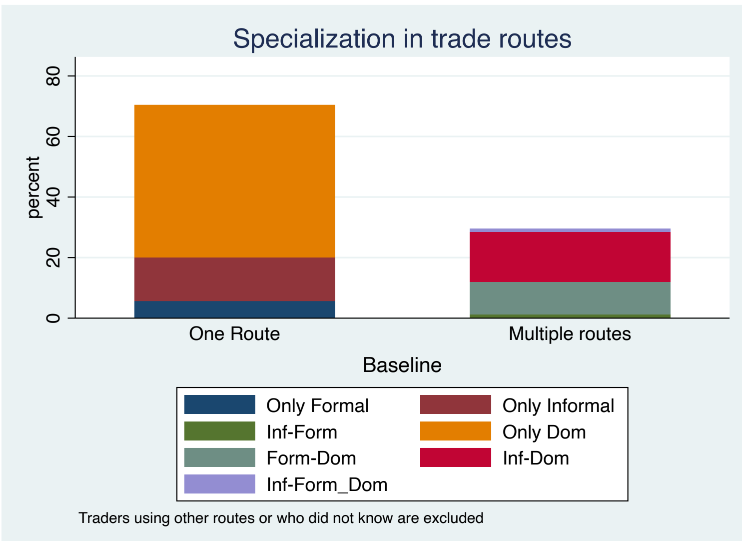


Figure 1.3: Market and Route Specialization



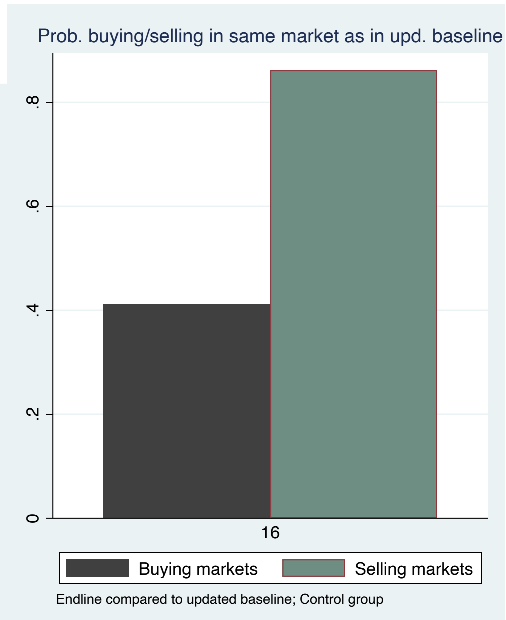
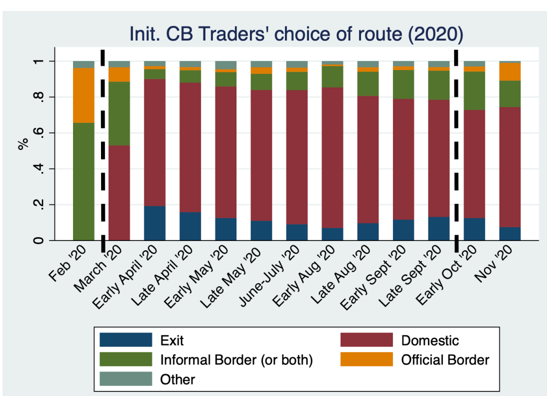


Figure 1.4: Border Closure effect on route choice - CB and Domestic traders



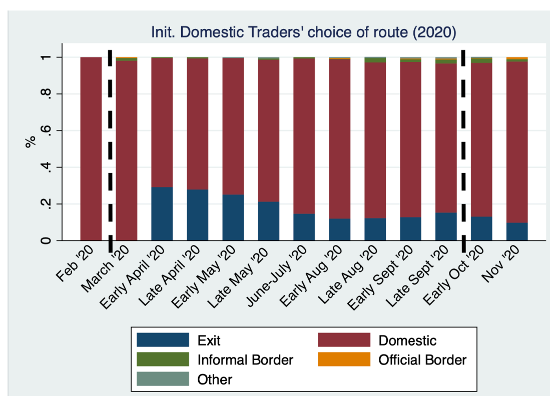


Figure 1.5: Route choice when border re-opens

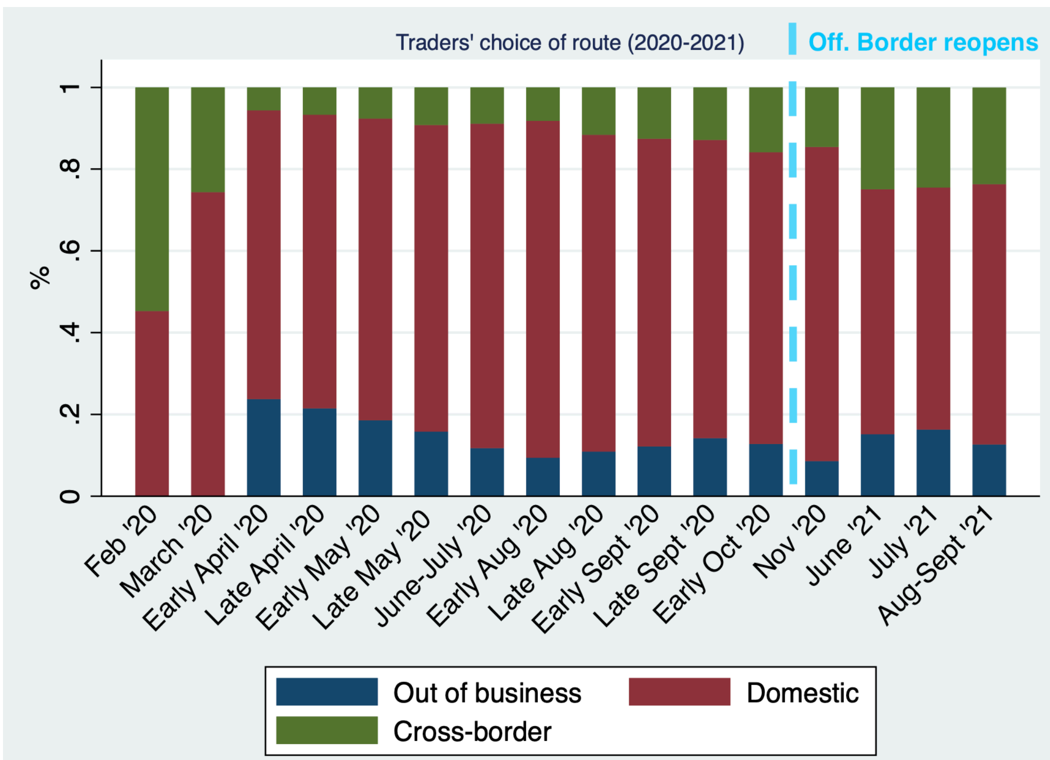


Figure 1.6: Overview of Model

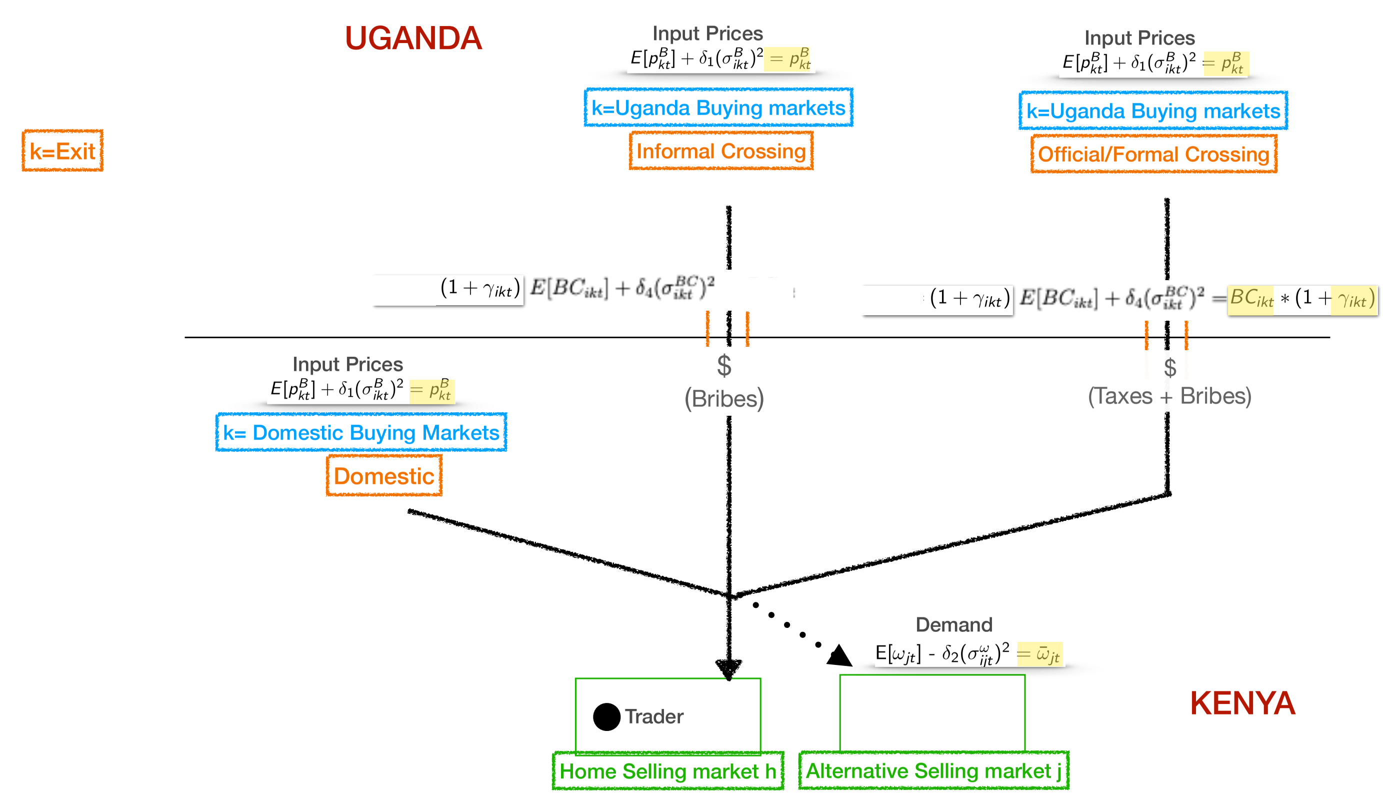


Figure 1.7: First Stage: Distribution of number of months of usage (Treatment Group)

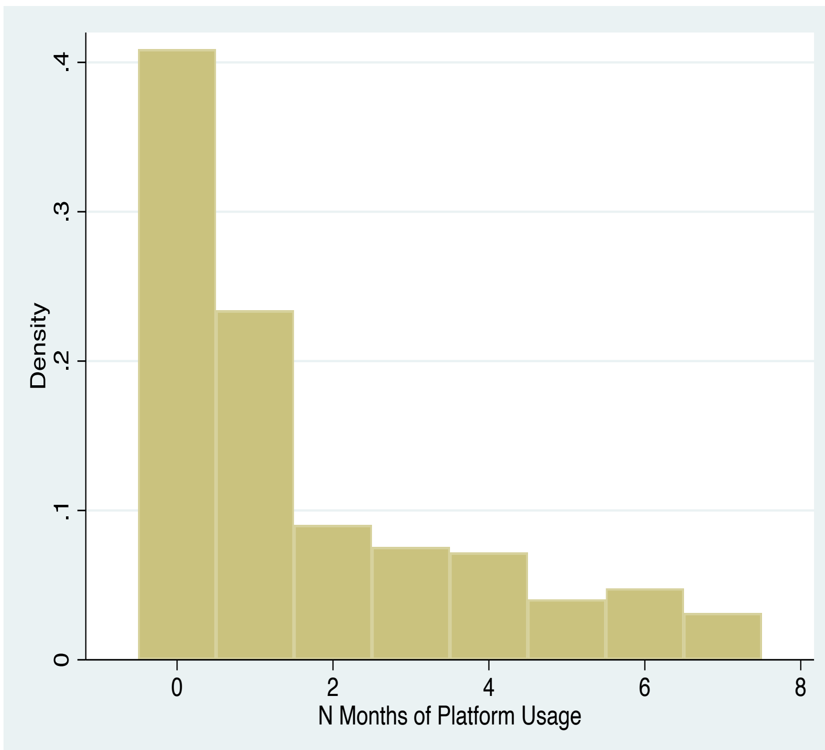


Figure 1.8: First Stage: N users, sessions and sessions per user (Treatment Group, conditional on usage)

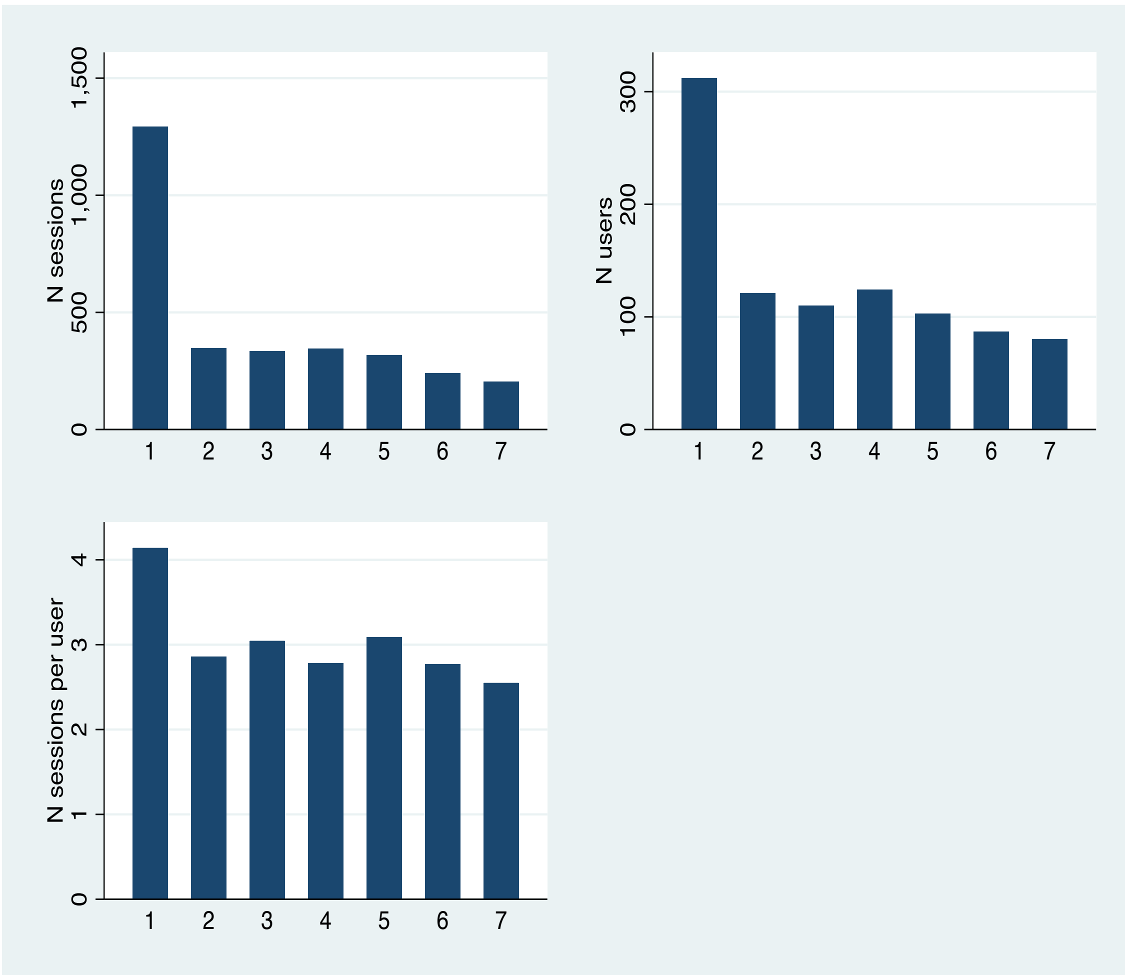
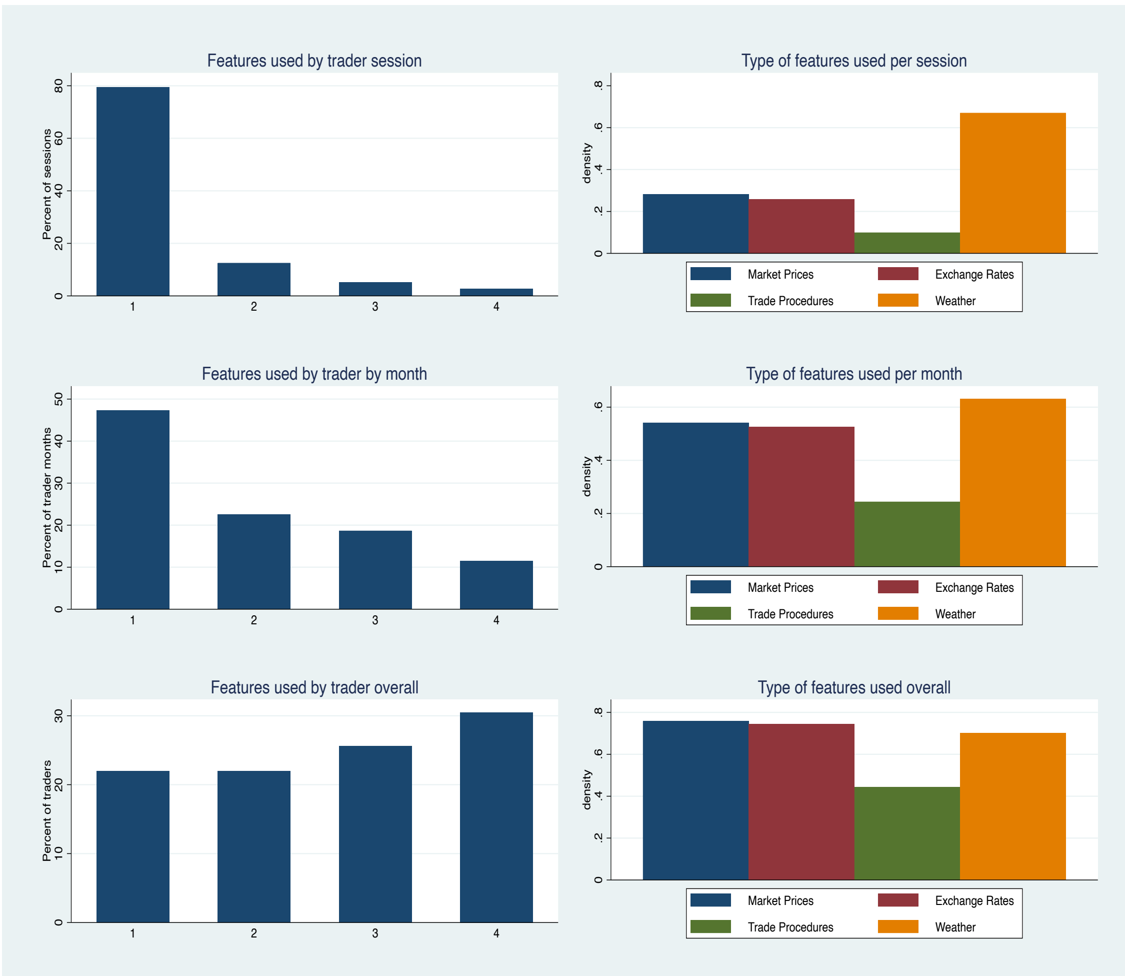


Figure 1.9: First Stage: Type of features requested (Treatment Group, conditional on usage)



Chapter 2

Private Input Suppliers as Information Agents for Technology Adoption in Agriculture

This chapter is coauthored with Manzoor H. Dar, Alain de Janvry, Kyle Emerick and Elisabeth Sadoulet

2.1 Introduction

Many people in poor countries rely on agriculture for their livelihoods. But they often use traditional technologies, despite the existence of more productive alternatives. This puzzle has sparked extensive research to understand which barriers constrain technology adoption. Researchers have focused on credit and insurance market failures and on information frictions (Feder et al., 1985; de Janvry et al., 2017; Magruder, 2018). Decades of research shows that learning plays an important role for adoption (Griliches, 1957; Conley and Udry, 2010; Fabregas et al., 2019; Gupta et al., 2020; Cole and Fernando, 2021).

Recognizing this, the public sector invests in agricultural extension — a process where government workers communicate information to selected contact farmers. This initial communication might trigger the flow of information through social networks. But the government system of outreach has not met expectations (Farrington, 1995; Anderson and Feder, 2007). Outreach via the private sector may be one way to improve the flow of information to farmers, who rely on commercial input dealers for advice.[1] In India, farmers rely on

[1] These dealers are in turn informed by the companies whose products they sell (Fites, 1996) and by public agricultural agencies and research institutions (Wolf et al., 2001).

input dealers for advice almost as often as they rely on their peers.[2] Motivated by profits, input dealers have incentives to provide information. As such, reaching farmers through the private sector could induce technology adoption.

We contrast two ways of disseminating information using a randomized experiment in Odisha India. The main intervention provides information and seeds for testing to private input suppliers. We compare this approach with a control arm that replicates business-as-usual government extension. Many countries have piloted various forms of private agricultural extension (Rivera and Alex, 2004). There is limited evidence on these reforms. We contribute by showing how information outreach can be more effective when delivered via the private sector, compared to the government channel.

Involving private-sector agents in providing public services can have ambiguous effects (Hart et al., 1997). Input suppliers have motives to spread information about technologies they sell for profit. Repeated interactions with their clients may discipline them to provide high-quality recommendations.[3] On the other hand, agrodealers might provide lower quality when it is hard to verify (Bold et al., 2017; Ashour et al., 2019). Or they may recommend products that maximize their own profits instead of customer welfare, as researchers have found in other sectors (Hubbard, 1998; Inderst and Ottaviani, 2009; Mullainathan et al., 2012; Chen et al., 2016; Anagol et al., 2017). Given these concerns, we seek to answer both whether informed agrodealers increase adoption and whether they do so for the farmers that stand to gain the most from the innovation.

Our at-scale experiment took place in 10 coastal districts of the Indian state of Odisha. Our sample consists of 72 blocks, covering an area with about 1.7 million farmers.[4] We consider the dissemination of Swarna-Sub1, a new and profitable flood-tolerant rice variety.[5] We partnered with the government extension service to support their conventional activities in 36 control blocks. This partnership included providing Swarna-Sub1 seed minikits to the contact farmers on whom they rely to spread information. It also involved carrying out large-scale "cluster" demonstrations where many farmers grow the new variety on contiguous plots of land, and organizing farmer field days to share results from demonstrations. These

[2]The 2018-2019 National Sample Survey finds that 20% of agricultural households rely on input dealers for technical advice (NSSO, 2021). This is only slightly lower than 23% that rely on other farmers and over 6 times the 3.1 percent that rely on extension agents.

[3]A potential downside of this approach is that markets for agricultural inputs in developing countries can be sparse. Aggarwal et al. (2018) show that travel costs to input suppliers play an important role in technology adoption for African smallholders. If these costs are too high, then few farmers will have contact with dealers and solving information frictions on the supply side of technology could be less effective.

[4]Blocks are the relevant administrative units for agricultural extension. Blocks in the experiment have an average of 136 villages, and each block has its own local agricultural extension office.

[5]Previous work shows that this innovation is profitable for farmers. By reducing risk, it induces them to invest more in early-season inputs. Notably, it has no yield penalty in normal years (Emerick et al., 2016).

are all activities the government extension service would do with adequate resources, but we supported them as part of the experiment to make sure that they were carried out to the full extent and that our control group reflects business-as-usual activities at their best.

We provided the exact same quantity of seeds and the same information to input dealers in the 36 treatment blocks. We did not support any conventional extension activities as was done in the control blocks. These dealers are highly local, small-scale businesses, selling seeds and often other inputs such as agro-chemicals. They were free to choose how to use the demonstration seeds. The key distinction between this treatment and the standard mode of agricultural extension (the control) is that information constraints are being relaxed on the supply side with private agrodealers, rather than on the demand side with farmers. The treatment tries to leverage the economic incentives created by the marketplace for private businesses in transmitting information to their clients. Dealers in our sample receive the same profit margin for Swarna-Sub1 as they do for other seeds. Thus, increasing quantities sold, both currently and in future seasons, is the motive for dealers to promote new seeds to farmers.[6]

Turning to results one year later, we find that the dealer-based treatment increases adoption of Swarna-Sub1 — the variety being introduced — by over 56 percent, i.e. from 6.3 to 9.8 percent of farmers. Furthermore, the average farmer in the treatment blocks cultivates 69 percent more land with the variety.

Consistent with these farm-level results, the treatment triggered a supply-side response on the seed market. By the 2018 season, two years after we introduced the seeds, dealers in treatment blocks were about 59 percent more likely to have them in stock. We find some evidence that informing agrodealers causes a change in local seed production. Treatment blocks produced 40-50 percent more Swarna-Sub1 seeds three years after the intervention.

Are farmers induced to adopt in the dealer treatment those with higher expected returns? To consider this, we look at heterogeneity according to past flood exposure — an important determinant of returns.[7] The technology provides higher yields when crops are flooded relative to other types of rice grown by farmers. It leaves yields unchanged when there is no flooding. We find that the dealer treatment only increases adoption for farmers that are the most exposed to flooding. The treatment more than doubled adoption for the highest risk farmers. It left adoption unchanged for lower risk farmers. This finding suggests that dealers may consider the benefits to farmers when making suggestions.

[6] Dealers in our setting principally sell seeds that are produced by the state-run seed corporation that fixes both wholesale and retail prices equally for all seed varieties.

[7] Flood tolerance is the key attribute of Swarna-Sub1. The technology provides higher yields when crops are flooded relative to other types of rice grown by farmers (Xu et al., 2006). It leaves yields unchanged when there is no flooding.

Are these results short-lived? The effects of information interventions in agriculture sometimes die out over time (Casaburi et al., 2019). Our main results come from a farm-level survey one year after the intervention. But the data we assembled on dealer-level supply and seed production cover up to four years after the intervention. We find no evidence that effects on these outcomes have declined over time. The persistent effects are consistent with the dealer treatment leading to adoption by farmers with high expected returns. We would have expected rapid disadoption if dealers push the technology to farmers with low returns. Dealers increase profit by market size. This creates a powerful incentive to seek profit via long-term consumer satisfaction.

Turning to mechanisms, our treatment was designed to reduce a particular learning friction. Namely, high-powered incentives do not exist in the traditional model of agricultural extension. Agricultural extension agents in that model rely on non-incentivized communication between selected contact farmers and farmers in their social networks. By contrast, providing new information to profit-motivated agrodealers results in them sharing this information with farmers.

The rest of our analysis shows supporting evidence for this mechanism where profit-motivated dealers give information to farmers. We show two separate pieces of evidence: one for information transmission and another for profit motives. Starting with information transmission, we sent "secret shoppers" to about 300 dealers. The purpose of these visits was to inquire about new rice varieties. This took place in the third season of the study — two years after the intervention. We find that the treatment changes what dealers say to potential customers. Dealers in treatment blocks are about 25 percent more likely to mention Swarna-Sub1 when listing the new varieties to consider. When asked for a specific recommendation, dealers in treatment blocks recommend older types of seeds at lowers rates. In some cases, they are more likely to recommend trying Swarna-Sub1.

We then ran an experiment to test whether profit motives can explain why dealers communicate information that leads to greater adoption. In partnership with a local NGO, we revisited dealers in all blocks during the fourth season (2019). Each dealer was randomized into one of two treatments. In the first treatment, someone visited the dealer and asked which farmers, locations, and varieties would be best for a demonstration where farmers would cultivate a new variety and then the NGO would organize a meeting with other villagers to explain its attributes. Importantly, the name of the dealer giving the recommendation would be advertised during the meeting. In the second treatment, the same demonstration was described, but the NGO would not name the dealer and would collect the harvest after the demonstration and redistribute it as seeds to other farmers.

The first treatment presents a clear profit opportunity to the dealer. Making the dealer's participation known can direct any demand created from the demonstration to them. The second treatment reduces these incentives by not broadcasting the dealer's participation. It also redistributes seeds, which would reduce demand through the demonstration.

The treatment highlighting profit opportunities changes the advice given by dealers. It causes them to suggest different locations, types of farmers, and seed varieties. Starting with location, when presented with a candidate list of villages for the program, dealers in the profit motive treatment are more likely to suggest a village outside that list. They often suggest their own village. Dealers in this treatment also spend more time thinking of farmers to recommend. They are more likely to suggest neighbors or other people in their own village. Finally, the treatment causes dealers to recommend less common seed varieties.

Taken together, these findings are suggestive of a mechanism where our treatment first informs dealers. These dealers are then motivated by profit to pass this information along to farmers. But we cannot rule out other mechanisms. In particular, the treatment provided dealers with demonstration seeds. They could have passed those seeds to better connected farmers, who then shared the information. Additionally, dealers are usually farmers themselves. They may also be better connected, leading to more information transmission independent of their role as dealers.

We contribute by providing evidence on how private-sector agrodealers can improve the delivery of information to farmers. Studies focus on improving the government system of spreading information. For instance, GIS monitoring of government agricultural workers increases their effort (Dal Bó et al., 2020). Strategic selection of contact farmers and providing them financial incentives increases adoption (Beaman and Dillon, 2018; BenYishay et al., 2020; BenYishay and Mobarak, 2018; Beaman et al., 2021). Training the government-selected contact farmers has less of an impact (Kondylis et al., 2017). One common theme across these studies is the focus on government outreach to farmers through extension agents and contact farmers. We take a different approach by considering whether agricultural extension can leverage the private sector. Input suppliers have rarely been considered as information agents in agricultural markets.[8] One related non-experimental study in Niger finds that setting up demonstration plots with agrodealers — as opposed to no demonstration plots at all — increases adoption of new seeds by farmers (Mamadou et al., 2019).

More broadly, the private sector delivers public services in developing countries, including giving medical advice (Das et al., 2016; Kwan et al., 2022), distributing subsidized food (Banerjee et al., 2019), and providing services such as water and education (Galiani et al., 2005; Romero et al., 2020). In agriculture, full outsourcing of extension to the private sector has met challenges (Anderson and Feder, 2004; Rivera and Alex, 2004). The approach has proved much more complex to implement than expected, subject to collusion, and too

[8]In the context of microcredit, Maitra et al. (2020) show that agricultural loans generate more benefits when private traders select recipients — rather than letting them self-select into group loans — perhaps due to incentive effects where private traders benefit themselves when loans cause farmers to harvest more output.

expensive for farmers. Moreover, the fee-for-service approach only works when the advice is a market good, i.e. customized information that is excludable. Our intervention differs from full outsourcing. In our case, there is no contracting between public and private entities.[9] Instead, we show how providing better information to private agrodealers can outperform the provision of information through government channels.

The rest of this paper is organized as follows. Section 2.2 gives more information on the setting and outlines the experiment. Section 2.3 describes the data collection. Section 2.4 presents the main results on how assisting input dealers to learn increases technology adoption by farmers, particularly those with the highest potential benefit. Section 2.5 shows evidence that dealers spread information to their customers and that profits motivate them to do so. Section 2.6 concludes.

2.2 Background and Design of Main Experiment

This section starts by providing background information on the standard methods used in agricultural extension. It also gives a description of how the public sector delivers information to farmers in our particular study area. We then outline the design of our main experiment to compare these standard methods with the more business-oriented approach of using agrodealers as information agents. The section concludes with a discussion of profits for agrodealers and how they can only benefit from recommending a particular seed if it will increase their aggregate quantity sold, either currently or in the future.

2.2.1 Public-Sector Agricultural Extension

Governments all over the world support agricultural extension services as a mode of information delivery. Ministries of agriculture typically have entire departments dedicated to providing these services. These departments oversee local offices that hire frontline extension agents whose role is to diffuse information about new agricultural technologies and practices to farmers. The specific techniques used by agents vary across contexts, but the basic methods are largely consistent, especially in poor countries. Agents usually work with selected "contact farmers" who are keen on trying new approaches and are best able to transmit knowledge to others in their social networks. They also organize farmer field days with cluster demonstration plots, where new technologies are implemented by multiple farmers, to boost the diffusion of information.

[9] Levin and Tadelis (2010) document how U.S municipalities are less likely to privatize services when performance measurement is difficult or there are holdup concerns. These issues are absent in our setting since we are not testing full outsourcing of agricultural extension.

In the context of our experiment, agricultural extension workers use these standard techniques. Each of the 10 districts in the sample is organized into blocks, where a block has an average of 136 villages. Each block has an agricultural office that is led by a Block Agricultural Officer (BAO). The BAO employs Assistant Agricultural Officers (AAO) and Village Agricultural Workers (VAW) who work in the field with farmers.

2.2.2 Experiment on Dealer-Based Extension

Our sample consists of 72 blocks in 10 flood-prone districts of Odisha.[10] We selected these areas because the promoted technology — a flood-tolerant rice variety called Swarna-Sub1 — is most suitable for flood-affected areas.[11] The blocks in the sample represent around 20 percent of the blocks in the state.

We randomly assigned 36 of these blocks to the treatment group where agrodealers were targeted to receive seeds and information. This randomization was stratified by district. The remaining 36 blocks serve as a comparison group where we supported the government extension service to carry out normal extension activities.

Figure C.1.1 displays the timeline of these interventions. Starting in May 2016 — about 6-8 weeks before planting time — we partnered with the government's extension service to introduce Swarna-Sub1 into control blocks. We did this in a way that mirrors three common practices in agricultural extension. First, field staff provided 10 seed minikits of 5 kilograms each to the BAO, who then helped identify contact farmers to use the kits. The BAO chose 2 villages and 5 farmers in each village. Each kit contained only seeds for testing and some basic information about Swarna-Sub1. Our field staff then delivered the kits to the recommended farmers. Second, we provided another 150 kg of seeds to the BAO so that he could set up a cluster demonstration where the seeds would be used by several farmers on a contiguous set of plots. Based on seeding rates in the region, 150 kg allows for cultivation of 5-10 acres. The BAO chose where to do the demonstration and which farmers to target. Official government guidelines for organizing these clusters suggest that they be carried out in sites that are easily accessible to be viewed by many farmers. Moreover, sites should be representative of average conditions in the area. Third, we helped the BAO carry out a farmer field day in November — at the time right before harvest. The BAO selected the location of the field day and whom to invite. The purpose of the field day was for extension staff to train farmers about Swarna-Sub1 and share information from the demonstrations.

[10] The districts are Bhadrak, Balasore, Cuttack, Ganjam, Kendrapara, Khorda, Jagatsinghpur, Jajpur, Nayagarh, and Puri.

[11] The technology has been shown to benefit farmers by reducing both yield losses when flooding occurs and therefore downside risk in any year, thereby increasing investment (Dar et al., 2013; Emerick et al., 2016).

The objective of such an active control group is twofold. First, it ensures that each block is endowed with the same quantity of seeds. Therefore, the dealer-based treatment only differs on *who* received the new seeds and information. Second, the demonstrations and partnerships with contact farmers may not have taken place without our involvement. Forcing these activities to happen makes the treatment-control comparison more meaningful. Most importantly, it sets a higher bar for the dealer-based treatment by eliminating any possibility that the new technology would not be promoted by the government extension service.

Turning to the 36 treatment blocks, we obtained a list of 2,087 seed suppliers from the state Department of Agriculture. These include suppliers of two types: private seed dealers and Primary Agricultural Cooperative Societies (PACS). PACS are farmer groups that handle credit, seed supply, and procurement of output for farmers. We did not include them in the intervention because their incentives are not the same as those of private dealers. Seed sales are usually handled by a member that is not the residual claimant on any profits from the sale. Despite being fewer in number relative to PACS, private dealers account for almost 60 percent of the seeds sold to farmers. The sample consists of 666 private dealers, 327 of which were located in the treatment blocks.

Armed with this list, our field staff entered each treatment block and located five dealers interested in receiving seed minikits and an informational pamphlet about Swarna-Sub1. In some blocks fewer than 5 dealers were available. We provided additional seed to each dealer in these cases to guarantee that a full 200 kilograms (the same amount as control blocks) were introduced. The list provided by the Department of Agriculture did not have enough locatable dealers in some cases. In these circumstances, our field staff provided the seeds to other local agrodealers.[12] Overall, seeds and information were provided to 151 dealers across the 36 treatment blocks.[13] 119 of these were from the original list.

Once provided with seeds and information, the dealers were left alone to decide how to use them. We asked dealers about their intended uses. They overwhelmingly stated that they would use the seeds for testing on their own farms and would provide them to good customers for testing.[14] Our intervention did not include any additional assistance to dealers. This approach differs from standard methods in agricultural extension where agents

[12]The list of 666 dealers includes only those that are registered with the state, a prerequisite for selling seeds from the state seed corporation. The dealers not included in our list could have been in the process of renewing their license or only selling seeds produced by private companies.

[13]Two dealers in one control block were provided seeds by mistake. All our analysis uses only the original random treatment assignment.

[14]Around 83 percent of dealers indicated they would try some of the seeds on their own, while 63 percent indicated that some of the seeds would be provided to their good customers. Other less common responses were to provide them to family members (9 percent) and friends (24 percent).

continually revisit their contact farmers. We allowed dealers to learn on their own because, in theory, they should be motivated to learn about a new product that could enhance their business. The goal of our treatment is to measure whether this motivation causes information to flow to farmers and ultimately increases adoption. Not intervening further ensures that our treatment effect is driven by any real-world incentives dealers have to learn, rather than heavy monitoring by our partners.

We tracked these activities in both arms of the experiment. For control blocks, we took GPS coordinates of farmers chosen by the BAO for minikits. We collected the names and villages of the demonstration farmers from the BAO. Lastly, one of our team members attended 29 of the 36 field days and took GPS coordinates. In treatment blocks we took photographs and GPS coordinates during seed delivery to dealers. Table B.1.1 uses these GPS coordinates to show how proximity to the different extension sources varies by treatment. In particular, farmers in control blocks are 11.7 kilometers from the nearest treated dealer. This distance falls to 4.36 kilometers for treatment farmers.

Dealers in our sample are small business entrepreneurs. Some operate out of their homes, while others maintain shops in rural towns. 44 percent of dealers sell only seeds, with fertilizers and pesticides being the most common inputs sold by the other dealers. They are highly local. The average block in our study has 540 hectares of rice area per seed supplier.[15] The median dealer in our data sells enough rice seed to cover roughly 162 ha, which implies that about 30 percent of area is planted with new seed each year. Farmers use seeds from their previous harvest for the remaining area. This creates an opportunity for dealers to expand sales by getting farmers to buy new seeds rather than use ones from the prior year.

Turning to the second season (2017), we ran an SMS messaging experiment to compare our intervention with this "lighter touch" information treatment. The random delivery of SMS messages allows us to test whether our dealer treatment substitutes (or complements) basic knowledge that can be easily transmitted via ICT technology. Furthermore, it allows us to compare the direct effects of the two approaches.

The message informed farmers that Swarna-Sub1 is a new variety that is suitable for medium-low land in terms of elevation, matures in 145 days, and can tolerate up to two weeks of flooding. The message also stated that it was being produced by OSSC and could be available at local dealers. As a sampling frame, we obtained mobile numbers for 75,616 farmers that had registered for the state government's Direct Benefit Transfer (DBT) scheme to obtain seed subsidies.[16] These farmers are located across the 261 gram panchayats (an

[15]We arrived at this estimate by taking the total number of dealers and cooperatives in each block that are registered with the state seed corporation. Block-level rice area from 2016-2017 was available from the Odisha Directorate of Economics and Statistics.

[16]Beginning in 2016 the state government started providing seed subsidies in the form of payments back to farmers. Farmers were required to register, provide bank account details, and pay the full price at the

administrative unit usually consisting of around eight villages) that cover our main estimation sample, as outlined below. The SMS treatment was randomized at the gram panchayat level, resulting in messages being delivered to 37,783 of the names on the list.

2.2.3 Motivation of dealers to recommend new seeds

Dealers can be motivated to recommend a new seed either because it has a higher profit margin, or because it will increase their quantity sold. The profit margin explanation cannot explain dealer behavior in our context. Around 84% of the seeds sold by dealers in our sample are produced by the state-run Odisha State Seed Corporation (OSSC). As licensed agents of this company, dealers pay the same wholesale price for all rice varieties. State regulation fixes equal retail prices across varieties. Thus, the margin for dealers is identical across all types of varieties. This model, where state seed corporations play a major role on seed production and retailing, is common throughout India.[17]

As a result of these fixed markups, it is optimal for dealers to invest effort in recommending varieties that will increase their aggregate sales. Convincing farmers to shift from buying one variety to another is not profitable for the dealer because of the equal margins. But getting farmers to purchase a new variety, instead of using their own seeds of an older variety, represents an increase in business. Put differently, dealers profit from recommending a new variety to promote seed replacement.

But dealer's incentives for providing advice go beyond encouraging seed replacement during a single season. Dealers sell seeds year after year. In this dynamic perspective, they benefit from making a recommendation today if it increases the size of their future business. This benefit can arise either because the same satisfied customers return, or because farmers communicate to others that they learned about a profitable new variety from the dealer. Providing good advice today is a way for dealers to increase the quantities they sell during future seasons.

These features of the setting suggest that incentives are in place for dealers to promote improved seed varieties. The incentives operate through increasing current or future quantities, not through differential profit margins. Similar incentives do not exist for the public agricultural extension system. Government extension workers are not paid for performance, which may lower effort (Dal Bó et al., 2020). Even when government workers successfully influence contact farmers, these farmers might not spread information without incentives for themselves (BenYishay and Mobarak, 2018). Expanding dealers' knowledge about new

time of seed purchase. The subsidy was then credited to their bank account after the transaction details had been entered into a mobile phone app by the seed dealer.

[17]For example, 80 percent of rice seeds in the state of Andhra Pradesh are purchased from retail outlets of the Andhra Pradesh State Seeds Development Corporation.

seeds is thus meant to overcome the incentive problems that exist in traditional government extension.

2.3 Data Collection for Main Experiment

This section describes the experimental data for testing whether intervening with agrodealers increases adoption by farmers. It also discusses the satellite data used to test whether adoption effects are larger for farmers with higher expected benefits from using the technology. We focus only on the data from the first two years of the study. We save the discussion on the additional data and the experiment on mechanisms for Section 2.5.

2.3.1 Survey on farmer technology adoption

We anticipated that dealers and contact farmers would use the demonstration minikits for learning in 2016 and any possible treatment effects could first be detected during year two (the 2017 season). Our main followup survey therefore took place in August-September 2017 — around 15 months after the interventions. Its purpose was to measure adoption of seed varieties by rice farmers. To minimize measurement error, we timed the survey to be right after planting.

Our sample consists of 7,200 farmers. These farmers were drawn from a random sample of 261 gram panchayats.[18] Before drawing this sample, we excluded gram panchayats that had any village within 1.5 kilometers of the block boundary.[19] We removed these areas to reduce any interference caused by farmers possibly obtaining seeds from other blocks. The 261 sample gram panchayats had 75,616 farmers registered in the DBT program for seed subsidies. Using this database as a sampling frame, we randomly drew 100 farmers from each block (amongst the sample gram panchayats). These farmers are spread across 1,333 villages.[20] Figure B.1.1 shows their geographic dispersion across the 10 districts in the experiment.

Survey teams succeeded in locating and surveying 6,653 (92 percent) of the farmers. Of these, 93 percent were currently cultivating rice. Table B.1.2 shows no significant differences in the probabilities of being surveyed or growing rice between treatment and control groups.

[18]We limited our data collection to a sample of gram panchayats to lower transportation costs for survey teams. The gram panchayats were identified using the 2011 Population Census of India (Asher and Novosad, 2019).

[19]Approximately 17.5 percent of the villages across the 72 blocks are within 1.5 km of another village in a different block.

[20]The farmer survey has almost no overlap with the dealers from treatment blocks. Using phone numbers of the the treated dealers, we found only one of them in the farmer sample.

The survey focused on which seed varieties were currently being used for rice cultivation. Surveyors went through a list of 30 varieties and asked farmers which ones they were currently using and the amount of land being grown.[21] In addition to these adoption data, we obtained information on contacts with agricultural extension agents during the last year, topics discussed during these conversations, whether the farmer had seen any seed demonstrations, and whether they had recently learned about Swarna-Sub1.

2.3.2 Data on supply responses

Any treatment effects on farmer-level uptake might occur simultaneously with supply responses by dealers.[22] To measure this, we surveyed seed dealers around the same time as the farmer survey. We timed the survey to be in September so that seed purchases would be recently completed and easier to recall for dealers. Dealers were asked which varieties they carried for the 2017 season, how much of each was sold, and whether they were selling seeds from private companies or from the state's seed corporation.

Our sample consists of 613 dealers from the list of dealers obtained prior to the experiment.[23] A large fraction could not be located or were no longer selling rice seeds. Specifically, 22.8 percent of them could not be reached. Of the 473 dealers located, 274 (58 percent) were selling rice seeds in the 2017 season. In results that follow, we show effects both for all dealers that were reached and those that remained in the seed business. Table B.1.3 shows that the likelihood of being located and the probably of selling rice seeds during the 2017 season are uncorrelated with treatment. Focusing on the treatment blocks, about 42 percent of the dealers surveyed received the intervention.

In addition to these dealer sales, we obtained data on the physical location of seed production. Seeds are grown by registered farmers that contract with the state to produce seeds that meet minimum certification standards. OSSC then collects, processes, and bags these seeds before selling them to farmers (via dealers and cooperatives) during the next season. The average block in our study had 32 seed growers per season from 2014 to 2019. We use records from a publicly available database that gives the location of each seed grower, the contracted area, the variety they produced, and the amount that was collected and processed (Odisha State Seed and Organic Products Certification Agency, 2020).

[21] Swarna-Sub1 — the variety introduced in the treatments — was 24th in this ordering. Asking about uptake in this way makes it less likely that responses reflect experimenter demands. Furthermore, farmers surveyed were not informed about the interventions that were carried out in their block a year earlier.

[22] This need not be the case if there was already excess capacity of Swarna-Sub1 seeds or if farmers obtain seeds outside of formal markets, such as from friends or relatives.

[23] There were 53 dealers on our list of 666 that had no contact details and thus we did not attempt to locate them.

Seed growers tend to be large farmers. They have incentives to produce the most profitable varieties for their land — just like other farmers.[24] As such, their production of a new variety depends on them being convinced of its potential. We therefore aggregate seed production at the block-season level and estimate the effect of the dealer treatment on the amount of Swarna-Sub1 produced in the block.

2.3.3 Flooding exposure for individual farmers

Returning to farmer-level information, we use remote sensing data to approximate flooding risk. These data help us predict which farmers are expected to benefit the most from Swarna-Sub1. Being able to observe a key determinant of returns makes it possible to test for heterogeneous treatment effects according to a proxy for predicted benefits. More simply, is there a tradeoff between intervening with private-sector agents and a technology reaching the right people? Or, does involving input suppliers in the diffusion of information cause technology to diffuse to high-return individuals?

We have GPS coordinates of the houses for 83 percent of the farmers that we surveyed in 2017.[25] These coordinates are matched to daily images of flooded areas from June to October for the period 2011 to 2017. We consider a household as exposed to flooding on a given day if its house is within one kilometer of any flooded area.[26] We then aggregate the total number of days of flood exposure across the 7 years as a measure of flooding risk — and hence as a proxy for the return to Swarna-Sub1.

The online appendix shows three characteristics of this variable. First, it varies substantially across the sample (Figure B.1.3). About 30 percent of households were not exposed to flooding. In contrast, 10 percent of households had flooding for 40 days or more. Second, this variation is partly driven by geographic characteristics. Particularly, Figure B.1.4 shows that flooding is more frequent in lower-elevation areas that are closer to rivers. These correlations provide verification that our measure at least partly reflects underlying determinants

[24]The contracts with OSSC are on an acreage basis. OSSC and the grower agree on the variety and OSSC purchases the output at a pre-determined price. The grower pays for all the inputs.

[25]The likelihood of missing GPS coordinates is uncorrelated with treatment. A regression of observing GPS coordinates on treatment has a coefficient estimate of -.018 and a t-statistic of 0.4.

[26]A different study in one of the same districts collected GPS coordinates of both houses and rice plots (Emerick and Dar, 2021, 2020). These data show that rice plots are within one kilometer of the household almost 90 percent of the time (Figure B.1.2). The images of flooding extent are processed from MODIS by the DFO Flood Observatory (floodobservatory.colorado.edu). Each image has a spatial resolution of 250m. A pixel is classified as flooded on days when a ratio of the bands detecting surface water to land exceeds a numeric threshold. Using the GIS coordinates of each household, we calculate the distance between the household and the nearest flooded pixel for each day during June-October for 2011-2017.

of flooding risk — not just recent flood shocks. Third, farmers exposed to more flooding tend to be smaller, poorer, and belong to low-caste social groups (Table B.1.4).

2.3.4 Descriptive Statistics

Table 2.1 shows descriptive statistics and verifies randomization balance. Panel A shows block-level characteristics, derived mostly from the 2011 Census. Most notably, the blocks have around 136 villages and an average population of 110,000. Beyond these, treatment and control blocks look similar on a number of other characteristics, including local Swarna-Sub1 seed production, caste distribution of the population, and elevation.

Panel B shows characteristics of the respondents from our 2017 survey. These characteristics were collected after the treatment, but are time invariant. Observables are mostly balanced for this sample that we use to estimate our main regressions.

2.4 Results of Dealer Extension Experiment

This section presents the results of the agrodealer experiment. After outlining the estimation strategy in Section 2.4.1, Section 2.4.2 shows that using dealers as information agents increases adoption by farmers. This finding is robust to different ways of measuring adoption and to including a battery of control variables. Section 2.4.3 tests whether this treatment effect varies by exposure to flooding risk — which is highly correlated with expected returns. We turn to effects on the supply side in Section 2.4.4. Particularly, we show effects on both dealer-level seed inventories and block-level production of Swarna-Sub1 seeds.

2.4.1 Estimation

Our main analysis consists of farmer- or dealer-level regressions of outcomes on the block-level treatment indicator:

$$y_{ibd} = \beta * Treatment_{bd} + \alpha_d + \varepsilon_{ibd}, \tag{2.1}$$

where i indexes farmers (or dealers), b indexes blocks, and d indexes districts. We include district fixed effects in all specifications because the treatment was stratified by districts. We cluster all standard errors by block. The analysis uses only the random variation we generated, but the online appendix shows that our results are robust to controlling for the covariates in Table 2.1.

2.4.2 Effect on Technology Adoption

Informing private input dealers and providing them with seeds to test leads to greater adoption by farmers when compared to conventional extension approaches used by the public sector. Table 2.2 shows this result. Each column gives treatment effects with clustered standard errors (parentheses) and p-values calculated by randomization inference (brackets).[27] Starting with Column 1, farmers in treatment blocks are 3.5 percentage points more likely to adopt Swarna-Sub1 a year after the treatment, compared to farmers in control blocks. Given an adoption rate of 6.3% in the control group, this implies the treatment leads to a 56% increase in uptake. Columns 2 and 3 add controls for pre-treatment covariates. Column 2 includes all the controls from Table 2.1, while Column 3 uses the post-double selection procedure in Belloni et al. (2014) to select controls that predict either the outcome or the treatment. Adding controls does not affect the results.

The treatment caused acreage cultivated to increase: farmers in treatment blocks planted an average of 0.06 more acres with Swarna-Sub1 compared to farmers in control blocks, a 69% increase (Column 4). Columns 5 and 6 show that acreage results are unaffected when adding control variables. This adoption effect operates on both the extensive and intensive margins: in addition to increasing the rate of adoption, the treatment increased the cultivated area of adopters (Table B.1.5).

Table B.1.6 shows that the level of contact with extension agents or with cluster demonstrations is very low, even with our reinforced extension service in control blocks and that farmers in treatment blocks were no less likely to be in contact with extensions workers, or to have observed a demonstration of Swarna-Sub1, compared to control farmers.[28] In other words, we do not find evidence of displacement at the expense of other traditional channels.

Following up on the idea of displacement, we look at whether the treatment displaced other new varieties, potentially lowering welfare if it caused a shift away from high-quality seeds. We find no such evidence. Table B.1.7 shows that the treatment had a negative effect on adoption of only two seed varieties — both of which were released over three decades ago. It does not appear that the increase in adoption caused by agrodealers corresponds to a shift away from newly released technologies.

Finally, we find no evidence that the SMS messages increased adoption (Table B.1.8). They also did not change the effectiveness of the dealer treatment. The adoption gains from the dealer treatment cannot be obtained with a "lighter touch" SMS messaging intervention, at least in our context.

[27] For randomization inference, we use the randomization-t values described in Young (2019). Resampling is clustered at the block level and we use 1,000 replications.

[28] Only 5.7% of farmers report contact with the agricultural extension worker during the last year. This number is in line with other studies that showed low levels of contact between extension workers and farmers.

2.4.3 Heterogeneity

The evidence on average adoption rates shows that helping private agrodealers learn is more effective than conventional approaches used in the public sector. A concern may be that, as private agents, dealers optimize behavior based on their own profits; in contrast with government extension agents who might factor in equity and may be better at targeting farmers who have high expected returns to adoption. It is however not obvious whether profit maximizing dealers will deliver inferior targeting. Profit maximization strategies and farmers benefiting from adoption could coincide and may lead to similar outcomes, especially if we consider the repeated interactions between dealers and farmers over time.

In our context, being exposed to frequent flooding gives an easy-to-observe measure of potential returns — given the flood tolerance property of the variety.[29] We show that treatment dealers were successful at targeting Swarna-Sub1 to farmers who could benefit the most from the new technology, i.e. farmers who live in flood prone areas.

Figure 2.2 separates the sample by the satellite-based measure of past flooding and shows that treatment effects only exist in approximately half the sample where there were at least 3 flood days from 2011 to 2017. Conversely, the dealer treatment had little or no effect on adoption in the bottom half of the sample.

In Table 2.3, we show how the treatment effect depends on flooding risk. Two results stand out. First, control farmers from flood prone areas are *less likely* to adopt Swarna-Sub1. The third row in the table shows that being a high-risk control farmer is associated with a 6% lower likelihood of adoption compared to low-risk control farmers. This estimate is merely a correlation. Farmers exposed to flooding differ in a number of ways that might directly influence adoption. Second, and more importantly, the dealer treatment was only effective in flood-prone areas, i.e. the interaction between treatment and flooding exposure is positive. Column 1 shows that the dealer treatment targets high-risk farmers increasing their adoption by 6.4%, while the effect of the treatment is only 0.8% for low-risk farmers (and not significant). The difference between the two treatment effects (the interaction term) is statistically significant at the 10% level. This interaction effect may be picking up other correlates of flood risk. Column 2 shows that the results are not sensitive to including interactions between the flooding risk variable and all the covariates in Table 2.1. We find similar results when looking at acreage in Columns 3 and 4. The intervention only increased acreage of Swarna-Sub1 for farmers that were most exposed to flooding.

Table 2.4 shows analogous results where we split the sample according to flood risk. For farmers facing the most risk, the treatment increases adoption by 5.9 percentage points. This

[29] Our analysis focuses on flooding risk as one determinant of returns because it is measurable in our data. However, we acknowledge that the actual benefits to a given farmer depend on a number of factors, some of which are harder to observe, such as risk aversion.

amounts to a more than doubling since only 3.6 percent of farmers adopted in the control group. Similarly, Column 2 shows that the dealer intervention increases Swarna-Sub1 area by 0.11 acres for high-risk farmers. The intervention had no effect on adoption or acreage in places where flooding is less frequent (Columns 3 and 4).

As another piece of evidence, Table B.1.9 shows that the average adopter in treatment blocks is more exposed to flooding. Specifically, they are more than twice as likely to be above the median in terms of flood exposure.[30]

There is no evidence that informing dealers prioritizes adoption by the wealthiest farmers, which might have been expected if agrodealers cater more to larger and wealthier farmers. In particular, Table B.1.10 shows that there is no treatment-effect heterogeneity according to farm size. Adoption is more likely by larger farmers, but this is equally true in treatment and control blocks. We also find no heterogeneity according to being below the poverty line or in a marginalized caste group.

2.4.4 Supply-side responses to the treatment

Recall that we only treated a fraction of the dealers in each block. More precisely, 42% of sample dealers in treatment blocks received seeds and information (Table B.1.11). These dealers were not randomized. Hence, our dealer-level analysis compares all private dealers in treatment blocks to those in control blocks. We therefore capture any direct effect of receiving the seeds and information and any spillovers — which of course could be either negative or positive.

There is some evidence that the treatment caused dealers to increase the *availability* of Swarna-Sub1. Columns 1-4 in Table 2.5 show results from one year after the treatment (year 2). Focusing on all dealers — including those that were no longer operating — the treatment has a small positive effect on the likelihood of carrying Swarna-Sub1 at any time during the season (Column 1) and the total amount the dealer reported selling throughout the year (Column 2). But both of these estimates are very imprecise, partly due to some dealers no longer being in business. Amongst the subset of active dealers, those in treatment blocks were 6.2 percentage points more likely to carry Swarna-Sub1, a 17 % increase (Column 3). Column 4 shows that dealers in treatment blocks sold 3.7 additional quintals (1 quintal =

[30] One possible concern is that we failed to obtain GPS coordinates for all farmers — they are missing for about 17 percent of the sample (1,110 respondents). In the online appendix we impute the locations of these houses using village locations in one of two ways. If we observe other households in that village, then we use the average latitude and longitude values from the observed households (603 farmers). If we observe no other households in the village, then we try to match the village to the 2011 Census and use the village centroid as an approximate household location (323 farmers). Figure B.1.5 shows that results are robust to including these observations in the flood-heterogeneity analysis.

100 kg), which represents a 59% increase in volume sold. But again, while larger, neither of these results are close to statistically significant.

Anticipating on an intervention done in year 3 (and described below), we find large and precise effects on stocking behavior (Column 5). 19.3% of dealers in control blocks had Swarna-Sub1 in stock when visited by the secret shopper.[31] This increases by 11.4 percentage points (59%) in treatment blocks. This large effect is being observed two years after the treatment. It also comes from a direct observation of what the dealer had available on a certain day, rather than an estimate from what they recalled after the season. Lastly, we collected secondary data from OSSC on which dealers carried Swarna-Sub1 during the 2020 season (year 5). Column 6 shows that the treatment effect persisted during that season.

This result could be driven by a number of things. First, it could come directly from the dealers that were treated and had their information sets updated. Table B.1.12 shows some evidence of this: there is a positive correlation between receiving the demonstration seeds and selling them during future years. Second, dealers talk to farmers. Any increase in knowledge of farmers could spread to other dealers, not only those that were treated. Third, dealers were provided with several minikits for testing. They could have shared those in a way that increased local knowledge. We cannot distinguish between these effects in the analysis.

We next test whether the treatment changed the extent of local seed *production*. Our data here amount to six observations per block: three from the period before our treatment could have triggered a production response (2014-2016) and three from the post-treatment period (2017-2019). We therefore estimate block-level regressions of the amount of Swarna-Sub1 seed produced on treatment and district and year fixed effects.

We find evidence that treatment blocks produced more Swarna-Sub1 seeds after the experiment. Columns 1 and 2 of Table 2.6 show regressions using the total amount of seed production (in quintals) and its log. Seed production is highly skewed (Figure B.1.6).[32] As a result, the point estimate is more precise in the log regression of Column 2. It shows that the treatment led to a 47.9 percent increase in the amount of seed production, conditional on some production taking place. Columns 3 and 4 show that results become more precise when conditioning on average annual production during the 2014 to 2016 period. In Column 3, treatment increased production by 79 quintals, or about 38%. Again the result is more precise in the log regression, as Column 4 shows that treatment blocks produced an average

[31] The probability of having Swarna-Sub1 available appears lower in year 3 for a couple of reasons: 1) availability was observed on a specific day when the shopper visited and not across the entire season and 2) visits by the secret shopper occurred later in the season when varieties were no longer in stock for some dealers.

[32] No Swarna-Sub1 seed is produced for just over 40 percent of block-year observations. At the same time, large amounts of seed are produced in a small number of blocks.

of 57 percent more Swarna-Sub1 seeds during the three years after the treatment.

Columns 5 and 6 show separate effects for the three seasons after the experiment. None of the year-specific effects are statistically distinguishable. This provides some evidence that the treatment effects have not deteriorated over time. The online appendix helps visualize the results by showing cumulative distribution functions of seed production by treatment (Figure B.1.6). They show a noticeable rightward shift in the distribution for the treatment blocks, particularly in the top quintile of the distribution.

These results should not necessarily be interpreted as local production increasing to meet growing demand of farmers. In fact, seeds are often processed outside of the block where they are grown and can be sold anywhere in the state after processing. It seems more likely that intervening with agrodealers caused more people to know enough about Swarna-Sub1 to cultivate it. This group includes seed producers, who are often large landholders, and can rely on some of the same sources of information as smaller farmers.[33]

2.5 Mechanisms and motivations behind dealers' role in increasing technology adoption

Our analysis up to this point shows that the information treatment targeted at agrodealers causes more farmers to adopt new technology. That is, the approach of informing private agents on the supply side of technology can outperform standard approaches where frontline government workers interact directly with selected farmers. This section explores the mechanisms that are at play. More specifically, we first want to understand *how* dealers increase adoption. Section 2.5.1 shows evidence of one channel: dealers communicate information and recommendations to farmers. Secondly, we look into *why* treated dealers exhibit this behavior. We test whether profit opportunities motivate dealers when making recommendations (Section 2.5.2).

2.5.1 Dealer communication with farmers

There are different ways dealers could have increased farm-level adoption. For one, they may advise clients to purchase Swarna-Sub1, playing an active and directed information-sharing role that goes beyond the traditional information sharing approach practiced by extension

[33] Table B.1.13 shows that there is a positive correlation between farmer-level adoption (from our survey) and block-level seed production. This is evidence that seed producers select varieties that are best suited for local conditions — and hence selected by farmers.

agents. Alternatively, they may have played more of an indirect role by informing people who are well connected, or even by giving demonstration seeds to better connected farmers.

We used the third year of the experiment (2018 season) to test whether dealers actively advised farmers about the new technology. Eliciting advice is not easy. Simply asking dealers whether they informed farmers and/or recommended Swarna-Sub1 likely suffers from experimenter demand effects.

We alleviate this concern with a unique strategy to elicit advice using "secret shoppers". First, an enumerator visited dealers in both treatment and control blocks during the time when farmers usually buy seeds. The enumerator was someone the dealer had not seen before and who did not identify himself as part of the research. Then, the enumerator followed a specific script to obtain advice from the dealer. The enumerator mentioned that his father from a nearby village was planning to cultivate rice and was looking for information on possible varieties to grow. The enumerator asked the dealer which varieties to consider, without mentioning the name of any particular one.[34] Dealers usually mentioned several varieties — which we describe as the dealers "listing" varieties. If the dealer did not make a specific recommendation the shopper asked him which one he would recommend. We refer to this outcome as a dealer recommendation. If the dealer asked about type of land, he was told medium-low in terms of elevation and hence risk of flooding. We also asked which varieties the dealer currently had in stock.

Given the costs of these visits, and the scattered nature of our sample, we focused on the dealers that we reached during the previous year and were selling rice seeds. The sample consists of 310 dealers, 15 of which were not from our list obtained at the beginning of the study, and 15 of which we did not reach. The sample for analysis therefore includes 280 dealers.

To assess whether dealers actively advised farmers, we look at how they interact with the secret shopper who visited their shops. If dealers do play an active role in promoting Swarna-Sub1 to farmers, dealers located in treatment blocks should be more likely to list and/or recommend Swarna-Sub1 compared to control blocks. If instead, dealers do not play this active role, we should not expect to see differential recommendations from dealers between control and treatment blocks.

Table 2.7 shows results, where the rows are for separate outcomes. As with the earlier dealer-level results, these results are "intention to treat" since not all dealers in treatment blocks were informed. While the shoppers had a clear script to follow, it was impossible for the conversation to follow the same path for every single dealer. We therefore show results with and without fixed effects for the different shoppers.

[34]We phrased the question in these general terms to avoid priming the dealers to think about any particular variety.

Most dealers list Swarna as a popular variety and this is not different between treatment and control blocks. However, dealers in treatment blocks were 12-13 percentage points more likely to list Swarna-Sub1 as a possibility to consider, a 25% increase given that control dealers list Swarna-Sub1 51% of the time. Furthermore, treatment dealers were about 4-7 percentage points more likely to recommend Swarna-Sub1 — a large albeit non significant effect. Not surprisingly, the increase in Swarna-Sub1 recommendations comes at the expense of recommendations for Swarna — the variety that is otherwise similar, but does not offer flood tolerance. Indeed, being located in the treatment block reduces the likelihood dealers recommend Swarna by 13 percentage points, a 31% decrease in recommendations for this older variety.

We previously showed that dealers in treatment blocks were more likely to be carrying Swarna-Sub1 at the time of these visits. The last four rows of the table show that listing and recommending Swarna-Sub1 go hand in hand with stocking it. In other words, the treatment causes dealers to both suggest the new variety to farmers and carry it in their shops. This is evidence that treatment dealers play a direct role in the increase of Swarna-Sub1 adoption by mentioning it to farmers.

An alternative interpretation of our results is that the intervention worked only because it increased supply, making it easier for farmers to obtain the seed. This explanation differs from a mechanism where the treatment triggers more spread of information. Our data allow for one test. Farmers obtain seeds from multiple sources, including dealers, other farmers, and agricultural cooperative societies. Our follow-up survey asked farmers where they obtained seeds. Table 2.8 shows evidence that the dealer intervention led to greater adoption from sources other than dealers. This provides evidence that the effect is not being driven only by increased supply from the treated dealers. Rather, outreach via the private sector leads to more informed farmers and thus increases demand.

2.5.2 Profit opportunities as motivation for spreading information

We conducted an additional small experiment during the fourth year (2019) to learn more about what motivates dealers to expend effort in making recommendations. As was discussed above, dealers can be motivated by profits from increasing quantities sold, even when their margins for Swarna-Sub1 are the same as for other varieties. Beyond these additional profits, dealers could be motivated by pro-social preferences for other farmers' well-being, or they could simply want to be known as somebody who gives good advice.

We test the role of profit incentives by partnering with a local NGO that visited the dealers in our sample and asked for recommendations for a new program. More specifically, the

NGO informed dealers that they would organize a seed demonstration in a village (located in the dealer's block) where farmers would cultivate a new variety and villagers would be invited to learn about it. They asked dealers for recommendations on the type of variety, which village to choose, and who would be the best farmers within the chosen village to select for cultivation. Dealers were randomly assigned to receive one of two different programs from the NGO. In the first group, the NGO made it clear that the dealer would have a profit opportunity. Dealers in that group were informed that their name would be displayed and advertised during the demonstration. Broadcasting the dealer's participation was chosen to increase the opportunity for him to profit from increased sales as a result of the demonstration. Dealers in the second group were asked for advice on a demonstration that was similar, but had two key differences. First, the dealer's name would not be displayed as part of the demonstration. Second, the NGO explained to dealers that the harvest would be collected after the demonstration and redistributed as seeds to other farmers in the village. Distributing seeds was meant to further limit profit opportunities, since it would reduce the need of farmers to buy seeds after the demonstration. We recorded the responses for four types of outcomes during the interview: dealers' effort invested in providing recommendations to the NGO, the varieties they suggest using, the village they recommend for the program, and lastly which farmers they suggest partnering with to carry out the demonstration.

Table 2.9 shows the results, which are divided into five groups. First, we find that dealers in the profit motivation treatment spent a non-significant 10 percent more time on making recommendations (row 1). The second row shows that this increase in effort appears mostly when, in the conversation, dealers are asked which specific farmers to rely upon for the demonstration: a 14% increase in time invested on an average of just over two minutes picking farmers to recommend. Second, the profit motivation treatment seemed to shift dealers away from suggesting popular seed varieties and toward selecting them based on land type. The fourth row shows that when asked why they suggested a particular variety, 16 percent of dealers report doing so because it is locally popular; and this falls by 7.1 percentage points, or 44% in the treatment group.[35] While the estimate in the next row is not significant, the treatment seems to cause dealers to recommend varieties based on their agronomic characteristics, suitability to the land, and how long they take to grow (duration).

Third, we find some evidence that the treatment changed how dealers recommend villages. Dealers were presented with a list of three randomly selected nearby villages to choose from. They were also given an option of recommending a village not on that list. The treatment increased the likelihood that dealers took this option by 12.7 percentage points (a 34 percent effect). Most dealers taking this option picked their own village. Amongst dealers that wanted to recommend a village not on their randomly selected list, 63 percent of the control

[35] Popular varieties mostly correspond to older seeds that farmers have been growing for a long time.

group picked their own village. This increased by 13.5 percentage points for the treatment group, although the difference is not quite statistically significant.

Fourth, dealers were asked to identify three farmers to grow the seeds as part of the demonstration. About 81 percent of dealers report that they selected existing clients. This falls by 9.2 percentage points with the treatment. Linking a dealer's name to the recommendation seems to instead cause them to suggest farmers that are villagers or close neighbors. The dealer's ability to observe the demonstration plot when it is cultivated by a close neighbor might explain this result. Dealers may put more value on having this ability when their name is linked to the demonstration.

Finally, dealers at the end of the visit were asked if they felt the demonstration would affect their business. Around 67% of the control group reported it would. This suggests that many dealers thought the demonstration would increase seed demand — even though their names would not have been identified. The treatment increased the perception that the demonstration would affect business by 8.6 percentage points, a 13 percent effect. This effect is modest, but it aligns with the other results showing how this subtle treatment changed some of the advice given by dealers.

These findings help interpret the results of our main experiment. They provide suggestive evidence that the advice given by dealers is motivated at least partly by their concerns about how it will affect their future profits. While helping to understand the incentives at play, this additional experiment has limitations. First, it does not mimic an everyday conversation between dealers and customers as closely as our secret shoppers did the previous year. Framing the conversation around an on-farm demonstration made it possible to experiment with the salience of profit opportunities. But it has the downside of not being the same type of conversation that would happen between a farmer and dealer. Second, the sample size is small, as it was limited to the active dealers from our original list. As such, some of the estimates are imprecisely estimated.

2.6 Discussion and conclusions

This article provides evidence on how intervening on the supply side of agricultural input markets can be more effective than public agricultural extension services when providing information to promote the adoption of new technologies. Intervening with private input suppliers, i.e. agrodealers, would be a substantial transformation of the standard methods currently used. Government workers most often try to spread knowledge via direct contacts with selected farmers, expecting those to, in turn, diffuse information in their social networks. Much of current research on information constraints tries to identify ways of optimizing this approach. Our paper instead provides an empirical test of a different approach where information is transmitted to private input suppliers.

We find that informing private agrodealers about a new and profitable seed variety, and giving them small amounts of demonstration seeds to test, causes farm-level adoption to increase by over 50 percent compared to the business-as-usual approach where government workers focus on outreach with selected farmers. Using the private-sector approach increases adoption most among farmers with higher expected benefits from the technology. This improvement in targeting suggests that there is an alignment of incentives between dealers and farmers: dealers benefit from inducing farmers to adopt the right technology. We also found that our treatment triggers a supply response on the seed market. It causes dealers to be more likely to keep the seed in stock and it increases local production of the seed.

Further evidence shows that these effects can be at least partly explained by dealers actively advising farmers. Dealers in treated locations are more likely to mention the new seed variety when asked what to grow by a "secret shopper". Unpacking the incentives of agrodealers is difficult, and something we cannot do perfectly with our experiments. But we find some evidence in our second experiment that profit motives create incentives for agrodealers to give advice.

Our findings thus show potential for a different approach to agricultural extension in developing countries: delivering information on the supply rather than the demand side of technology. There might be drawbacks to this type of approach in some contexts. Particularly, if the technology is not a product that agrodealers can sell for a profit, then using them as information agents may not increase adoption. But it appears that when profit motives exist, as in the case of our experiment, making input suppliers better informed can improve the practice of agricultural extension.

Tables

Table 2.1: Summary Statistics and Covariate Balance

	Means		
	Control	Treatment	p-value
Panel A: Block Characteristics (N=72)			
Number Villages	136.0	135.4	0.878
	(64.04)	(60.52)	
Population	110687.2	120997.2	0.296
	(38991.2)	(49184.3)	
Annual Swarna-Sub1 Seed Production	298.9	191.9	0.289
	(421.2)	(264.6)	
Share Scheduled Caste	0.209	0.214	0.617
	(0.0492)	(0.0522)	
Share Scheduled Tribe	0.0462	0.0286	0.136
	(0.0767)	(0.0441)	
Elevation (Meters)	23.51	19.72	0.175
	(24.43)	(19.35)	
Literacy Rate	0.727	0.737	0.425
	(0.0726)	(0.0475)	
Share Agricultural	0.636	0.651	0.537
	(0.141)	(0.0931)	
Child Sex Ratio, 0-6 yrs	1.072	1.069	0.746
	(0.0229)	(0.0214)	
Panel B: Farmer Characteristics (N=6653)			
Age	49.42	49.83	0.401
	(11.12)	(11.38)	
Years Education	7.930	7.818	0.303
	(4.467)	(4.533)	
Below Poverty Line Card	0.471	0.510	0.245
	(0.499)	(0.500)	
Female Farmer	0.0794	0.0809	0.714
	(0.270)	(0.273)	
Scheduled Tribe	0.0160	0.00569	0.0336
	(0.125)	(0.0752)	
Scheduled Caste	0.150	0.153	0.853
	(0.357)	(0.360)	

Panel A shows means and standard deviations of block-level characteristics from the 2011 Census of India (with exceptions of elevation and annual Swarna-Sub1 seed production). Elevation is calculated from satellite data and the annual Swarna-Sub1 seed production is the average annual amount of seed processed from registered growers in the block from 2014 to 2016. It is measured in quintals (1 quintal = 100 kg). The literacy rate is defined as the number of literate individuals divided by the population older than 6 years old. The share agricultural is defined as the number of people working in agriculture divided by the working population. The child sex ratio is the numer of male children 0-6 years old divided by the number of female childen 0-6 years old. Panel B shows means and standard deviations of characteristics from our household survey. The p values in column 3 are for the treatment variable in regressions of each characteristic on treatment and district (strata) fixed effects. Panel A uses robust standard errors while Panel B clusters errors at the block level.

Table 2.2: Treatment Effects on Technology Adoption

	Adoption			Acres		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.0348*	0.0324**	0.0312**	0.0641**	0.0611**	0.0640**
	(0.0193)	(0.0147)	(0.0146)	(0.0311)	(0.0251)	(0.0251)
	[0.0949]	[0.0709]	[0.0739]	[0.0649]	[0.0340]	[0.0250]
Dependent Variable Control Mean	0.0634	0.0631	0.0631	0.0932	0.0926	0.0926
R-Squared	0.0280	0.0626	0.0617	0.0265	0.0449	0.0437
District Fixed Effects	X	X	X	X	X	X
Controls		X			X	
LASSO picked controls			X			X
Observations	6653	6599	6601	6653	6599	6601

The table shows the main treatment effects on adoption and acreage. All regressions use the data from the follow-up survey with farmers in August/September of 2017. The dependent variables are whether the farmer was currently using Swarna-Sub1 (columns 1-3) and the acreage cultivated with Swarna-Sub1 (columns 4-6). Columns 2 and 5 include all the control variables in the balance table. Columns 3 and 6 only include the control variables that are selected by an adaptive LASSO (ridge) regression. The standard error (clustered at the block level) is reported in parentheses below each point estimate. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels. P-values calculated by randomization inference are reported in brackets below each standard error.

Table 2.3: Heterogeneous effects by flooding risk

	Adoption		Acres	
	(1)	(2)	(3)	(4)
Treatment	0.00788	-0.00404	0.0260	0.00194
	(0.0251)	(0.0218)	(0.0426)	(0.0389)
Treatment * Above Median Risk	0.0563*	0.0589*	0.0909	0.116**
	(0.0330)	(0.0304)	(0.0590)	(0.0576)
Above Median Risk	-0.0595***	-0.0644***	-0.0955***	-0.103**
	(0.0218)	(0.0242)	(0.0344)	(0.0411)
District FE	Yes	Yes	Yes	Yes
Treat*Controls	No	Yes	No	Yes
Mean in Control	0.063	0.063	0.089	0.088
P-value Treat + Treat*High Risk	0.018	0.014	0.009	0.002
Number of Observations	5536	5495	5536	5495
R squared	0.036	0.072	0.033	0.054

The table shows heterogeneous treatment effects by flooding exposure. The dependent variable in columns 1-2 is an indicator for adopting Swarna-Sub1. Flooding risk is calculated by using satellite images from 2011-2017 (June-October) to count the total number of days where flooding was detected within 1 km of the farmer's house. Above median risk is a binary variable indicating a farmer that is above the median for the days of flood exposure. The dependent variable in columns 3 and 4 is the acreage cultivated with Swarna-Sub1. Columns 2 and 4 include interactions between all the control variables in Table 2.1 and the above-median risk variable. Standard errors are clustered at the block level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Table 2.4: Separate effects by past flood exposure

	High Risk		Low Risk	
	(1)	(2)	(3)	(4)
	Adoption	Acres	Adoption	Acres
Treatment	0.0586**	0.111**	0.00222	0.0174
	(0.0255)	(0.0437)	(0.0256)	(0.0426)
	[0.0539]	[0.0170]	[0.949]	[0.743]
Dependent Variable Control Mean	0.0357	0.0517	0.0831	0.116
R-Squared	0.0600	0.0416	0.0326	0.0320
District Fixed Effects	X	X	X	X
Observations	2508	2508	3028	3028

The table shows the main treatment effects on adoption and acreage separate for farmers with above- and below-median past flood exposure. All regressions use the data from the follow-up survey with farmers in August/September of 2017. The dependent variables are whether the farmer was currently using Swarna-Sub1 (columns 1 and 3) and the acreage cultivated with Swarna-Sub1 (columns 2 and 4). The standard error (clustered at the block level) is reported in parentheses below each point estimate. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels. P-values calculated by randomization inference are reported in brackets below each standard error.

Table 2.5: Treatment Effects on Supplying Seeds by Dealers

	Year 2: All		Year 2: In business		Year 3	Year 5
	(1) Carry	(2) Quantity	(3) Carry	(4) Quantity	(5) In Stock	(6) Carry
Treatment	0.026	1.633	0.062	3.662	0.114**	0.102*
	(0.052)	(1.676)	(0.083)	(2.656)	(0.057)	(0.060)
Dependent Variable Control Mean	0.242	3.793	0.397	6.231	0.193	0.252
R-Squared	0.045	0.037	0.114	0.117	0.097	0.116
District Fixed Effects	X	X	X	X	X	X
Observations	473	472	274	273	280	252

The table shows treatment effects on Swarna-Sub1 inventories from a survey of dealers (Columns 1-4), the secret shopper visit (Column 5), and an online database with dealer-level inventories (Column 6). Columns 1 and 2 are for the sample of dealers that were located and surveyed during year 2 (September 2017). Columns 3 and 4 are for the subset of those dealers that were actively in the seed business during that same season. Column 5 is for dealers that were visited in the secret shopper sample during year 3 (June 2018). Column 6 is for the 252 of these dealers for which we were able to obtain their license numbers. These numbers were matched to an online database with dealer inventories The standard errors in each regression are clustered at the block level. The dependent variables are an indicator for whether the dealer reported carrying Swarna-Sub1 at any time during the season (columns 1 and 3), the total quantity sold throughout the season (columns 2 and 4), an indicator for whether the dealer had Swarna-Sub1 in stock when visited by the secret shopper (column 5), and an indicator for whether the dealer showed positive inventory in the online database at any time during the season (column 6). Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Table 2.6: Treatment Effects on Local Seed Production

	Time Period: 2017-2019					
	(1) Amount	(2) Log	(3) Amount	(4) Log	(5) Amount	(6) Log
Treatment	14.014	0.479*	79.494	0.568***		
	(81.938)	(0.262)	(52.494)	(0.209)		
2014-2016 Production			0.736***	0.002***	0.736***	0.002***
			(0.122)	(0.000)	(0.123)	(0.000)
Treatment X Year=2017					76.218	0.575**
					(68.242)	(0.283)
Treatment X Year=2018					53.207	0.740***
					(96.331)	(0.250)
Treatment X Year=2019					109.057	0.300
					(67.453)	(0.364)
Dependent Variable Control Mean	209.511		209.511		209.511	
p-value: 2017=2018					0.670	0.544
p-value: 2017=2019					0.753	0.486
p-value: 2018=2019					0.667	0.314
R-Squared	0.149	0.391	0.435	0.532	0.436	0.536
District Fixed Effects	X	X	X	X	X	X
Year Fixed Effects	X	X	X	X	X	X
Observations	216	124	216	124	216	124

The table shows the effects of the dealer treatment on block-level seed production of Swarna-Sub1. Publicly available data on producer-level production of certified Swarna-Sub1 seeds were matched to the blocks in the experiment. The unit of observation in each regression is the block-season, where years range from 2014 to 2019. All columns are for the 2017-2019 period, 1-3 years after the intervention. The dependent variables are the annual amount of Swarna-Sub1 seed processed from growers in the block (columns 1, 3, and 5) and its logged value (columns 2, 4, and 6). Columns 3-6 control for the average annual production from 2014-2016 (the pre-period outcome). The standard errors in each regression are clustered at the block level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Table 2.7: Treatment Effects on Providing Advice to Secret Shoppers

	Control Mean (1)	Treatment Effect (2)	Treatment Effect w/ Shopper FE (3)
Listed Swarna	0.834	-0.0157	0.0129
		(0.0353)	(0.0337)
Listed Swarna-Sub1	0.510	0.120	0.130**
		(0.0740)	(0.0634)
Recommended Swarna-Sub1	0.297	0.0425	0.0701
		(0.0517)	(0.0484)
Recommended Swarna	0.421	-0.116**	-0.131**
		(0.0467)	(0.0495)
Listed Swarna-Sub1 & No Stock	0.352	-0.00892	0.00120
		(0.0576)	(0.0476)
Listed Swarna-Sub1 & Stocked	0.159	0.129**	0.129**
		(0.0514)	(0.0550)
Recommended Swarna-Sub1 & No Stock	0.179	-0.0166	-0.00249
		(0.0376)	(0.0419)
Recommended Swarna-Sub1 & Stocked	0.117	0.0591	0.0726*
		(0.0395)	(0.0406)

The table shows treatment effects on the type of information provided by dealers when they were visited by a secret shopper during year three (June 2018). Each row shows results from a separate regression (N=280) of that outcome variable on the block-level treatment indicator and strata fixed effects. Column 1 shows the control mean, while the coefficient estimate on the treatment indicator, and its standard error, are presented in Column 2. Column 3 shows results when also including a shopper fixed effect. Listing Swarna and Swarna-Sub1 (rows 1 and 2) are binary variables for whether the dealer included that variety when listing good varieties to try. Recommending Swarna and Swarna-Sub1 (rows 3 and 4) are binary variables for whether the dealer recommended that variety when asked to make a specific recommendation. Rows 5-8 show effects on listing / recommending the varieties and whether or not they were currently in stock with the dealer. The standard errors in each regression are clustered at the block level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Table 2.8: Treatment Effects on Adoption from Dealers and Other Sources

	(1) Dealers	(2) Other Sources
Treatment	0.00587	0.0289**
	(0.0110)	(0.0136)
	[0.650]	[0.0619]
Dependent Variable Control Mean	0.0223	0.0410
R-Squared	0.0107	0.0391
District Fixed Effects	X	X
Observations	6653	6653

The table shows the main treatment effects on adoption separately for whether the farmer reported getting the seeds from a dealer (column 1) or other sources (column 2). All regressions use the data from the follow-up survey with farmers in August/September of 2017. The dependent variable in column 1 is a binary variable equal to 1 for farmers that adopted Swarna-Sub1 and reported that they got seeds from dealers. The dependent variable in column 2 is a binary variable equal to 1 for farmers that adopted Swarna-Sub1 and reported they got seeds from any source other than dealers. The standard error (clustered at the block level) is reported in parentheses below each point estimate. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels. P-values calculated by randomization inference are reported in brackets below each standard error.

Table 2.9: Results on profit motivation and dealer recommendations

	Control Mean	Estimate
Timing		
Total time spent (minutes)	5.093	0.526 (0.406)
Time spent picking farmers	2.200	0.303* (0.166)
Variety Selection		
Number varieties recommended	2.893	-0.269 (0.212)
Selected popular	0.157	-0.071* (0.039)
Selected for duration/land type	0.179	0.078 (0.048)
Village Selection		
Chose outside of list	0.371	0.127** (0.056)
Picked own village	0.493	0.069 (0.059)
Picked own village if chose outside	0.635	0.135 (0.086)
Farmer Selection		
Chose because client	0.807	-0.092* (0.047)
Chose because villager/neighbor	0.121	0.091** (0.041)
Business Perception		
Felt treatment would affect business	0.671	0.086* (0.052)

The table shows the effects of the profit motivation treatment on suggestions to an NGO for carrying out on-farm demonstrations. The treatment was informing dealers that their name would be affiliated with the demonstration and the control group is not providing this information and explaining that the seeds from the demonstration would be distributed to other villagers. Column 1 shows the mean of the outcome in the control group, while column 2 shows the point estimate and standard error from a regression of the outcome on the treatment and district fixed effects. The unit of randomization is the dealer. Therefore, robust standard errors are in parentheses for column 2. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Figures

Figure 2.1: Timeline of interventions and data collection

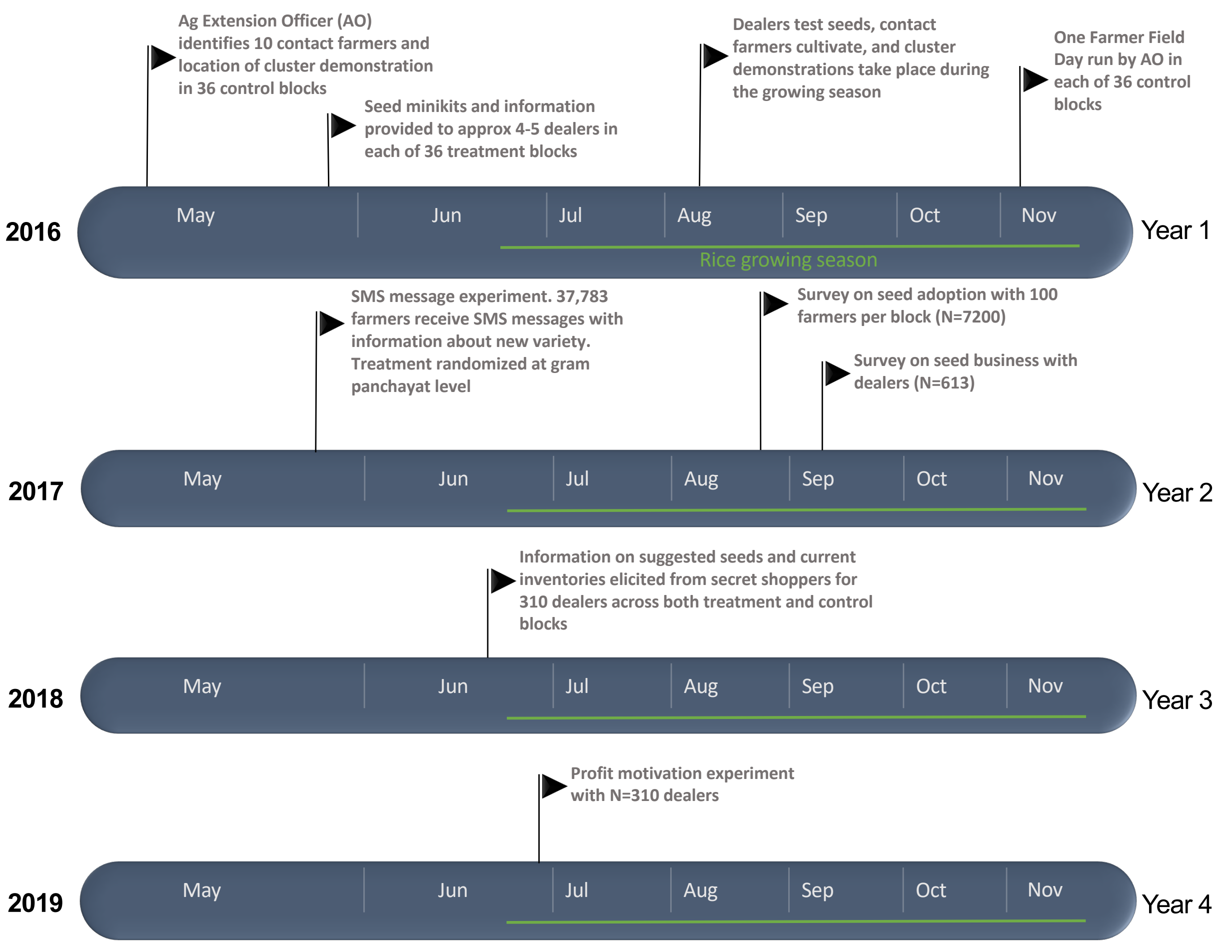
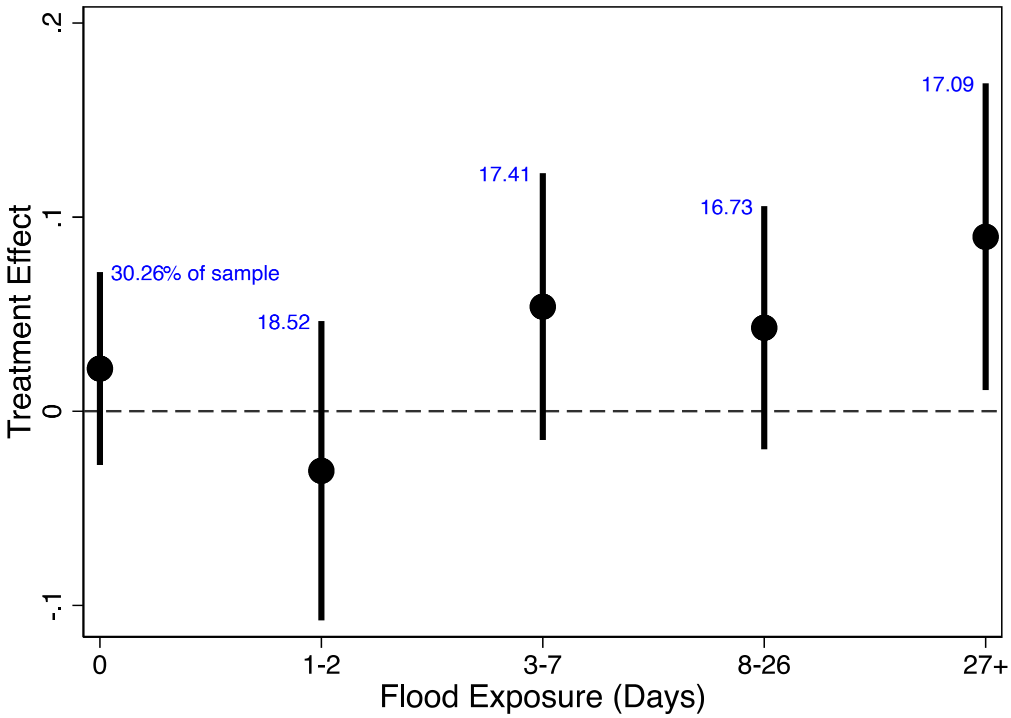


Figure 2.2: Treatment effects by flood exposure



Notes: The figure shows treatment effects from a single regression of adoption on separate treatment indicators for different levels of flood exposure and district fixed effects. The 5 bins of flood exposure correspond to households with no exposure from 2011-2017 and then an approximately equal division of households with at least one day of exposure. The percentage of observations in each bin is denoted in blue. The dots are the treatment effects of dealer-based extension and the vertical lines denote 95 percent confidence intervals.

Chapter 3

Poisoned by Policy: The Impact of the Flint Water Crisis on Political Participation

This chapter is coauthored with Kate Pennington

3.1 Introduction

In 2014, the city of Flint, Michigan experienced a massive government failure that exposed one in twelve households to lead in their tap water. The criminal misconduct that caused the crisis has been widely understood as an instance of institutional racism (Michigan Civil Rights Commission, 2017). An investigative task force concluded, "Flint residents, who are majority black or African-American and among the most impoverished of any metropolitan area in the United States, did not enjoy the same degree of protection from environmental and health hazards as that provided to other communities" (Flint Water Advisory Task Force, 2016). As such, the crisis created variation in individual perceptions of the government failure through differences in both personal lead exposure and racial identity. In this paper, we study the impact of the government failure on voting, new voter registration, and candidate choice through the channels of personal lead exposure and race. This setting offers an opportunity to explore the interplay between personal experiences and group identities in shaping individual perceptions and responses to shocks.

We combine address-level lead test results with voting records from the Michigan Qualified Voter File and precinct-level outcomes for mayoral elections from 2005-2015.[1] We

[1]We focus on mayoral elections for three reasons. First, the 2015 mayoral election is the only election

identify the causal impact of knowledge of personal lead exposure by exploiting random variation in the timing of receiving lead test results relative to the mayoral election that took place during the crisis. Because residents had to opt into testing, we avoid sample selection issues by restricting the sample to voters who opted in. "Treated" voters learned their lead test results before the election while control voters found out afterwards.[2] There are two potential treatments: receiving a positive test result or receiving a negative test result. We present results separately for each group.[3] Since race is not randomly assigned, we do not have a parallel causal identification strategy; instead, we explore racial heterogeneity in voters' response to contrast the responses of Black and White residents.

We find that both race and the knowledge of test results play a role in voters' response to the government failure. White voters did not change their voting behavior regardless of treatment. In contrast, Black voters who did not know their lead test results before the election increased turnout by 1-2 percentage points in both the positive and negative groups. Receiving a negative lead test result before the election drove an additional 14 percentage point spike in turnout among Black voters, while receiving a positive test result before the election had no significant impact.

We find the same pattern when we explore whether lead tests drove new voters to register. Using a proportional hazards model, we evaluate the monthly probability of registering to vote within the year after testing positive for lead, compared to testing negative. We find that testing positive has no impact on the probability of registering to vote for either Black or White Flint residents. However, Black residents accelerated their voter registration by 16 to 33% relative to Whites.

Finally, we use precinct-level voting results from the Genesee County Clerks Office to test whether lead test results caused changes in candidate choice. We take advantage of the fact that the same candidate ran in the two elections before the crisis and the first election after. Before the crisis, he won the majority in nearly every precinct. After the crisis, White voters continued to vote for the incumbent while Black voters sharply rejected him: an

that took place during widespread lead testing, which allows us to exploit variation in voters' information about their exposure. Second, it functioned as a referendum on the city government's handling of the water crisis, while the motivation to turn out for statewide and federal elections is influenced by many factors we cannot control for. Finally, the incumbent mayor who ran in 2015 also ran and won the two previous mayoral elections, improving comparability across those three elections.

[2]Although water quality fell universally after the switch to river water (it became noticeably more brackish and bad-tasting), the presence of dangerous lead levels in water is not detectable by appearance or taste. To find out if they were exposed to lead at home, residents needed to conduct a home lead test.

[3]We assume that voters had no priors about their lead status prior to receiving test results. Unlike many settings for studying information updating, the Flint water crisis was a one-time shock that made residents think about a health hazard they had not previously considered. Intuitively and identically for our identification strategy, one could also assume that all residents shared the prior of no exposure to lead.

additional percentage point of Black residents predicts a 0.51 percentage point decrease in the incumbent's vote share.

Our results highlight the central role of race in determining voters' response to the Flint water crisis. Despite personally experiencing a massive government failure, White voters did not change their behavior while Black voters did. We explore and reject alternative mechanisms and outline directions for future research.

We interpret these findings as strong suggestive evidence that the Flint water crisis made racial identity more salient for voter turnout and candidate choice. In a very simple framework, the crisis can affect behavior through three steps: signal, interpretation, and reaction. Racial identity could affect any of these steps: perhaps it sent voters different signals based on their race (for example, "This was targeted at you" versus "This was not targeted at you"); perhaps voters interpreted the crisis differently by race (for example, "This is part of a long history of government racism" versus "This was a fluke"); and finally, even if Black and White voters received the same signal and interpretation, they might react differently by race (for example, "I need to engage" versus "I don't need to engage") .[4] While we cannot differentiate between changing signals and interpretations here, the observed reactions make clear that the way government performance impacts political participation is determined not only by individual experiences of the current event, but also by history, identity, and the relationship of that identity to government performance.

This work contributes to a growing literature that documents how differences in the participation of Black and White Americans in institutions like voting and medical care are driven by historical events that shape beliefs (Williams, 2021; Alsan and Wanamaker, 2018; Hersh and Nall, 2016; Hutchings and Valentino, 2004). It also highlights the importance of looking at the intersection of identities in order to understand voter turnout (Uhlaner and Scola, 2016). In this study, we found that race mattered more than belonging to the most-harmed group.

We also contribute to a long, rich literature on voter motivation that seeks to explain why the theory of rational voting (Downs, 1957) cannot predict observed levels and patterns of voter turnout (Geys, 2006). This literature shows that emotional events like wartime casualties can impact voter turnout (Koch and Nicholson, 2016; Karol and Miguel, 2007) and that some voters are motivated by social pressures (Feddersen and Sandroni, 2009; Jankowski, 2002, 2007; Jones and Hudson, 2000; Panagopoulos, 2013) or sensitive to various costs of voting (Gerber and Green, 2000; Highton, 1997; Gentzkow, 2006; Falck et al., 2014). We shed light on the basic question of how government quality affects political participation and how group identity affects the reaction to experiences of government quality. We show

[4]These hypothetical signals and interpretations are based both on media reporting and on qualitative interviews with Flint residents conducted in 2017.

that race is a primary frame through which voters understand their government, and that Black and White voters see government very differently. Future research should explore how and why these differences in experience produce the differences in political participation we document.

As such, this work underscores a new recognition in the field of economics that race needs to be studied directly. Only 0.2 percent of the articles in the top ten economics journals published between 2010 and 2020 explore issues of race, racial inequality, and racism (Cihak et al., 2020). The majority of economics research on race analyzes racial disparities as the result of either a taste for discrimination (Becker, 1971) or statistical discrimination (Arrow, 1998; Phelps, 1972) without addressing the role of institutional discrimination (Small and Pager, 2020; Lang and Kahn-Lang Spitzer, 2020). This paper joins a growing body of work that demonstrates the necessity of explicitly studying the construction and operationalization of race, including the perception of discrimination, the role of past discrimination on contemporary outcomes, and the impact of organizational and legal discrimination (Small and Pager, 2020).

3.2 Data and Setting

Flint's fortunes have risen and fallen with General Motors production, booming with the founding of General Motors (GM) in 1908. At its 1978 peak, GM employed roughly 80,000 residents – 71% of the population over 16. But when GM began downsizing in the 1980s, Flint sputtered into a sharp decline. Today, GM employs fewer than 8,000 residents and Flint is one of the United States' poorest cities. Roughly 40 percent of residents live below the poverty line. Flint is also one of Michigan's few majority Black cities, and racial inequality is a visible and defining problem. The median income for White-headed households is $29,492 and $22,231 for Black-headed households.[5]

Flint's economic contraction has caused the city serious financial difficulties. In 2011, Michigan Governor Rick Snyder used a controversial state law to suspend the municipal government and appoint an emergency financial manager.[6] To cut costs, the emergency manager terminated Flint's water contract with Detroit and joined a new water authority

[5]U.S. Census Bureau (2015). American Community Survey 1-year estimates. Retrieved from Census Reporter Profile page for Flint, MI.

[6]Public Act 72, passed in 1990, granted the Governor the power to appoint an emergency manager on the grounds that the local government is a branch of the state government. In 2011, Gov. Snyder signed Public Act 4 to strengthen the EM law. Voters rejected it by referendum in 2012, but the state legislature responded by passing a new version of the law, PA 436, that included language protecting it from repeal (Local Financial Stability and Choice Act, 2012).

that planned to pipe water from Lake Huron. During construction, Flint would extract water from the Flint River.

It is routine to add corrosion control chemicals to corrosive surface water sources like rivers to prevent them from leaching metals from pipes. Yet the water management team decided not to without conducting a single water safety test. Although the municipal water mains were up to code, many of the service lines that conduct water from the mains into buildings still contained lead (see Appendix Figure C.1.1). One in twelve households was exposed to lead in their tap water as the river water stripped lead from old pipes in their homes. One resident faced lead concentrations of 13,200 parts per billion (ppb) – more than double the EPA's threshold for classification as hazardous waste. The maximum legal concentration is 15 ppb.

Any level of exposure to lead is harmful. It accumulates in the brain, liver, kidneys, and bones, causing cognitive damage and increased risk of ADHD (Chen et al., 2007), increased school discipline incidents (Aizer and Currie, 2019) and lower test scores (Aizer et al., 2018), aggressive behavior, and arrests (Wright et al., 2008; Needleman, 2004; Needleman et al., 1990). Early-life exposures have permanent consequences for later-life employment and earnings (Currie and Almond, 2011), and they can be physically inherited across generations. Pregnant women whose bones contain lead can pass it to their fetus, increasing the risk of miscarriage, stillbirth, premature birth and low birthweight, minor malformations, and cognitive damage.

After the switch to river water, the percentage of children with elevated blood lead levels in Flint doubled (Hanna-Attisha et al., 2016). Yet the Department of Environmental Quality, state, and city governments concealed, ignored, and denied the danger. Michigan's Attorney General has brought criminal charges against 15 state and local officials, including Governor Snyder and two emergency managers. The director of the Michigan Department of Health and Human Services faces a felony count of misconduct in office and Michigan's chief medical executive has been charged with obstruction of justice and lying to a police officer. Five officials have been charged with involuntary manslaughter for twelve deaths caused by Legionnaires' disease, a respiratory disease caused by bacteria in the river water (Ridley, 2016; Office of the Attorney General, 2021).

After major public outcry, Flint's municipal government was restored to power on April 30, 2015 and it reconnected to Detroit water on October 16. But now that the pipes have been damaged, even the gentler Detroit water can leach lead. The city plans to replace all lead service lines. As of 2020, it had completed nearly 10,000 pipe replacements and most residents have access to safe water – although distrust remains high (Robertson, 2020).

We study the impact of this massive government failure on political participation by combining data from four sources: individual voting records from the Michigan Qualified Voter file (Department of Elections, 2017), address-level lead test data provided by the City of Flint and independent water quality researchers (Taking Action on Flint Water, 2016), precinct-level election results from the Genesee County Clerks Office (Genesee County, 2015), and public use 2014 American Community Survey demographic data at the blockgroup level.

The MQVF is a centralized voter registration database of all 7.3 million voters currently registered in Michigan, available by Freedom of Information Act request. It maintains a file of voters' personal information (name, street address, sex, date of birth, date of registration) and a separate file of their voting history (election date, election ID, and jurisdiction voted in[7]). Since race is not included in the voter file, we impute it by mapping individuals' observables to the demographic distribution in their Census block.[8] Using this information, we construct a panel of individual voter registration and turnout histories to study Flint's mayoral elections from 2005-2015.

One drawback of the MQVF is that it cannot track voters' locations perfectly. First, it only updates voter addresses when and if they re-register in a new location. We cannot identify moves that occur without reregistration. Second, when a voter reregisters, their new address overwrites their former address. Consequently, we can (imperfectly) track voters' migration across jurisdictions, but not across street addresses.[9]

This inability to track voters' moving decisions could threaten identification if an influx of newcomers arrived or a large number of Flint residents left after the crisis. We are not concerned about the first issue because Flint's population has been steadily declining for more than three decades. However, the second issue may still be a concern. Since we do not observe the street address of the voters who left Flint after the crisis, we cannot identify voters who may have 'voted with their feet' after receiving a positive lead test result. We drop all voters who observably left Flint after the crisis.[10] To avoid including voters who

[7]More precisely, a file of defunct voter registrations is also available. Voting locations can be inferred at the election administrator level, which is usually the jurisdiction, and merged into the voter's history data by voter ID.

[8]We use the R package wru, which follows the procedure developed by Imai and Khanna (2016) to predict individual race/ethnicity using surname, first name, middle name, geolocation, and other attributes, such as gender and age. The method utilizes Bayes' Rule to compute the posterior probability of each racial category for any given individual. We acknowledge that imputation introduces noise, but it is the best we can do given that race is not self-reported in the voter file.

[9]If a voter stops voting, we cannot infer whether they stopped turning out, died, or moved out of state. The Department of Elections addresses this problem through formal cleanings in the January following each even-year general election. State and federal law requires election officials to send a notice to voters if there is "reliable information" that they have moved. Common forms of reliable information include giving up a Michigan driver's license to another state and mailings (such as an ID card) coming back undeliverable. If the voter does not respond to the notice and two Federal election cycles pass, the voter's registration is canceled. In January 2017, around 130,000 voter registrations were canceled for this reason. Source: email from MI Bureau of Elections, April 13, 2017.

[10]We identify 2,339 voters who were registered to vote in Flint before the crisis but re-registered elsewhere

have died, we also drop voters older than 82.[11] After limiting the set of voters and elections, we observe 83,516 Flint voters who have faced 2 million turnout decisions. We note that we have found no evidence of special Get Out the Vote efforts for the 2015 mayoral election either in the media or in conversations with residents.

Lead tests were conducted on-demand by the City of Flint and independent water quality researchers from September 2015 through June 2018. About one quarter of Flint's residential addresses were tested. We match 11,675 out of 13,410 tested addresses with voters, yielding 31,093 voters matched with a lead test. Table 3.1 reports summary statistics. Although Flint is very racially segregated, the incidence of lead pipes is spread across both Black and White neighborhoods (see Appendix Figures C.1.2 and C.1.3 for maps of demographic characteristics and lead tests). Every precinct contains homes that have tested positive for lead, with the heaviest concentrations in the southeast precincts.

This pattern is driven by Flint's relatively uniform building age: the city was built in two main bursts to serve GM expansions in the 1920s and 1940s. Voters who tested positive for lead tend to live in slightly older buildings than voters who tested negative. Most voters live in neighborhoods, defined by radii from their home address, where 5-20% of the service lines are lead, and there is substantial variation in both the number and the percent of lead service lines within a given radius.

3.3 Impacts on individual turnout

We exploit variation in the timing of voters' knowledge of their lead test results relative to the 2015 mayoral election to identify the causal impact of knowledge of lead status on the probability of voting. Our specification needs to address several potential sources of selection. First, voters who opt into lead testing may be systematically different from voters who never get tested. Second, voters who get tested very early in the crisis may be different from those who get tested later. In particular, voters who get tested early may be a more 'engaged' type, who are already more likely to vote. Third, voters who test positive for lead may be systematically different from those who test negative if having lead pipes is not fully random.

To address these issues, we first limit the sample to the 31,093 voters who got a lead test at some point from September 2015 to June 2018. Then we further divide the sample into the group of people who test positive versus negative. We consider testing positive and testing negative to be two different treatments, and we study each of them using a difference

after the crisis. Of these known leavers, 31% are Black and 65% are White.

[11] Average life expectancy was 76 years in Genesee County in 2014, according to the US Health Map published by the Institute for Health Metrics and Evaluation.

in difference specification that compares differences in voting probabilities using timing of the election relative to the crisis (pre-post crisis) and the timing of the test relative to the election (informed-uninformed). Specifically, we look at the differential change in voting patterns before and after the Flint crisis between individuals who were informed of their lead test compared to those who weren't.[12]

In Figure 3.1, we provide visual evidence of parallel trends in turnout among voters with different lead information status in the two elections before the crisis. Our regressions compare groups shown in the same color: first, the informed and uninformed lead groups shown in red, and second, the informed and uninformed nonlead groups shown in blue. In the post-crisis election, turnout deviates from trend for voters who were informed of their lead status before the election.[13]

Our main specification is:

$$Voted_{ijb} = \alpha_1 Informed_{ij} + \alpha_2 Post_j + \beta Informed_{ij} \cdot Post_j + \omega_1 X_i + \gamma_b + u_{ijb} \tag{3.1}$$

where $Voted_{ijb} = 1$ if voter i living in city block b voted in election j, $Informed_{ij} = 1$ if person i received test results before election j, and $Post_j = 1$ if the election occurred after the water switch. The main coefficient of interest is β. We include γ_b city block fixed effects and X_i individual controls including sex, birthyear, and voting habit as measured by the number of elections person i voted in before August 2014.[14] We cluster standard errors at the voter level.

Table 3.2 presents estimates of the impact of lead test results on individual turnout. Column (1) displays results from running (3.1) on the full sample. We find that receiving lead test results had no impact on the probability of voting on average.

However, this null effect masks significant racial heterogeneity. Although the MQVF does not include individual race, we use 2010 Census block data on racial distribution to impute the probability that voter i is Black, White, or Other using their name, sex, and birthyear.

Looking first at how voter turnout evolved after the crisis, Black voters who were not aware of their lead status increased turnout by 1-2 percentage points on average (Columns 2 and 5), whereas uninformed White voters did not change their voting behavior after the crisis (Columns 3 and 6). Salience of government failure (proxied by getting test results)

[12]Our main results include all voters who ever opted into lead testing, so there might be selection with respect to the timing of requesting a lead test. As a robustness check, we show that the results are stable if we limit the sample to the voters who received lead test results within 2 months of the 2015 mayoral election (see Appendix Table C.2.5.

[13]Because our model considers a single treatment event and does not use two-way fixed effects, it does not create the problems associated with difference-in-differences models documented in the recent literature.

[14]Our results are robust to a number of different measures of voting habit, including percent of past elections voted in, or counts and percents by types of election.

therefore impacts Black voters' turnout decisions. Column 5 reveals that Black voters who tested negative increased turnout by 14 percentage points (46%) if they were informed of their test result before the election.[15]

3.3.1 Potential mechanisms

To summarize our main results: (1) Knowledge of test status increases voter turnout significantly for Black voters, not for White voters; (2) The positive effect on voter turnout for informed Black voters is driven by voters who tested safe for lead; (3) Black voters who did not know their test status increase voter turnout after the crisis, White voters do not. This third result should be interpreted as a time trend, not as a causal relationship between race and turnout.

Next, we explore potential explanations for this striking racial divergence. We test three alternatives to the explanation that the Flint water crisis made racial identity more salient. First, we test whether the divergence is driven by a racial income gap, which might make dealing with lead exposure more burdensome for Black households. Second, we explore whether neighborhood lead exposure is an omitted variable correlated with individual race. If voters are motivated by sympathy for their neighbors' experiences, and race is correlated with neighborhood lead exposure, then individual race might be picking up an effect of the intensity of exposure to the government failure.[16] Media reporting and our own conversations with Flint residents suggest that this may be the case, both indicating that Black residents believed both that the crisis was motivated by institutional racism and that Black neighborhoods were more likely to be poisoned. Finally, we explore whether it is neighborhood racial composition that matters, rather than individual race.

To explore each of these potential mechanisms, we run the following specification separately for the subsample who tested positive and who tested negative:

---

[15]In Appendix Tables C.2.3 and C.2.4, we show that there is no meaningful heterogeneity by gender within race. Appendix Table C.2.6 shows that the results are similar even if we expand the sample to include all Flint voters, even those who never opted into lead testing.

[16]We also explore whether neighborhood lead tests have a direct effect on turnout by including them as explanatory variables rather than as sources of heterogeneity in the main effects. We find that voters who live in neighborhoods with higher concentrations of positive lead tests also differentially reduce turnout after the crisis, although the effect is very small. The effects of individual lead tests results by race remain qualitatively similar. Full results are included in the online appendix.

$$
\begin{aligned}
Voted_{ijb} = & \alpha_1 Informed_{ij} + \alpha_2 Post_j + \beta Informed_{ij} \cdot Post_j + \\
& \gamma_1 Informed_{ij} \cdot Post_j \cdot Black_{ij} + \gamma_2 Post_j \cdot Black_{ij} + \gamma_3 Informed_{ij} \cdot Black_{ij} + \\
& \gamma_4 Black_{ij} + \delta_1 Mechanism_{ij} + \delta_2 Informed_{ij} \cdot Mechanism_{ij} + \\
& \delta_3 Post_j \cdot Mechanism_{ij} + \delta_4 Informed_{ij} \cdot Post_j \cdot Mechanism_{ij} + \omega_1 X_i + \gamma_b + u_{ib}
\end{aligned} \tag{3.2}
$$

If the mechanism is driving the racial result, then it would change γ_1. If not, then adding the mechanism will leave γ_1 unchanged.

We find that including these mechanisms does not meaningfully change γ_1, suggesting that leaving them out is not causing omitted variable bias (Appendix Figure C.2.1 plots the γ_1 coefficients from the main specification and each mechanism and Appendix Table C.2.1 presents results in tables). We interpret this as suggestive evidence that the differential effects of individual race are not explained by differences in neighborhood income, lead exposure, or racial composition.[17] These results reject simple models of voter turnout as a function of individual lead test results alone.

Instead, the results suggest that individual race fundamentally shapes how voters experienced the shock. This interpretation is supported by large differentials in polls of Black versus White Michiganders on their satisfaction with the government's handling of the crisis.[18]

3.4 Impacts on new voter registration

Did the Flint water crisis motivate new voters to register? We focus on the sample of people who were not registered to vote at the time of their lead test. Data on people who are tested and eventually register is taken from the MQVF. Data on people who are tested and never register comes from lead test records that we could not match with voters. Conservatively, we assume that one potential voter lives at each of these unmatched addresses, yielding a data set with 4,753 people of whom 2,381 tested positive for lead. We evaluate the monthly hazard of registering to vote within the year after testing positive for lead, compared to testing negative.

[17] Median income does not explain the change in turnout within race, either. Table C.2.2 shows results from running 3.2 on Black and White subsamples. This provides evidence against one hypothesis for the divergence between Black voters informed of a positive lead test and Black voters informed of a safe lead test: that all Black voters were motivated to vote more after the crisis, but the financial burden of dealing with lead exposure prevented the informed lead group from turning out.

[18] A statewide telephone poll of 600 likely voters, conducted January 23-26 2016, found that 48% of White respondents viewed Governor Snyder favorably compared to 16% of Black respondents. Free Press, 2016.

Our estimating equation is:

$$days\_to\_registration_{ibq} = \lambda_{q,b} + \alpha_1 L_i + \alpha_2 Black_i + \delta L_i \cdot Black_i + \beta X_i + e_{ibq} \tag{3.3}$$

where the outcome variable is days between the lead test and registration, $\lambda_{q,b}$ is the baseline hazard that person i who got tested in quarter q and lives in Census blockgroup b registers to vote, and L_i is a dummy equal to one if they tested positive for lead. Individual controls X_i include sex and birthyear.

The first column of Table 3.3 presents results. We find that α_1 is not significant: testing positive for lead did not have a significant impact on time to registration. However, being Black is associated with an acceleration of registration of 32.8%. The interaction of lead status and race is not significant, but its magnitude suggests that lead test results make no difference for voter registration for Black voters versus White voters.

We find the same result when we use a factor that combines individual race and lead test results. Individuals can fall into group Black, Positive; Black, Negative; White, Positive; or White, Negative:

$$days\_to\_registration_{ibq} = \lambda_{q,b} + \alpha group_i + \beta X_i + u_{ibq} \tag{3.4}$$

Column 2 shows that we find no difference in time to registration for informed-lead versus informed-safe White residents. In contrast, we find a large difference for Blacks versus Whites, regardless of lead test results. Informed-lead Blacks accelerate registration by 31.8% compared to informed-safe Whites, and informed-safe Blacks accelerate registration by 32.8%. These responses are statistically indistinguishable. Again, the major divergence is by race, not by lead information.

3.5 Impacts on candidate choice

Finally, we investigate the effect of treatment intensity on voter turnout and candidate choice at the precinct level. We count the number of positive lead tests in each precinct using the variable *n_treated* and study its impact on outcomes for precinct p in election j:

$$y_{pj} = \beta_1 Post_j + \beta_2 Post_j \cdot ntreated_j + \beta_2 Post \cdot pct\_Black_p + \gamma_p + \epsilon_{pj} \tag{3.5}$$

We study impacts on total voter turnout and votes for the incumbent mayor.

Figure 3.2 visualizes the results by plotting the predicted values for each outcome against the number of informed lead households and the percent of Black residents, respectively. Panel 3.2a shows that before the crisis, voter turnout for mayoral elections sat around 15 percent for all precincts. After the crisis, turnout rose differentially in precincts with higher numbers of informed lead voters, but the increase is not statistically significant. In contrast, the divergence in voter turnout by race is much starker. Panel 3.2c shows that there is no correlation between racial composition and turnout in the pre-period. After the crisis, white voters became less likely to vote and black voters became more likely to vote.

Next, we investigate the impact of the crisis on candidate choice. Flint's mayor at the time of the crisis had been in office for two terms. A White male career politician, he nonetheless received more than half of the votes in nearly every precinct in the pre-crisis elections. Because the city was under emergency management during the crisis, he was not directly culpable; but he is strongly associated with the established system and tweeted that the water was safe to drink and that the crisis was overblown. He ran again in the first mayoral election after the crisis against a female Black newcomer who campaigned nearly exclusively on the crisis.

Again, the incumbent's vote share changed more by race than by lead test status. Panel 3.2b shows that the incumbent continued to perform well in precincts with high numbers of informed lead households, although he received a lower share of votes in precincts with fewer informed lead households. Panel 3.2d shows a stark change in his performance in Black precincts. Before the crisis, he received the majority of the vote in nearly every precinct. After, his vote share declined linearly with the Black share of the population.

Together, these results indicate that knowledge of personal damage caused by the government failure did impact voter turnout. However, the magnitude of these government performance effects is dwarfed by the magnitude of the response by race.

3.6 Conclusion

This paper contributes to the literature on voter motivation by exploiting a natural experiment in the quality of governance experienced at the individual level. We find that voters' response to their individual experience was strongly mediated by their race. We show that the water crisis led to racial divergence in turnout, new voter registration, and candidate choice. Relative turnout increased among voters living in more Black neighborhoods, new Black voters were quicker to register, and the incumbent's vote share plummeted linearly with the share of Blacks in a precinct.

The Flint water crisis has been widely understood in both government and media investigations as an extreme example of systemic government racism, permitted – if not intentionally motivated – by discrimination against Blacks (Michigan Civil Rights Commission, 2017). We consistently heard this view expressed by Flint residents in qualitative interviews in July 2017. Our results suggest that voters' interpretation of new information is strongly influenced by their race and historical context. In particular, Black voters interpret government behavior through a lens informed by the United States' long history of racist policy. Future research should explore how and why these differences in perception produce the differences in political participation we document.

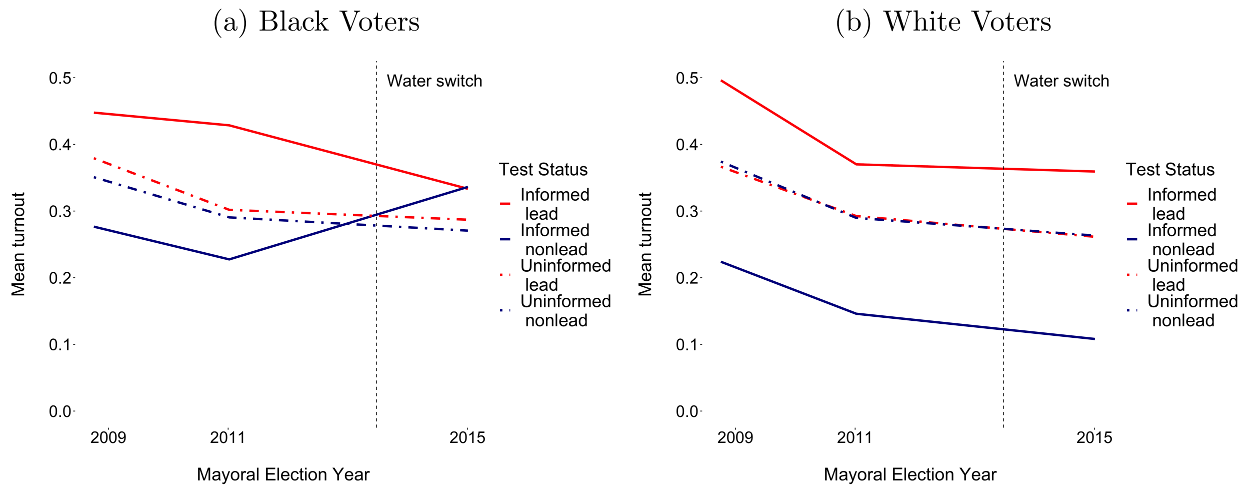
Tables and Figures

Table 3.1: Summary statistics for Flint voters who opted into lead testing

	Informed Lead			Informed Safe			Uninformed Lead			Uninformed Safe		
	Mean	s.d.	N	Mean	s.d.	N	Mean	s.d.	N	Mean	s.d.	N
Sex	0.52	0.50	800	0.50	0.50	103	0.53	0.50	14730	0.55	0.50	12540
Birthyear	1967	18.44	800	1961	16.92	103	1964	17.32	14731	1966	17.41	12540
Black	0.34	0.47	800	0.47	0.50	103	0.57	0.50	14731	0.53	0.50	12540
White	0.63	0.48	800	0.44	0.50	103	0.36	0.48	14731	0.40	0.49	12540
N neighbors informed lead in...												
0.1km	1.10	1.09	800	0.49	0.92	103	0.19	0.56	14731	0.14	0.48	12540
0.25km	2.10	2.83	800	1.28	1.92	103	1.01	2.03	14731	0.78	1.60	12540
0.5km	4.11	6.30	800	3.79	4.59	103	3.27	4.96	14731	2.69	3.89	12540
1km	8.84	10.25	800	11.38	8.76	103	10.07	9.05	14731	8.49	8.46	12540
% neighbors informed lead in...												
0.1km	4.94	4.05	800	1.55	2.73	103	0.58	1.88	14731	0.41	1.46	12540
0.25km	1.41	1.69	800	0.79	1.25	103	0.60	1.21	14731	0.46	0.95	12540
0.5km	0.75	1.11	800	0.66	0.78	103	0.56	0.88	14731	0.47	0.70	12540
1km	0.49	0.56	800	0.62	0.50	103	0.53	0.53	14731	0.49	0.54	12540
Year built of residence	1940	17.37	331	1947	14.24	98	1942	19.20	11217	1951	17.61	9266
Building value ($1000)	66.1	61.1	355	56.6	51.9	99	54.6	48.7	12217	48.9	44.4	9266
N moved residence	274	123	672	182	130	103	210	151	14711	196	139	12399
Median Income ($1000)	28.7	17.5	672	30.8	14.2	103	29.0	16.2	14711	27.2	13.6	12399
% Owners	0.50	0.22	672	0.64	0.19	103	0.56	0.20	14711	0.55	0.20	12399
% Renters	0.50	0.22	672	0.36	0.19	103	0.44	0.20	14711	0.45	0.20	12399
Median Rent	622	155	638	679	224	98	689	173	13517	677	173	11617
N White	468	452	672	516	376	103	438	387	14711	382	356	12399
% White	0.42	0.35	672	0.56	0.31	103	0.43	0.33	14711	0.38	0.32	12399
N Black	676	462	672	353	317	103	528	365	14711	578	368	12399
% Black	0.59	0.35	672	0.42	0.31	103	0.57	0.33	14711	0.62	0.32	12399
% Under 18	0.20	0.09	672	0.21	0.10	103	0.25	0.10	14711	0.26	0.10	12399
% Over 75	0.07	0.04	672	0.09	0.05	103	0.09	0.05	14711	0.09	0.06	12399
% below HS education	0.16	0.09	672	0.16	0.10	103	0.16	0.09	14711	0.17	0.09	12399
% with HS education	0.31	0.11	672	0.34	0.13	103	0.34	0.12	14711	0.35	0.11	12399
% with BA	0.08	0.06	672	0.09	0.07	103	0.09	0.07	14711	0.08	0.06	12399

This table reports summary statistics for the main sample used in the study, which includes Flint voters who received a lead test during the study period. Informed describes voters who received their test results before the 2015 mayoral election, and uninformed refers to voters who received their results after. Sex and birthyear come from the Michigan Qualified Voter File. Individual race is imputed using the R package wru based on individual Census block, name, and birthyear. Neighborhood lead test results are calculated using individuals' addresses from the voter file and residential lead test results. Year built of residence and residence building value come from City of Flint shapefiles. The remaining variables come from the public use 2014 American Community Survey (ACS).

Figure 3.1: Voter Turnout by Race and Lead Test Status



Note: This figure plots trends in voter turnout by subgroups defined by race and the timing of finding out their lead test status. Voters are labeled 'informed' if they received lead test results before the 2016 mayoral election and 'uninformed' if they received their results after. Turnout data come from the Michigan Qualified Voter file; race is imputed using public use 2014 ACS files; and lead test information comes from publicly posted residential lead test results.

Table 3.2: Difference in differences results

	Tested Positive			Tested Negative		
	All (1)	Black (2)	White (3)	All (4)	Black (5)	White (6)
Informed	0.03**	0.02	0.01	0.01	−0.05	0.11**
	(0.01)	(0.02)	(0.02)	(0.02)	(0.03)	(0.04)
Post	0.01	0.01**	−0.00	0.01**	0.02**	0.00
	(0.00)	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)
Informed·Post	0.00	−0.03	0.02	0.06	0.14**	−0.07
	(0.01)	(0.02)	(0.03)	(0.04)	(0.05)	(0.06)
Num. obs.	41517	22961	11414	32645	17357	8767
R^2 (full model)	0.51	0.53	0.51	0.48	0.49	0.50
R^2 (proj model)	0.50	0.51	0.48	0.47	0.46	0.47
Adj. R^2 (full model)	0.51	0.52	0.50	0.48	0.48	0.49
Adj. R^2 (proj model)	0.49	0.50	0.47	0.46	0.46	0.45
Num. groups: block	209	201	185	204	193	183
Dep. var. control mean	0.306	0.324	0.307	0.292	0.305	0.314

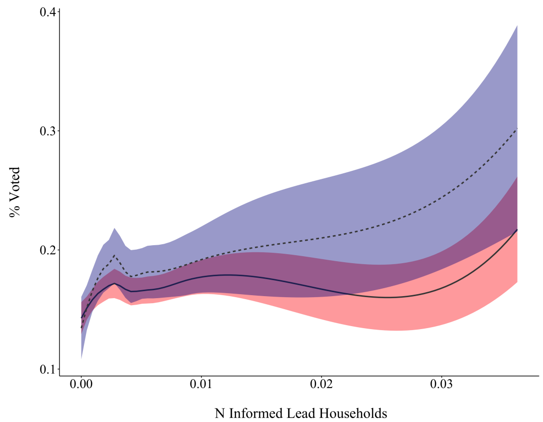
Note: This table reports results from specification 3.1 on the subsamples named in the column descriptions. The variable $Informed$ is equal to 1 for residents who received lead test results before the 2015 mayoral election. The variable *Post* is equal to one for the 2015 mayoral election, which occurred after the water crisis began. Results for residents informed of positive lead tests are shown in columns 1-3, and results for residents informed of negative lead tests are shown in columns 4-6. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$.

Table 3.3: Impact of Lead Status and Race on Time to Registration

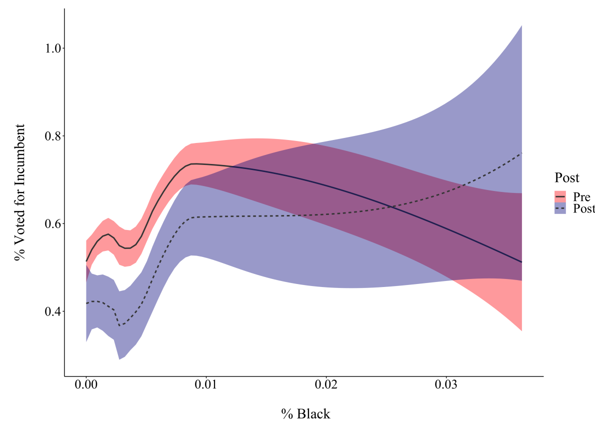
	Dep var: Days to registration	
	(1)	(2)
Positive	1.158	
	(1.150)	
Black	1.328*	
	(1.147)	
Positive·Black	0.857	
	(1.189)	
Black, Positive		1.318*
		1.140
Black, Negative		1.328*
		(0.137)
White, Positive		1.158
		(1.50)
AIC	2642.401	2643.607
R^2	0.005	0.006
Max. R^2	0.862	0.862
Num. events	1336	1336
Num. obs.	1336	1336
PH test	0.383	0.408

Both specifications use cross-sectional data on all people who were unregistered at the time of their lead test, include Census block and quarter-of-test strata, and control for individual age and sex. $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

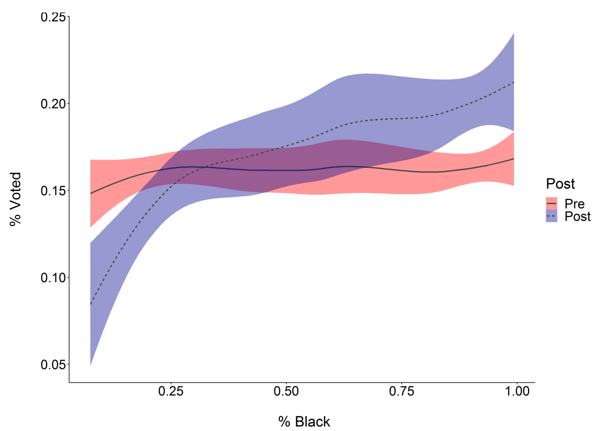
Figure 3.2: Precint-Level Results



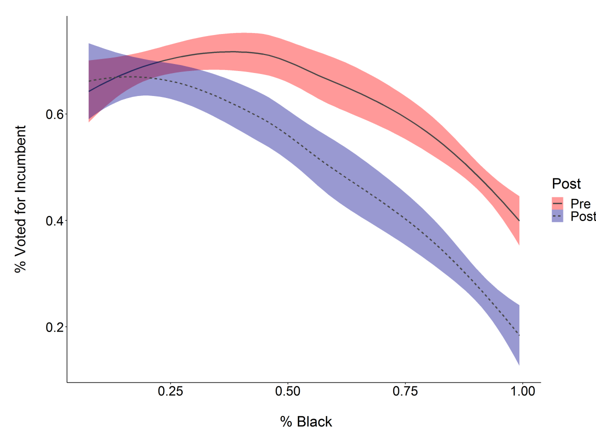
(a) Test Results and Voter Turnout



(b) Test Results and the Vote Share of the Incumbent



(c) Race and Voter Turnout



(d) Race and the Vote Share of the Incumbent

Bibliography

Ackello-Ogutu, C. (1997). Unrecorded cross-border trade between kenya and uganda implications for food security.

Aggarwal, S., Giera, B., Jeong, D., Robinson, J., and Spearot, A. (2018). Market access, trade costs, and technology adoption: evidence from northern tanzania. Technical report, National Bureau of Economic Research.

Aizer, A. and Currie, J. (2019). Lead and juvenile delinquency: new evidence from linked birth, school, and juvenile detention records. *Review of Economics and Statistics*, 101(4):575–587.

Aizer, A., Currie, J., Simon, P., and Vivier, P. (2018). Do low levels of blood lead reduce children's future test scores? *American Economic Journal: Applied Economics*, 10(1):307–41.

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 Mbiti, I. M. (2010). Mobile phones and economic development in africa. *Journal of Economic Perspectives*, 24(3):207–32.

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

Alsan, M. and Wanamaker, M. (2018). Tuskegee and the health of black men. *The quarterly journal of economics*, 133(1):407–455.

Anagol, S., Cole, S., and Sarkar, S. (2017). Understanding the advice of commissions-motivated agents: Evidence from the indian life insurance market. *Review of Economics and Statistics*, 99(1):1–15.

Anderson, J. R. and Feder, G. (2004). Agricultural extension: Good intentions and hard realities. *The World Bank Research Observer*, 19(1):41–60.

Anderson, J. R. and Feder, G. (2007). Agricultural extension. *Handbook of Agricultural Economics*, 3:2343–2378.

Arrow, K. J. (1998). What has economics to say about racial discrimination? *Journal of economic perspectives*, 12(2):91–100.

Asher, S. and Novosad, P. (2019). Socioeconomic High-resolution Rural-Urban Geographic Dataset for India (SHRUG). Technical report. https://doi.org/10.7910/DVN/DPESAK.

Ashour, M., Gilligan, D. O., Hoel, J. B., and Karachiwalla, N. I. (2019). Do beliefs about herbicide quality correspond with actual quality in local markets? evidence from uganda. *The Journal of Development Studies*, 55(6):1285–1306.

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

Atkin, D., Osman, A., and Khandelwal, A. (2017). Exporting and firm performance: Evidence from a randomized trial. *Quarterly Journal of Economics*, 132(3):551–615.

Banerjee, A., Hanna, R., Kyle, J., Olken, B. A., and Sumarto, S. (2019). Private outsourcing and competition: Subsidized food distribution in indonesia. *Journal of Political Economy*, 127(1):101–137.

Beaman, L., BenYishay, A., Mobarak, M., and Magruder, J. (2021). Can network theory based targeting increase technology adoption? *American Economic Review*, 111(6):1918–1943.

Beaman, L. and Dillon, A. (2018). Diffusion of agricultural information within social networks: Evidence on gender inequalities from mali. *Journal of Development Economics*, 133:147–161.

Becker, G. S. (1971). *The Economics of Discrimination*. University of Chicago Press.

Belloni, A., Chernozhukov, V., and Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*, 81(2):608–650.

Bensassi, S., Jarreau, J., and Mitaritonna, C. (2019). Regional integration and informal trade in africa: Evidence from benin's borders. *Journal of African Economies*, 28:89–118.

BenYishay, A., Jones, M., Kondylis, F., and Mobarak, A. M. (2020). Gender gaps in technology diffusion. *Journal of Development Economics*, 143:102380.

BenYishay, A. and Mobarak, A. M. (2018). Social learning and incentives for experimentation and communication. *The Review of Economic Studies*, 86(3):976–1009.

Bergquist, L. F. and Dinerstein, M. (2020). Competition and entry in agricultural markets: Experimental evidence from kenya. *American Economic Review*, 110(12):3705–47.

Bold, T., Kaizzi, K. C., Svensson, J., and Yanagizawa-Drott, D. (2017). Lemon technologies and adoption: measurement, theory and evidence from agricultural markets in uganda. *The Quarterly Journal of Economics*, 132(3):1055–1100.

Brenton, P., Gamberoni, E., and Sear, C. (2013). Women and trade in africa : Realizing the potential.

Casaburi, L., Kremer, M., Mullainathan, S., and Ramrattan, R. (2019). Harnessing ict to increase agricultural production: Evidence from kenya. *Working Paper*.

Chen, A., Cai, B., Dietrich, K. N., Radcliffe, J., and Rogan, W. J. (2007). Lead exposure, iq, and behavior in urban 5-to 7-year-olds: does lead affect behavior only by lowering iq? *Pediatrics*, 119(3):e650–e658.

Chen, B. K., Gertler, P. J., and Yang, C.-Y. (2016). Physician ownership of complementary medical services. *Journal of Public Economics*, 144:27–39.

Cihak, M., Mlachila, M., and Sahay, R. (2020). Race in economics. *International Monetary Fund Finance & Development Blog*, Fall.

Cole, S. A. and Fernando, A. N. (2021). Mobile'izing agricultural advice: Technology adoption, diffusion, and sustainability. *The Economic Journal*, 131(663).

Conley, T. G. and Udry, C. R. (2010). Learning about a new technology: Pineapple in ghana. *American Economic Review*, pages 35–69.

Croke, K., Mora, M. E. G., Goldstein, M., Mensah, E., and O'Sullivan, M. (2021). Up before dawn: Experimental evidence from a cross-border trader training at the drc-rwanda border. *Economic Development and Cultural Change*.

Currie, J. and Almond, D. (2011). Human capital development before age five. *Handbook of labor economics*, 4:1315–1486.

Dal Bó, E., Finan, F., Li, N. Y., and Schechter, L. (2020). Government decentralization under changing state capacity: Experimental evidence from paraguay. *forthcoming, Econometrica.*

Dar, M. H., de Janvry, A., Emerick, K., Raitzer, D., and Sadoulet, E. (2013). Flood-tolerant rice reduces yield variability and raises expected yield, differentially benefitting socially disadvantaged groups. *Scientific Reports*, 3.

Das, J., Holla, A., Mohpal, A., and Muralidharan, K. (2016). Quality and accountability in health care delivery: audit-study evidence from primary care in india. *American Economic Review*, 106(12):3765–99.

de Janvry, A., Sadoulet, E., and Suri, T. (2017). Field experiments in developing country agriculture. In *Handbook of Economic Field Experiments*, volume 2, pages 427–466. Elsevier.

Department of Elections (2017). Michigan qualified voter file. Technical report, Michigan. Accessed via FOIA request.

Downs, A. (1957). An economic theory of political action in a democracy. *Journal of Political Economy*, 65(2):135–150.

Emerick, K. and Dar, M. H. (2020). Replication data for: Farmer field days and demonstrator selection for increasing technology adoption. Technical report, Harvard Dataverse. https://doi.org/10.7910/DVN/KSPAQI.

Emerick, K. and Dar, M. H. (2021). Farmer field days and demonstrator selection for increasing technology adoption. *The Review of Economics and Statistics*, 103(4):680–693.

Emerick, K., de Janvry, A., Sadoulet, E., and Dar, M. H. (2016). Technological innovations, downside risk, and the modernization of agriculture. *American Economic Review*, 106(6):1537–1561.

Fabregas, R., Kremer, M., and Schilbach, F. (2019). Realizing the potential of digital development: The case of agricultural advice. *Science*, 366(6471).

Fafchamps, M. and Minten, B. (2012). Impact of sms-based agricultural information on indian farmers. *World Bank Economic Review*, 26(3):383–414.

Falck, O., Gold, R., and Heblich, S. (2014). E-lections: Voting behavior and the internet. *American Economic Review*, 104(7):2238–2265.

Farrington, J. (1995). The changing public role in agricultural extension. *Food Policy*, 20(6):537–544.

Feddersen, T. and Sandroni, A. (2009). The foundations of warm-glow theory. Technical report, Citeseer.

Feder, G., Just, R. E., and Zilberman, D. (1985). Adoption of agricultural innovations in developing countries: A survey. *Economic Development and Cultural Change*, pages 255–298.

Fites, D. V. (1996). Make your dealers your partners. *Harvard Business Review*, 74(2):84.

Flint Water Advisory Task Force (2016). *Flint Water Advisory Task Force Final Report*. Office of Governor Rick Snyder.

Galiani, S., Gertler, P., and Schargrodsky, E. (2005). Water for life: The impact of the privatization of water services on child mortality. *Journal of Political Economy*, 113(1):83–120.

Genesee County (2009-2015). Election results archive. Technical report, Elections Division. Link: https://www.gc4me.com/departments/county$_c lerks1/election_d ivision.php$ (accessed February 8, 2016).

Gentzkow, M. (2006). Television and voter turnout. *The Quarterly Journal of Economics*, 121(3):931–972.

Gerber, A. S. and Green, D. P. (2000). The effects of canvassing, telephone calls, and direct mail on voter turnout: A field experiment. *American political science review*, 94(03):653–663.

Geys, B. (2006). Rational theories of voter turnout: a review. *Political Studies Review*, 4(1):16–35.

Global Administrative Areas (2018). Gadm database of global administrative areas, version 3.6. Technical report, Accessed May 6, 2018. https://gadm.org/data.html.

Grant, M. and Startz, M. (2022). Cutting out the middleman: The structure of chains of intermediation. Technical report.

Griliches, Z. (1957). Hybrid corn: An exploration in the economics of technological change. *Econometrica, Journal of the Econometric Society*, pages 501–522.

Gupta, A., Ponticelli, J., and Tesei, A. (2020). Information, technology adoption and productivity: The role of mobile phones in agriculture. *Unpublished.*

Hadley, L. and Rowlatt, M. (2019). Innovating past data collection obstacles for east africa's women cross-border traders: Evidence from sauti east africa.

Hadley, L., Rowlatt, M., Li, Q., and Hakspiel, J. (2018). Needs assessment report - women in trade information platforms requirements assessment (kenya).

Hanna-Attisha, M., LaChance, J., Sadler, R. C., and Champney Schnepp, A. (2016). Elevated blood lead levels in children associated with the flint drinking water crisis: a spatial analysis of risk and public health response. *American journal of public health*, 106(2):283–290.

Hart, O., Shleifer, A., and Vishny, R. W. (1997). The proper scope of government: theory and an application to prisons. *The Quarterly Journal of Economics*, 112(4):1127–1161.

Hersh, E. D. and Nall, C. (2016). The Primacy of Race in the Geography of Income-Based Voting: New Evidence from Public Voting Records. *American Journal of Political Science*, 60(2):289–303. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ajps.12179.

Highton, B. (1997). Easy registration and voter turnout. *The Journal of Politics*, 59(02).

Hjort, J., Iyer, V., and de Rochambeau, G. (2020). Informational barriers to market access: Experimental evidence from liberian firms.

Hubbard, T. N. (1998). An empirical examination of moral hazard in the vehicle inspection market. *The RAND Journal of Economics*, pages 406–426.

Hutchings, V. L. and Valentino, N. A. (2004). The centrality of race in american politics. *Annual Review of Political Science*, 7(1):383–408.

Iacovone, L. and McKenzie, D. (2022). Shortening supply chains: Experimental evidence from fruit and vegetable vendors in bogota. *Economic Development and Cultural Change*, 71(1):111–149.

Imai, K. and Khanna, K. (2016). Improving ecological inference by predicting individual ethnicity from voter registration records. *Political Analysis*, 24(2):263–272.

Inderst, R. and Ottaviani, M. (2009). Misselling through agents. *American Economic Review*, 99(3):883–908.

Jankowski, R. (2002). Buying a lottery ticket to help the poor: Altruism, civic duty and self-interest in the decision to vote. *Rationality and Society*, 14(1):55–77.

Jankowski, R. (2007). Altruism and the decision to vote: Explaining and testing high voter turnout. *Rationality and Society*, 19(1).

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.

Jones, P. and Hudson, J. (2000). Civic duty and expressive voting: Is virtue its own reward? *Kyklos*, 53(1):3–16.

Karol, D. and Miguel, E. (2007). The electoral cost of war: Iraq casualties and the 2004 us presidential election. *Journal of Politics*, 69(3):633–648.

Klopp, J., Canagarajah, R., Wahome, L., Baraza, B., Trimble, M., Ozcan, B., Spanierman, R., Hadley, L., and Orega, D. (2022a). A cross border trade reporting experiment from kenya small-scale cross border traders, corruption and mobile phones.

Klopp, J. M., Trimble, M., and Wiseman, E. (2022b). Corruption, gender, and small-scale cross-border trade in east africa: A review. *Development Policy Review*, 40(5):e12610.

Koch, M. T. and Nicholson, S. P. (2016). Death and turnout: The human costs of war and voter participation in democracies. *American Journal of Political Science*, 60(4):932–946.

Kondylis, F., Mueller, V., and Zhu, J. (2017). Seeing is believing? evidence from an extension network experiment. *Journal of Development Economics*, 125:1–20.

Kwan, A., Boone, C. E., Sulis, G., and Gertler, P. J. (2022). Do private providers give patients what they demand, even if it is inappropriate? a randomised study using unannounced standardised patients in kenya. *BMJ open*, 12(3):e058746.

Lang, K. and Kahn-Lang Spitzer, A. (2020). Race discrimination: An economic perspective. *Journal of Economic Perspectives*, 34(2):68–89.

Levin, J. and Tadelis, S. (2010). Contracting for government services: Theory and evidence from us cities. *The Journal of Industrial Economics*, 58(3):507–541.

Little, P. D. (2005). Unofficial trade when states are weak: the case of cross-border commerce in the horn of africa.

Local Financial Stability and Choice Act (2012). Pa 436.

Magruder, J. R. (2018). An assessment of experimental evidence on agricultural technology adoption in developing countries. *Annual Review of Resource Economics*, 10:299–316.

Maitra, P., Mitra, S., Mookherjee, D., and Visaria, S. (2020). Decentralized targeting of agricultural credit programs: Private versus political intermediaries. *Unpublished*.

Mamadou, A., Osei, R. D., and Osei-Akoto, I. (2019). Impact of reinforcing agro dealer networks on agricultural productivity in niger. *Journal of Agriculture and Sustainability*, 13.

Michigan Civil Rights Commission (2017). The flint water crisis: Systemic racism through the lens of flint.

Mitra, S., Mookherjee, D., Torero, M., and Visaria, S. (2018). Asymmetric information and middleman margins: An experiment with indian potato farmers. *The Review of Economics and Statistics*, 100:1–13.

Mullainathan, S., Noeth, M., and Schoar, A. (2012). The market for financial advice: An audit study. Technical report, National Bureau of Economic Research.

Needleman, H. (2004). Lead poisoning. *Annu. Rev. Med.*, 55:209–222.

Needleman, H. L., Schell, A., Bellinger, D., Leviton, A., and Allred, E. N. (1990). The long-term effects of exposure to low doses of lead in childhood: an 11-year follow-up report. *New England journal of medicine*, 322(2):83–88.

NSSO (2021). Situation assessment of agricultural households and land and livestock holdings of households in rural india, 2019.

Odisha State Seed and Organic Products Certification Agency (2020). Ossopca seed grower list. Technical report, Accessed April 10, 2020. https://ossopca.nic.in/Publicinformation/Dist$_{g}rowerer_{R}eport.aspx$.

Office of the Attorney General (2021). Nine indicted on criminal charges in flint water crisis investigation. *State of Michigan*.

Panagopoulos, C. (2013). Positive social pressure and prosocial motivation: Evidence from a large-scale field experiment on voter mobilization. *Political Psychology*, 34(2):265–275.

Phelps, E. S. (1972). The statistical theory of discrimination. *American Economic Review*, 62(4):659–661.

Reid, O. and Weigel, J. (2019). The supply of bribes: Evidence from roadway tolls in the d.r. congo.

Ridley, G. (2016). Flint water crisis: Criminal charges against three state and city workers. *MLive*.

Rivera, W. and Alex, G. (2004). Volume 2. privatization of extension systems: Case studies of international initiatives, agriculture and rural development. Technical report, Discussion Paper 9. World Bank, Washington, DC.

Robertson, D. (2020). Flint has clean water now. why won't people drink it? *Politico*.

Romero, M., Sandefur, J., and Sandholtz, W. A. (2020). Outsourcing education: Experimental evidence from liberia. *American Economic Review*, 110(2):364–400.

Sauti-East-Africa (2017). Customer delivery report.

Sequeira, S. and Djankov, S. (2014). Corruption and firm behavior: Evidence from african ports. *Journal of International Economics*, 94:277–294.

Siu, J. (2020). Formalising informal cross-border trade: Evidence from one-stop-border-posts in uganda.

Small, M. L. and Pager, D. (2020). Sociological perspectives on racial discrimination. *Journal of Economic Perspectives*, 34(2):49–67.

Southern-Africa-Trust (2008). Optimizing regional integration in southern africa - assessing informal cross-border trade in sadc.

Startz, M. (2021). The value of face-to-face: Search and contracting problems in nigerian trade.

Stigler, G. (1961). The economics of information. *The Journal of Political Economy*, 69(3):213–225.

Stiglitz, J. (1989). Imperfect information in the product market. In Schmalensee, R. and Willig, R., editors, *Handbook of Industrial Organization*, volume 1, chapter 13, pages 769–847. Elsevier, 1 edition.

Taking Action on Flint Water (2016). Residential and business sampling. Technical report, State of Michigan. Link: https://www.michigan.gov/flintwater/sampling/residential-and-business-sampling (accessed August 26, 2016).

Titeca, K. and Kimanuka, C. (2012). *Walking in the Dark: informal Cross-border trade in the great lakes region.*

Uhlaner, C. J. and Scola, B. (2016). Collective representation as a mobilizer: Race/ethnicity, gender, and their intersections at the state level. *State Politics & Policy Quarterly*, 16(2):227–263.

Williams, J. (2021). Historical lynchings and contemporary voting behavior of blacks. *American Economic Journal: Applied Economics*.

Wolf, S., Just, D., and Zilberman, D. (2001). Between data and decisions: the organization of agricultural economic information systems. *Research Policy*, 30(1):121–141.

Wright, J. P., Dietrich, K. N., Ris, M. D., Hornung, R. W., Wessel, S. D., Lanphear, B. P., Ho, M., and Rae, M. N. (2008). Association of prenatal and childhood blood lead concentrations with criminal arrests in early adulthood. *PLoS Medicine*, 5(5):e101.

Xu, K., Xu, X., Fukao, T., Canlas, P., Maghirang-Rodriguez, R., Heuer, S., Ismail, A. M., Bailey-Serres, J., Ronald, P. C., and Mackill, D. J. (2006). Sub1a is an ethylene-response-factor-like gene that confers submergence tolerance to rice. *Nature*, 442(7103):705–708.

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

Appendix A

Border Trade and Information Frictions: Evidence from Informal Traders in Kenya

A.1 Model

A.1.1 Trader's Maximization Problem

A.1.1.1 Trader's Utility

Trader i maximizes her (risk-averse) utility by maximizing expected revenues and minimizing expected costs. Her utility is a standard profit function including a quadratic term in price gaps between selling and buying markets and in border costs. The utility for trader i, using market route k at time t, is as follows. Note that trader i's utility for market route k at time t includes quantities sold in home market h and alternative market j, conditional on using market route k (which is why quantities vary by market route k).

$$
\begin{aligned}
&MaxV_{ikt} = \\
&E[\sum_{m=h,j}\{[p^S_{mtk}(q_{imtk}) - p^B_{kt}(q_{ijtk} + q_{ihtk})] * q_{imtk} + \delta[p^S_{mtk}(q_{imtk}) - p^B_{kt}(q_{ijtk} + q_{ihtk})]^2 * q_{imtk}\} - \\
&[\delta_3 BC_{ikt}(1 + \gamma_{ikt}) + \delta_4 BC^2_{ikt} + \mu d_{ikt}] * (q_{ihtk} + q_{ijtk}) + c_{iht}q_{ihtk} + c_{ijt}q_{ijtk} + \lambda_{ik} + u_{ikt}]
\end{aligned}
\tag{A.1}
$$

Trader i has full information about random demand shocks ω_{ht} in her home market h but faces uncertainty about random demand shocks ω_{jt} in the other selling market j. Passing through the expectations, expected prices become $E[p_{jtk}] = \alpha E[\omega_{jt}]q_{ijtk}^{\frac{1}{v}}$ and $p_{htk} = \alpha\omega_{ht}q_{ihtk}^{\frac{1}{v}}$ and trader i's utility simplifies to the following:

$$MaxV_{ikt} = \\ [\alpha E[\omega_{jt}]q_{ijtk}^{\frac{1}{v}} - \delta_2(\sigma_{ijt}^{\omega})^2] * q_{ijtk} + \alpha\omega_{ht}q_{ihtk}^{\frac{1}{v}} * q_{ihtk} - [E[p_{kt}^B(q_{ijtk} + q_{ihtk})] + \delta_1(\sigma_{ikt}^B)^2] * (q_{ihtk} + \\ q_{ijtk}) - [\delta_3 E[BC_{ikt}](1 + \gamma_{ikt}) + \delta_4(\sigma_{ikt}^{BC})^2 + \mu d_{ikt}] * (q_{ihtk} + q_{ijtk}) + c_{iht}q_{ihtk} + c_{ijt}q_{ijtk} + \lambda_{ik} + u_{ikt} \tag{A.2}$$

with

- Expected revenues from selling in market h and j, conditional on using market route k: $[\alpha E[\omega_{jt}]q_{ijtk}^{\frac{1}{v}} - \delta_2(\sigma_{ijt}^{\omega})^2] * q_{ijtk} + \alpha\omega_{ht}q_{ihtk}^{\frac{1}{v}} * q_{ihtk}$
- Expected costs split between (i) purchasing costs $E[p_{kt}^B] * (q_{ihtk} + q_{ijtk})$, (ii) border costs $\delta_3 E[BC_{ikt}] + \delta_4(\sigma_{ikt}^{BC})^2$ (Tariffs if k = Uganda/Formal and Bribes if k = Uganda/Informal), (iii) bargaining power $1 + \gamma_{ikt}$ (with $\gamma_{ikt} \geq 0$) (iv) distance μd_{ikt} and (v) selling market-specific marginal cost for home market c_{ijt} and alternative market c_{iht}
- λ_{ik} utility associated with using buying market-route k. λ_{ik} includes supplier relationship, experience/comparative advantage, access to information or fixed costs
- $\delta_1 \geq 0$, $\delta_2 \geq 0$, $\delta_3 \geq 0$, $\delta_4 \geq 0$

A.1.1.2 Order of Maximization

Order of maximization (backwards induction):
Step 1: For each possible market-route k, trader chooses optimal quantities q_{ihtk}^* and q_{ijtk}^* to sell in home market h and alternative selling market j, conditional on using market route k
Step 2: Taking optimal quantity for market route k as given $q_{ikt} = q_{ihtk} + q_{ijtk}$, trader i chooses which market route k^* to use (Uganda/Formal, Uganda/Informal, Kenya/Domestic, Exit) to maximize utility V

A.1.2 Step 1: Solving for Prices and Quantities in Selling Markets

A.1.2.1 Solving for Prices

Trader i chooses q_{ihtk} and q_{ijtk} by maximizing V_{ikt}. Trader i therefore computes an optimal pair of quantities sold in home and alternative market q_{ihtk} and q_{ijtk} for each of the four alternative market routes k.
Following standard FOCs :

$$\frac{\partial V_{ikt}}{\partial q_{ijtk}} = \alpha E[\omega_{jt}] q_{ijtk}^{\frac{1}{v}} - \delta_2 (\sigma_{ijt}^S)^2 + \frac{\partial \alpha E[\omega_{jt}] q_{ijtk}^{\frac{1}{v}}}{\partial q_{ijtk}} * q_{ijtk} - E[p_{kt}^B(q_{ihtk} + q_{ijtk}, q_{-ikt})] - \delta_1 (\sigma_{ik}^B)^2 -$$
$$\frac{\partial E[p_{kt}^B(q_{ihtk} + q_{ijtk}, q_{-ikt})]}{\partial q_{ijtk}} * (q_{ijtk} + q_{ihtk}) - \delta_3 E[BC_{ikt}](1+\gamma_{ikt}) - \delta_4 (\sigma_{ikt}^{BC})^2 - \mu d_{ikt} - c_{ijt} = 0$$
$$\frac{\partial V_{ikt}}{\partial q_{ihtk}} = \alpha \omega_{ht} q_{ihtk}^{\frac{1}{v}} + \frac{\partial \alpha \omega_{ht} q_{ihtk}^{\frac{1}{v}}}{\partial q_{ihtk}} * q_{ihtk} - E[p_{kt}^B(q_{ihtk} + q_{ijtk}, q_{-ikt})] - \delta_1 (\sigma_{ik}^B)^2 -$$
$$\frac{\partial E[p_{kt}^B(q_{ihtk} + q_{ijt}, q_{-ikt})]}{\partial q_{ihtk}} * (q_{ihtk} + q_{ijtk}) - \delta_3 E[BC_{ikt}](1+\gamma_{ikt}) - \delta_4 (\sigma_{ikt}^{BC})^2 - \mu d_{ikt} - c_{iht} = 0 \tag{A.3}$$

Solving for p_{jtk} and p_{htk}:

Following the standard monopoly optimal pricing strategy, setting the mark-up over marginal costs as a function of the price elasticity of demand in the selling market, I solve for price[1] as a function of the price elasticity of demand in selling market v and price elasticity of supply (elasticity of marginal cost)[2] ϵ_{kt}^{Buy}

$$E[p_{jtk}^S] = \frac{1}{1+1/v} * [E[p_{kt}^B(q_{ihtk} + q_{ijtk})](1 + \frac{2}{\epsilon_{kt}^{Buy}}) + \delta_3 E[BC_{ikt}](1+\gamma_{ikt}) + \delta_4 (\sigma_{ikt}^{BC})^2 +$$
$$\mu d_{ikt} + c_{ijt} + \delta_1 (\sigma_{ikt}^B)^2 + \delta_2 (\sigma_{ijt}^{\omega})^2]$$
$$p_{htk}^S = \frac{1}{1+1/v} * [E[p_{kt}^B(q_{ihtk} + q_{ijtk})](1 + \frac{2}{\epsilon_{kt}^{Buy}}) + \delta_3 E[BC_{ikt}](1+\gamma_{ikt}) + \delta_4 (\sigma_{ikt}^{BC})^2 + \mu d_{ikt} +$$
$$c_{iht} + \delta_1 (\sigma_{ikt}^B)^2] \tag{A.4}$$

---

[1] See appendix for derivations of the model without demand structure and for full derivations with demand structure

[2] Again, I am assuming the partial effect on expectation of price is the same as the partial effect on price

Using the structure imposed in buying markets (equation 1.2):

$$E[p^S_{jtk}] = \frac{1}{1+1/v} * [\zeta E[b_{kt}]Q^B_{kt}(1+\frac{2}{\epsilon^{Buy}_{kt}}) + \delta_3 E[BC_{ikt}](1+\gamma_{ikt}) + \delta_4(\sigma^{BC}_{ikt})^2 + \mu d_{ikt} + c_{ijt} + \delta_1(\sigma^B_{ikt})^2 + \delta_2(\sigma^{\omega}_{ijt})^2]$$
$$p^S_{htk} = \frac{1}{1+1/v} * [\zeta E[b_{kt}]Q^B_{kt}(1+\frac{2}{\epsilon^{Buy}_{kt}}) + \delta_3 E[BC_{ikt}](1+\gamma_{ikt}) + \delta_4(\sigma^{BC}_{ikt})^2 + \mu d_{ikt} + c_{iht} + \delta_1(\sigma^B_{ikt})^2] \tag{A.5}$$

with $1/v = 1/\epsilon^{Sell}_{htk} = \frac{\partial p^S_{htk}(q_{ihtk})}{\partial q_{ihtk}} * \frac{q_{ihtk}}{p^S_{htk}} = 1/\ \epsilon^{Sell}_{jtk} = \frac{\partial p^S_{jtk}(q_{ijtk})}{\partial q_{ijtk}} * \frac{q_{ijtk}}{p^S_{jtk}}$ as I am assuming partial effect on expectation of price is the same as partial effect on price $\frac{\partial E[p^S_{jtk}(q_{ijtk})]}{\partial q_{ijtk}} * \frac{q_{ijtk}}{E[p^S_{jtk}]} = \frac{\partial p^S_{jtk}(q_{ijtk})}{\partial q_{ijtk}} * \frac{q_{ijtk}}{p^S_{jtk}}$; and $1/\epsilon^{Buy}_{kt} = \frac{\partial p^B_{kt}(q_{ijtk}+q_{ihtk})}{\partial q_{ihtk}} * \frac{q_{ihtk}}{p^B_{kt}} = \frac{\partial p^B_{kt}(q_{ijtk}+q_{ihtk})}{\partial q_{ihtk}} * \frac{q_{ijtk}}{p^B_{kt}}$

A.1.2.2 Selling Market Entry Conditions

Traders enter selling market m if their expected profits from selling in market m are positive. The entry condition for home market and for alternative market:

$$p^S_{htk}(q_{ihtk}) - E[p^B_{kt}(q_{ihtk}+q_{ijtk})] - \delta_1(\sigma^B_{ikt})^2 - \delta_3 E[BC_{ikt}](1+\gamma_{ikt}) - \delta_4(\sigma^{BC}_{ikt})^2 - \mu d_{ikt} \geq c_{iht}$$
$$E[p^S_{jtk}(q_{ijtk})] - \delta_2(\sigma^{\omega}_{ijt})^2 - E[p^B_{kt}(q_{ihtk}+q_{ijtk})] - \delta_1(\sigma^B_{ikt})^2 - \delta_3 E[BC_{ikt}](1+\gamma_{ikt}) - \delta_4(\sigma^{BC}_{ikt})^2 - \mu d_{ikt} \geq c_{ijt} \tag{A.6}$$

A.1.2.3 Solving for Quantities

Using the price function (1.1) and the optimal price expressions from the optimization (A.4), I solve for quantities, including market entry conditions (A.6):

$$q_{ihtk} = [\frac{1}{\alpha\omega_{htk}} * \frac{1}{1+1/v} * [E[p^B_{kt}(q_{ihtk}+q_{ijtk})](1+\frac{2}{\epsilon^{Buy}_{kt}}) + \delta_3 E[BC_{ikt}](1+\gamma_{ikt}) + \delta_4(\sigma^{BC}_{ikt})^2 + d_{ikt} + c_{iht} + \delta_1(\sigma^B_{ikt})^2]^v \tag{A.7}$$

with $\alpha\omega_{ht}q^{1/v}_{htk} - E[p^B_{kt}(q_{ihtk}+q_{ijtk})] - \delta_1(\sigma^B_{ikt})^2 - \delta_3 E[BC_{ikt}](1+\gamma_{ikt}) - \delta_4(\sigma^{BC}_{ikt})^2 - \mu d_{ikt} \geq c_{iht}$

$$q_{ijt} = \begin{cases} 0 \text{ if } E[\omega_{jt}]\alpha_{jt}q_{ijtk}^{1/v} - E[p_{kt}^B(q_{ijtk}+q_{ihtk})] - \delta_1(\sigma_{ikt}^B)^2] - \delta_3 E[BC_{ikt}](1+\gamma_{ikt}) - \\ \delta_4(\sigma_{ikt}^{BC})^2 - \mu d_{ikt} - \delta_2(\sigma_{ijt}^{\omega})^2 < c_{ijt} \\ \\ [\frac{1}{\alpha E[\omega_{jt}]} * \frac{1}{1+\frac{1}{v}} * [E[p_{kt}^B(q_{ihtk}+q_{ijtk})](1+\frac{2}{\epsilon_{kt}^{Buy}}) + \delta_3 E[BC_{ikt}](1+\gamma_{ikt}) + \delta_4(\sigma_{ikt}^{BC})^2 + \\ \mu d_{ikt} + c_{ijt} + \delta_1(\sigma_k^B)^2 + \delta_2(\sigma_{ijt}^{\omega})^2]^v \quad \text{ if } E[\omega_{jt}]\alpha_{jt}q_{ijt}^{1/v} - E[p_{kt}^B(q_{ihtk}+q_{ijtk})] - \\ \delta_1(\sigma_{ikt}^B)^2 - \delta_3 E[BC_{ikt}](1+\gamma_{ikt}) - \delta_4(\sigma_{ikt}^{BC})^2 - \mu d_{ikt} - \delta_2(\sigma_{ijt}^{\omega})^2 \geq c_{ijt} \end{cases} \tag{A.8}$$

Using structure put on selling markets in equation 1.2:

$$q_{ihtk} = [\frac{1}{\alpha\omega_{htk}} * \frac{1}{1+1/v} * [\zeta E[b_{kt}]Q_{kt}^B](1+\frac{2}{\epsilon_{kt}^{Buy}}) + \delta_3 E[BC_{ikt}](1+\gamma_{ikt}) + \delta_4(\sigma_{ikt}^{BC})^2 + d_{ikt} + c_{iht} + \delta_1(\sigma_{ikt}^B)^2]^v \tag{A.9}$$

with $\alpha\omega_{ht}q_{htk}^{1/v} - [\zeta E[b_{kt}]Q_{kt}^B] - \delta_1(\sigma_{ikt}^B)^2 - \delta_3 E[BC_{ikt}](1+\gamma_{ikt}) - \delta_4(\sigma_{ikt}^{BC})^2 - \mu d_{ikt} \geq c_{iht}$

$$q_{ijt} = \begin{cases} 0 \text{ if } E[\omega_{jt}]\alpha_{jt}q_{ijtk}^{1/v} - [\zeta E[b_{kt}]Q_{kt}^B] - \delta_1(\sigma_{ikt}^B)^2] - \delta_3 E[BC_{ikt}](1+\gamma_{ikt}) - \\ \delta_4(\sigma_{ikt}^{BC})^2 - \mu d_{ikt} - \delta_2(\sigma_{ijt}^{\omega})^2 < c_{ijt} \\ \\ [\frac{1}{\alpha E[\omega_{jt}]} * \frac{1}{1+1/v} * [\zeta E[b_{kt}]Q_{kt}^B](1+\frac{2}{\epsilon_{kt}^{Buy}}) + \delta_3 E[BC_{ikt}](1+\gamma_{ikt}) + \delta_4(\sigma_{ikt}^{BC})^2 + \\ \mu d_{ikt} + c_{ijt} + \delta_1(\sigma_k^B)^2 + \delta_2(\sigma_{ijt}^{\omega})^2]]^v \quad \text{ if } E[\omega_{jt}]\alpha_{jt}q_{ijt}^{1/v} - [\zeta E[b_{kt}]Q_{kt}^B] - \\ \delta_1(\sigma_{ikt}^B)^2 - \delta_3 E[BC_{ikt}](1+\gamma_{ikt}) - \delta_4(\sigma_{ikt}^{BC})^2 - \mu d_{ikt} - \delta_2(\sigma_{ijt}^{\omega})^2 \geq c_{ijt} \end{cases} \tag{A.10}$$

A.1.3 Step 2: Choosing Buying Market and Route

A.1.3.1 Choice model

Trader i will compare her utility across each market route, taking the optimal quantity for each route as given.

Trader i will pick buying market route k' iff $V_{ik't} \geq V_{ikt}$

$$[E[p^S_{jtk'}(q_{ijtk'})] - \delta_2(\sigma^{\omega}_{ijt})^2] * q_{ijtk'} + p^S_{htk'}(q_{ihtk'}) * q_{ihtk'} - E[p^B_{k't}(q_{ihtk'} + q_{ijtk'})] - \delta_1(\sigma^B_{ik't})^2 - \delta_3 BC_{ik't}(1 + \gamma_{ik't}) - \delta_4(\sigma^{BC}_{ik't})^2 - \mu d_{ik't} + \lambda_{ik'} \geq [E[p^S_{jtk}(q_{ijtk})] - \delta_2(\sigma^{\omega}_{ijt})^2] * q_{ijtk} + p^S_{htk}(q_{ihtk}) * q_{ihtk} - E[p^B_{kt}(q_{ihtk} + q_{ijtk})] - \delta_1(\sigma^B_{ikt})^2 - \delta_3 BC_{ikt}(1 + \gamma_{ikt}) - \delta_4(\sigma^{BC}_{ikt})^2 - \mu d_{ikt} + \lambda_{ik} \tag{A.11}$$

So, increased profits from lower marginal costs in a new buying market route need to be larger than the lost utility from switching market routes $\lambda_{ik} - \lambda_{ik'}$.

There is a $\bar{\lambda}_i$, at which $V_{ik't} = V_{ikt}$. And if $\lambda_{ik} - \lambda_{ik'} \leq \bar{\lambda}_i$, trader i switches to k'. $\lambda_{ik} = \hat{\lambda_k} + \lambda'_{ik}$ with λ'_{ik} being an unobserved/random term that follows an extreme value distribution.

A.1.3.2 Choice Probabilities

Using a Mixed Logit Model, the probability of choosing buying market route k is :

$Prob(Y_{it=k}) = \int \frac{exp(V_{ikt}(\beta))}{\Sigma exp(V_{ikt}(\beta))} * f(\beta|\theta) * d\beta$ with β coefficients in V and θ parameters for the mixing distribution, estimated through simulations.

A.2 Model under Cournot competition

In this section, I relax the assumption of firms being monopolies and instead solve for Cournot competition (and perfect competition). I also relax the second order effect of the increase in quantity in market h on the marginal cost of purchasing goods for market j (the direct effect of an increase in quantity in market h on the cost of purchasing goods for market j remains).[3]

In this setting, there are N identical firms selling a homogeneous good. $P_{mtk} = \alpha * Q_{mtk}^{1/v}$ with Q being the sum of all individual firms quantities sold in market m. Each firm chooses quantity, taking as given the quantity of other firms (i.e. taking into account other firms' maximized quantity).

Following standard FOCs :

$$\begin{aligned}
\frac{\partial V_{ikt}}{\partial q_{ijtk}} = \alpha E[\omega_{jt}] Q_{jtk}^{\frac{1}{v}} - \delta_2 (\sigma_{ijt}^S)^2 + \frac{\partial \alpha E[\omega_{jt}] q_{ijtk}^{\frac{1}{v}}}{\partial Q_{jtk}} * q_{ijtk} - E[p_{kt}^B(q_{ihtk} + q_{ijtk}, q_{-ikt})] - \delta_1 (\sigma_{ik}^B)^2 - \\
\frac{\partial E[p_{kt}^B(q_{ihtk} + q_{ijtk}, q_{-ikt})]}{\partial q_{ijtk}} * (q_{ijtk}) - \delta_3 E[BC_{ikt}](1+\gamma_{ikt}) - \delta_4 (\sigma_{ikt}^{BC})^2 - \mu d_{ikt} - c_{ijt} = 0 \\
\frac{\partial V_{ikt}}{\partial q_{ihtk}} = \alpha \omega_{ht} Q_{htk}^{\frac{1}{v}} + \frac{\partial \alpha \omega_{ht} Q_{htk}^{\frac{1}{v}}}{\partial q_{ihtk}} * q_{ihtk} - E[p_{kt}^B(q_{ihtk} + q_{ijtk}, q_{-ikt})] - \delta_1 (\sigma_{ik}^B)^2 - \\
\frac{\partial E[p_{kt}^B(q_{ihtk} + q_{ijt}, q_{-ikt})]}{\partial q_{ihtk}} * (q_{ihtk}) - \delta_3 E[BC_{ikt}](1+\gamma_{ikt}) - \delta_4 (\sigma_{ikt}^{BC})^2 - \mu d_{ikt} - c_{iht} = 0
\end{aligned} \tag{A.12}$$

Since $q_1^* = q_{*2} = ... = q_N^* \implies Q_{htk} = N q_i$ and $Q_{jtk} = N q_i$

Substituting this in the first order conditions to get optimal quantities:
$q_{ijtk}^* = \frac{1}{N}[MC(q_{ijtk}) * \frac{1}{\alpha} * \frac{1}{1+\frac{1}{Nv}}]^v$
$q_{ihtk}^* = \frac{1}{N}[MC(q_{ihtk}) * \frac{1}{\alpha} * \frac{1}{1+\frac{1}{Nv}}]^v$

with $MC(q_{ijtk}) = \delta_2 (\sigma_{ijt}^S)^2 + E[p_{kt}^B(q_{ihtk} + q_{ijtk}, q_{-ikt})] + \delta_1 (\sigma_{ik}^B)^2 + \frac{\partial E[p_{kt}^B(q_{ihtk}+q_{ijtk}, q_{-ikt})]}{\partial q_{ijtk}} * (q_{ijtk})$
$+ \delta_3 E[BC_{ikt}](1+\gamma_{ikt}) + \delta_4 (\sigma_{ikt}^{BC})^2 + \mu d_{ikt} + c_{ijt}$

and $MC(q_{ihtk}) = E[p_{kt}^B(q_{ihtk} + q_{ijtk}, q_{-ikt})] + \delta_1 (\sigma_{ik}^B)^2 + \frac{\partial E[p_{kt}^B(q_{ihtk}+q_{ijt}, q_{-ikt})]}{\partial q_{ihtk}} * (q_{ihtk})$
$+ \delta_3 E[BC_{ikt}](1+\gamma_{ikt}) + \delta_4 (\sigma_{ikt}^{BC})^2 + \mu d_{ikt} + c_{iht} = 0$

[3]This simplification was done to simplify the algebra but does not affect the predictions or results.

Solving for market quantities:

$Q_{jtk} = N * q^*_{ijtk} = [MC(q_{ijtk}) * \frac{1}{\alpha} * \frac{1}{1+\frac{1}{Nv}}]^v$

$Q_{htk} = N * q^*_{ihtk} = [MC(q_{ihtk}) * \frac{1}{\alpha} * \frac{1}{1+\frac{1}{Nv}}]^v$

Solving for the associated market prices :

$P_{jtk} = MC(q_{ijtk}) * \frac{1}{1+\frac{1}{Nv}}$

$P_{htk} = MC(q_{ihtk}) * \frac{1}{1+\frac{1}{Nv}}$

Comparing Cournot solutions to monopolies

Not surprisingly, each firm's quantity under Cournot is smaller than the monopoly's quantity. However, market level quantity under Cournot is larger than the monopoly's quantity. Prices in Cournot are lower than under the monopoly's assumption.

Perfect competition

As N become large, the Cournot solution approximates perfect competition. Indeed as N increases:

Solving for market quantities:

$Q_{jtk} = [MC(q_{ijtk}) * \frac{1}{\alpha}]^v$

$Q_{htk} = [MC(q_{ihtk}) * \frac{1}{\alpha}]^v$

Solving for the associated market prices :

$P_{jtk} = MC(q_{ijtk})$

$P_{htk} = MC(q_{ihtk})$

Note that we find that MC equals price, common to a competitive equilibrium.

A.3 Tables

Table A.3.1: Business Registration

	mean
Personal KRA Pin Number	0.40
Business KRA Pin Number	0.02
Personal and Business KRA Pin Number	0.01
No KRA Pin Number	0.57
Observations	954

Table A.3.2: Traders' Costs by crossing type - Audit Study Experiment

	Official Crossing mean	Informal Crossing (Marachi) mean
Value of good (Kshs '00)	107.79	144.53
Costs (Kshs '00)		
Crossing Bribes	1.09	1.57
Total Crossing Costs (excl.transport)	1.94	1.81
Crossing Transport Costs	0.08	0.65
Experience		
Waiting Time	20.66	6.81
N border agents faced	1.76	1.72
N trucks at crossing	10.61	0.51
N traders at crossing	16.74	8.12
Filled certificate	0.03	0.00
Observations	38	209

Table A.3.3: Costs by trader size

	(1) Sales Kshs
Total purchase costs	1.135***
	[0.026]
Total costs	0.728**
	[0.351]
Constant	12997.709***
	[3491.072]
Dep Var Mean	1.63e+05
R-Squared	.835
Observations	826

Note: Standard errors robust (reported in brackets).
* $p<0.1$, ** $p<0.05$, *** $p<0.01$.
Data from second baseline

Table A.3.4: RCT: balance table

	Control	Treatment	P-value of Diff
Baseline			
Gender	0.20	0.18	0.406
Age	39.58	37.09	0.380
N formal associations	0.56	0.61	0.313
N informal associations	1.24	1.22	0.847
Trading main source of income (past 12 months)	0.96	0.94	0.178
Inventory	0.44	0.42	0.637
Value of Inventory	16626.14	16048.02	0.851
Domestic	0.46	0.44	0.417
Official	0.21	0.17	0.055*
Informal	0.32	0.39	0.015**
Updated Baseline			
Attrition Midline	0.17	0.20	0.183
Out of business Midline	0.04	0.04	0.965
Domestic (Midline)	0.73	0.70	0.265
Official (Midline)	0.08	0.08	0.684
Informal (Midline)	0.16	0.20	0.113

Note: * p<0.1, ** p<0.05, *** p<0.01.

Table A.3.5: Attrition Balance

	(1) Midline	(2) Round 1	(3) Round 2	(4) Round 3	(5) Endline
Treatment	-0.029	0.015	0.001	0.027	0.027
	[0.023]	[0.023]	[0.023]	[0.023]	[0.024]
Dep Var Mean	0.818	0.811	0.808	0.810	0.785
Observations	1166	1165	1166	1166	1166

* p<0.1, ** p<0.05, *** p<0.01.

Table A.3.6: First Stage: Number of alternatives looked up in platform

	Market Prices		Exchange Rates		Trade Procedure		Weather	
	mean	sd	mean	sd	mean	sd	mean	sd
N alternatives by session	1.75	1.50	1.34	0.65	1.16	0.53	1.25	0.87
N alternatives by day	1.87	1.70	1.44	0.78	1.23	0.67	1.40	1.11
N alternatives by month	2.71	3.35	2.21	2.02	1.49	1.22	4.18	6.11
N alternatives overall	4.67	7.61	3.86	5.39	2.01	2.25	9.16	15.85
N unique alternatives by session	1.75	1.50	1.34	0.65	1.16	0.53	1.25	0.87
N unique alternatives by day	1.85	1.65	1.41	0.74	1.19	0.57	1.30	0.99
N unique alternatives by month	2.46	2.58	1.96	1.55	1.31	0.73	1.87	2.21
N unique alternatives overall	3.75	4.61	2.90	3.55	1.50	0.92	2.77	3.53

Table A.3.7: RCT Results: Switching to new markets

	Ratio upd. baseline selling markets		Ratio upd. baseline buying markets	
	(1) Round 1-3	(2) Endline	(3) Round 1-3	(4) Endline
Treatment	-0.032*	-0.067**	0.032	0.019
	[0.019]	[0.026]	[0.023]	[0.033]
Dep Var Mean (Control)	0.697	0.798	0.368	0.383
R-Squared	.305	.008	.093	0
Pre-Period				
Round FE	X		X	
Observations	7963	846	7579	837

Note: Standard errors robust or clustered as trader level (reported in brackets).
* p<0.1, ** p<0.05, *** p<0.01.

Table A.3.8: RCT: Sales - Levels

	Sales		Profits		Stock	
	(1) Rounds 1-3	(2) Endline	(3) Rounds 1-3	(4) Endline	(5) Rounds 1-3	(6) Endline
Treatment	4984.250**	14402.266**	712.951**	2283.362*	-618.428	927.209
	[2068.393]	[6614.729]	[356.012]	[1177.428]	[863.044]	[1213.227]
Dep Var Mean (Control)	24616.063	68568.821	4270.341	11376.329	8883.181	8897.763
R-Squared	.008	.005	.007	.004	.001	.001
Pre-Period						
RoundFE	X		X		X	
Observations	2790	898	2792	895	2806	906

Note: Standard errors robust or clustered as trader level (reported in brackets).
* p<0.1, ** p<0.05, *** p<0.01.

Table A.3.9: RCT: Costs - Levels

	Total		Per Sales	
	(1) Purch. Costs	(2) Oth. Costs	(3) Purch. Costs	(4) Oth. Costs
Treatment	3759.665**	196.612	0.038	-0.010
	[1760.577]	[146.113]	[0.046]	[0.007]
Dep Var Mean (Control)	22003.278	1904.950	1.033	0.120
R-Squared	.006	.004	.001	.002
Pre-Period				
Round FE	X	X	X	X
Observations	2795	2799	2402	2398

Note: Standard errors robust or clustered as trader level (reported in brackets).
Similar results with costs in levels
* p<0.1, ** p<0.05, *** p<0.01.

Table A.3.10: Reasons for staying domestic : Updated Baseline and Endline

	Updated Baseline mean	Endline mean
Lack of contacts	0.14	0.18
Better prices in dom. markets	0.47	0.44
Expensive tarrifs and fees	0.16	0.09
Difficult Importing/Exporting process	0.20	0.27
Not enough knowledge on procedures	0.16	0.11
Covid 19	0.12	0.12
Border Closure	0.09	0.04
Existing ties with suppliers	0.19	0.36
Observations	613	297

Table A.3.11: Types of traders: baseline-updated baseline-endline

CB-CB-CB	0.02
CB-Dom-CB	0.06
CB-DomCB-CB	0.09
CB-CB-Dom	0.00
CB-Dom-Dom	0.19
CB-DomCB-Dom	0.04
Dom-CB-CB	0.00
Dom-Dom-CB	0.01
Dom-DomCB-CB	0.01
Dom-CB-Dom	0.00
Dom-Dom-Dom	0.30
Dom-DomCB-Dom	0.01
Complete Exit During/Post BC	0.27
Observations	1137

Table A.3.12: Selling market prices (Reported Data)

	(1) Price Levels	(2) Price Logs	(3) CPI	(4) Log CPI	(5) CPI	(6) Log CPI
Intensity Treat x Post	-13.312	-0.234***	-15.527	-0.623	-2.335*	-0.193***
	[10.398]	[0.076]	[16.115]	[0.639]	[1.386]	[0.067]
Intensity Treat	9.988	0.174**	15.063	0.358	1.799	0.044
	[11.491]	[0.080]	[15.428]	[0.600]	[1.206]	[0.043]
Post	4.925	0.116***	-0.307	-0.222	1.048	0.110**
	[3.928]	[0.031]	[6.636]	[0.304]	[0.878]	[0.049]
Dependent Variable Control Mean	77.419	4.035	22.874	1.532	2.497	0.191
R-Squared	.711	.716	.024	.049	.01	.041
Market Cluster	X	X				
Product FE	X	X				
Market Weight			X	X		
Area Weight			.		X	X
Observations	1223	1266	232	232	250	250

Note: Standard errors robust or clustered at market level (reported in brackets).
* p<0.1, ** p<0.05, *** p<0.01.
Markets outside sampleframe included, updated baseline, HF rounds and endline used.

Table A.3.13: Buying market prices (Reported Data)

	(1) Price Levels	(2) Price Logs	(3) CPI	(4) Log CPI	(5) CPI	(6) Log CPI
Intensity Treat x Post	-2.926	0.011	0.570	0.066	-0.468	-0.123
	[10.101]	[0.132]	[8.087]	[0.581]	[1.457]	[0.139]
Intensity Treat	3.656	-0.017	9.723	0.642	-1.100	-0.019
	[12.498]	[0.150]	[7.396]	[0.522]	[1.345]	[0.123]
Post	3.356	0.087	0.559	0.132	0.297	0.090
	[4.275]	[0.067]	[3.279]	[0.255]	[0.856]	[0.068]
Dependent Variable Control Mean	58.326	3.734	12.733	1.215	3.750	0.318
R-Squared	.71	.715	.073	.052	.036	.026
Market Cluster	X	X				
Product FE	X	X				
Market Weight			X	X		
Area Weight					X	X
Observations	797	812	236	236	263	263

Note: Standard errors robust or clustered at market level (reported in brackets).
* p<0.1, ** p<0.05, *** p<0.01.
Markets outside sampleframe included; updated baseline, HF rounds and endline used.

A.4 Figures

Figure A.4.1: Definition of informal traders

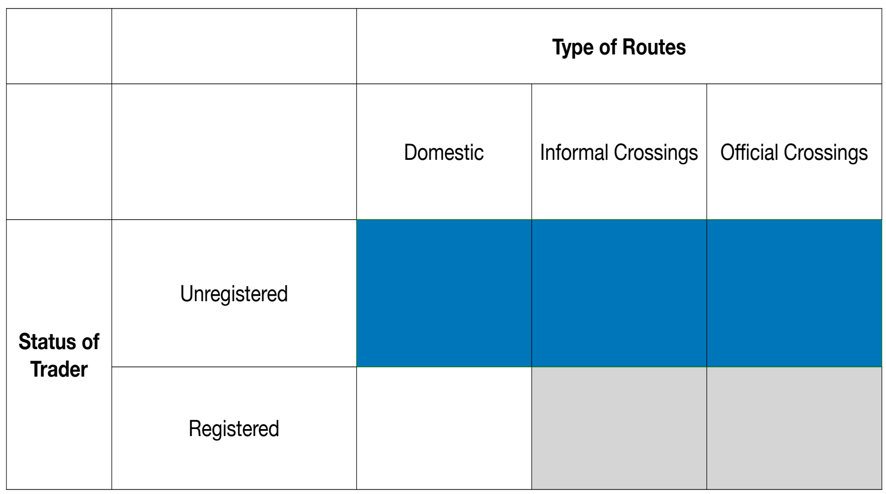


Figure A.4.2: Attrition across rounds

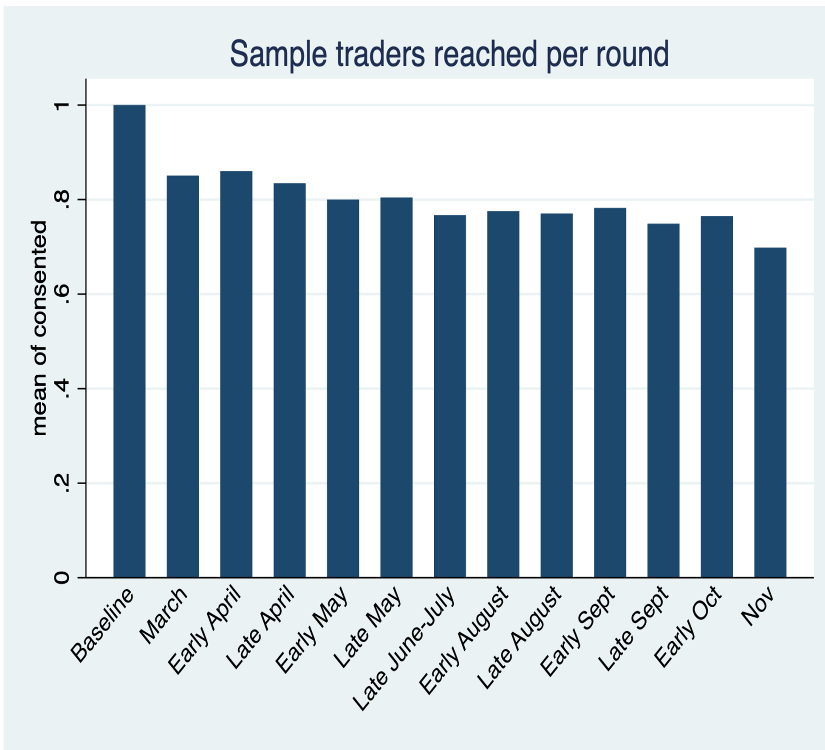


Figure A.4.3: Sample Composition

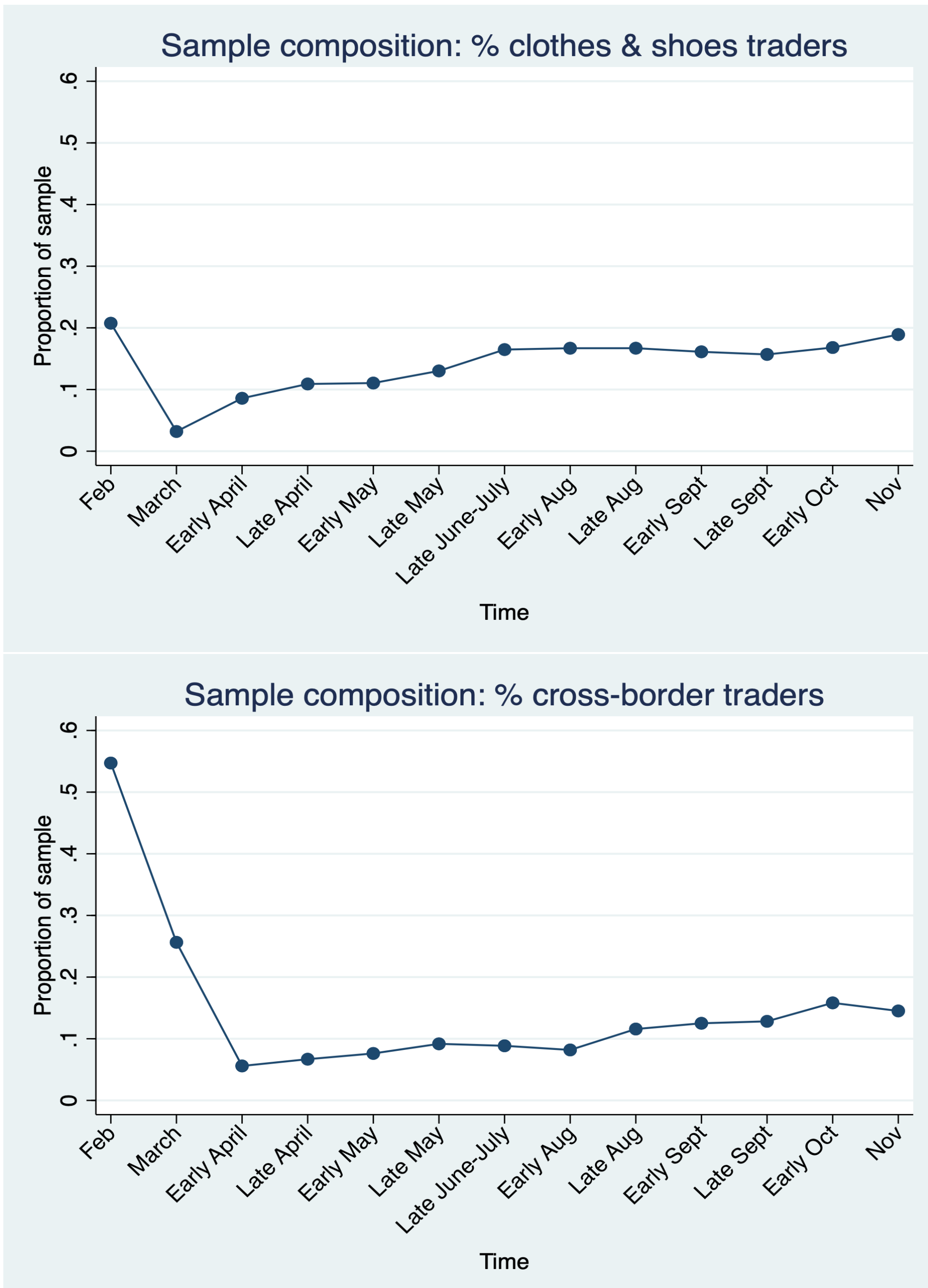


Figure A.4.4: Determinants of being Out of business by industry and trader type

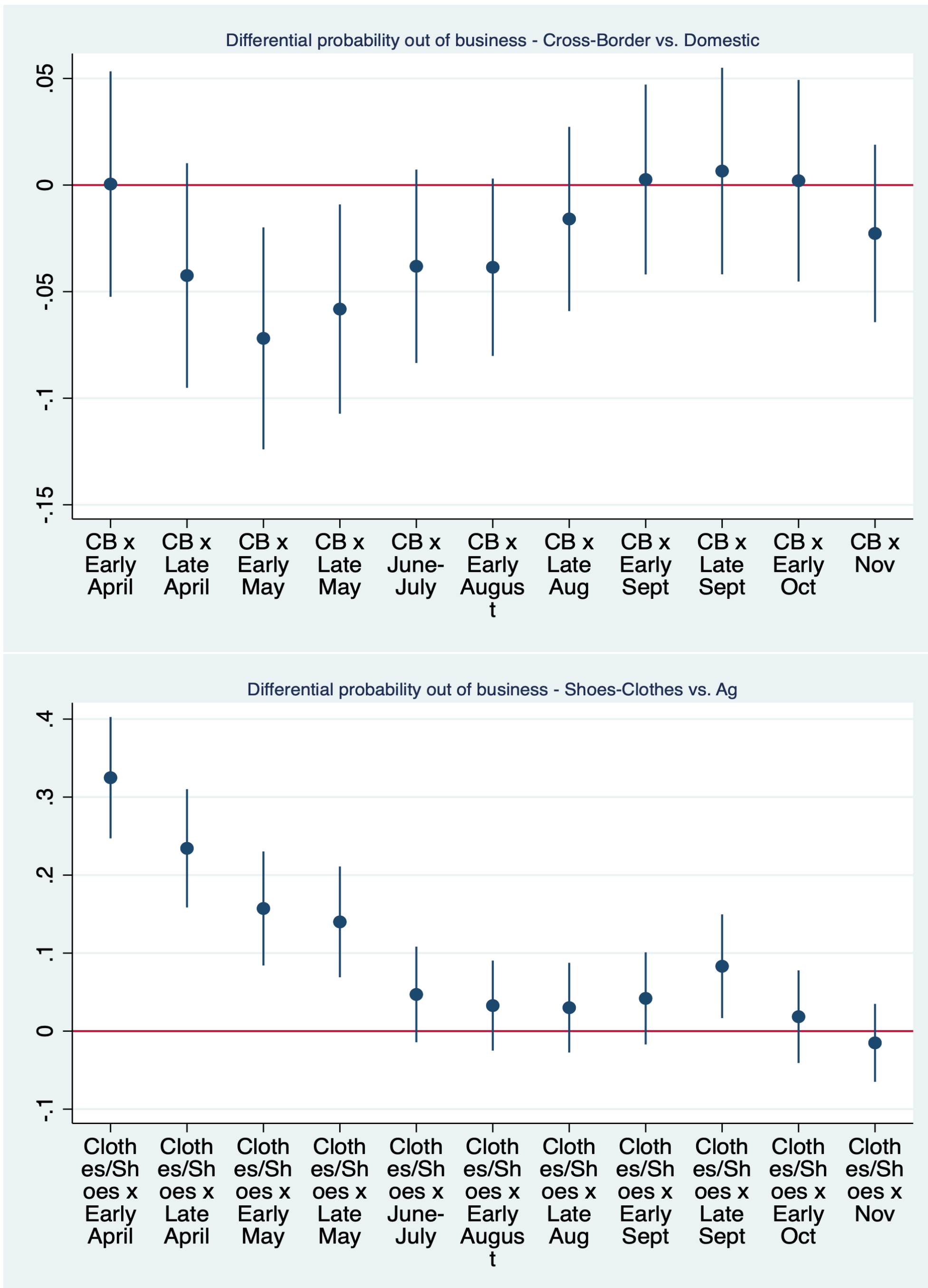


Figure A.4.5: Platform

Figure A.4.6: Increase in corruption and harassment during border closure

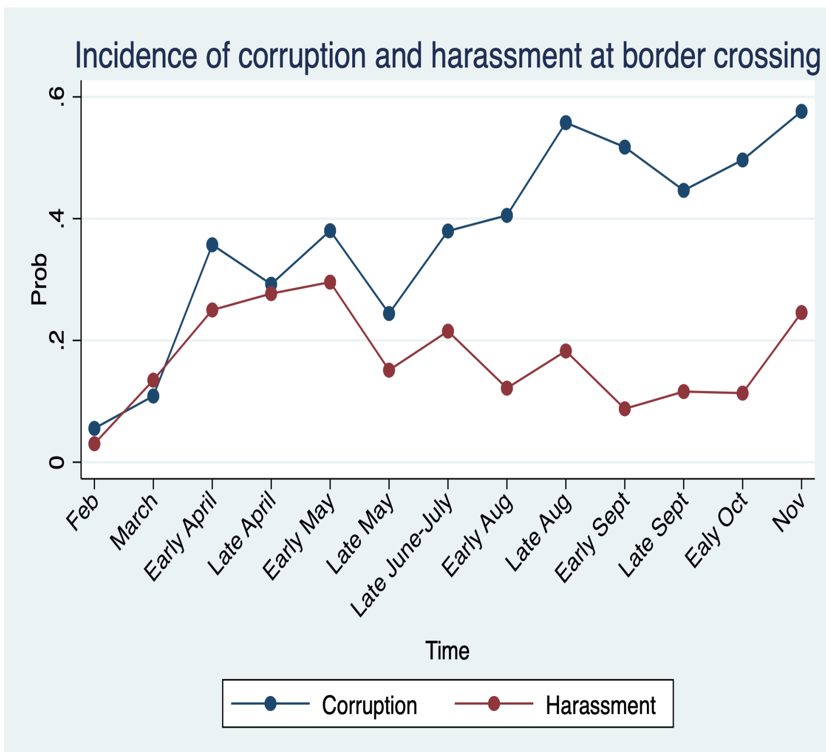
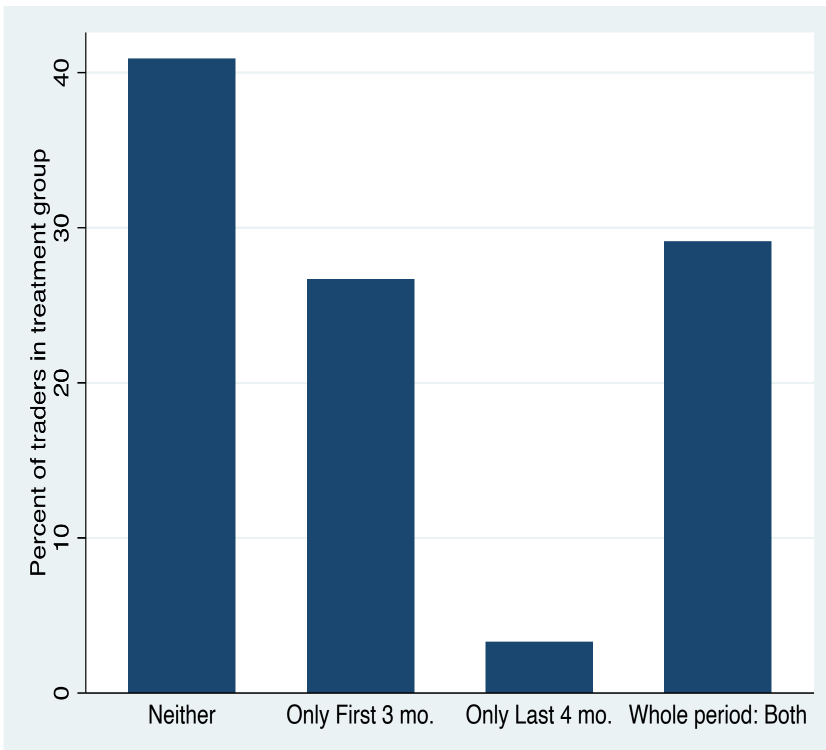


Figure A.4.7: First Stage: Distribution of usage early months versus late months



Appendix B

Private Input Suppliers as Information Agents for Technology Adoption in Agriculture

B.1 Appendix

Table B.1.1: Distance to the nearest interventions by treatment status

	(1) Govt. Extension	(2) Treated Dealers
Treatment	6.5386***	-7.3254***
	(0.9246)	(1.6138)
District FE	Yes	Yes
Mean in Control	4.922	11.688
Number of Observations	5536	5536
R squared	0.426	0.278

The dependent variables are distances (measured in km) between the farmer's house and the nearest activities supported by the research. Column 1 uses the distance between the farmer's home and the nearest Swarna-Sub1 cultivation through the government extension in the control blocks (any of the seeds distributed through the BAO or the farmer field day). Column 2 uses the distance between the farmer's home and the nearest dealer that was provided seeds. The coefficients in the table verify that farmers in treated blocks were further from the government extension activities and closer to dealers that were provided seeds. For instance, the coefficient in column 2 indicates that farmers in treated blocks were 7.32 km closer to the nearest dealer receiving seeds. Farmers in control blocks were 11.7 km from the nearest treated dealer (the control mean). This falls by 7.32 km in the treated blocks. The standard errors in both columns are clustered at the block level. Asterisks indicate that coefficient is statistically significant at the 1% ***, 5% **, and 10% * levels.

Table B.1.2: Relationship between treatment assignment, non-response, and growing rice among the sample of farmers

	(1) Not Surveyed	(2) Grows Rice
Treatment	-0.010	0.024
	(0.013)	(0.015)
Dependent Variable Control Mean	0.079	0.920
R-Squared	0.043	0.011
District Fixed Effects	X	X
Observations	7200	6653

The table shows the difference in the rate of non-response and currently growing rice across treatment and control groups. All regressions use the data from the follow-up survey with farmers in August/September 2017. The dependent variables are indicator variables for not being surveyed (column 1) and an indicator for growing rice during the 2017 season (column 2). The standard errors in each regression are clustered at the block level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Table B.1.3: Relationship between treatment assignment, non-response, and selling seeds among the sample of dealers

	(1) Located	(2) In Business
Dealer-Based Extension	0.035	-0.041
	(0.046)	(0.055)
Dependent Variable Control Mean	0.745	0.610
R-Squared	0.316	0.050
District Fixed Effects	X	X
Observations	613	473

The table shows the difference in the rate that dealers were not surveyed (column 1) and the difference in being in the rice seed business (column 2) across treatment and control groups. All regressions use the data from the follow-up survey with dealers around September 2017. The dependent variables are indicator variables for not being surveyed (column 1) and an indicator for currently selling rice seeds among those surveyed (column 2). The standard errors in each regression are clustered at the block level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Table B.1.4: Correlation between flood exposure and socioeconomic characteristics

	(1)	(2)	(3)	(4)	(5)	(6)
Area Cultivated (Acres)	-1.104	-0.593**				
	(0.666)	(0.242)				
Below Poverty Line Card			3.269	1.987		
			(2.016)	(1.451)		
Scheduled Tribe or Caste					4.183*	5.136**
					(2.113)	(2.366)
Dependent Variable Control Mean	16.075	16.075	17.374	17.374	17.353	17.353
R-Squared	.004	.129	.002	.142	.001	.144
District Fixed Effects		X		X		X
Observations	5134	5134	5521	5521	5529	5529

The table shows the relationship between flood exposure and household characteristics from the 2017 survey. The dependent variable in all regressions is the total number of days of flood exposure during the growing seasons from 2011-2017, measured by matching satellite data to the GPS coordinates of the household. All standard errors are clustered at the block level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Table B.1.5: Tobit estimates of extensive and intensive margins of adoption

	(1) Adoption	(2) Acres for Adopters
Treatment	0.034*	0.117*
	(0.019)	(0.061)
Dependent Variable Control Mean	0.063	1.470
District Fixed Effects	X	X
Observations	6653	6653

The table shows marginal effects from Tobit regressions of area cultivated with Swarna-Sub1 on strata fixed effects and treatment. All regressions use the data from the follow-up survey with farmers in August/September of 2017. Both columns show average marginal effects and delta-method standard errors. Column 1 shows the marginal effect on the probability of adoption, while column 2 shows the marginal effect on acreage cultivated, conditional on adoption. The standard errors in each regression are clustered at the block level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Table B.1.6: Effects on learning-related outcomes

	(1) Extension Contact	(2) Saw Demonstration	(3) Learned during last 24 months
Treatment	0.013	0.003	0.018
	(0.010)	(0.012)	(0.017)
Dependent Variable Control Mean	0.057	0.043	0.090
R-Squared	0.016	0.031	0.191
District Fixed Effects	X	X	X
Observations	6120	6653	6653

The table shows treatment effects on contact with extension workers and learning about Swarna-Sub1. All regressions use the data from the follow-up survey with farmers in August/September of 2017. The dependent variables are an indicator for whether the farmer had any contact with the Village Agricultural Worker during the last year (column 1), whether the farmer had seen a demonstration of a new seed variety (column 2), and whether the farmer had learned about Swarna-Sub1 in the last 24 months (column 3). The standard errors in each regression are clustered at the block level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Table B.1.7: Effects on adoption of different rice varieties

	Control Mean	Estimate
Pooja	0.376	0.011
		(0.048)
CR 1018	0.053	0.010
		(0.021)
MTU 1001	0.053	0.010
		(0.026)
Swarna	0.433	-0.041
		(0.049)
Sarala	0.099	-0.047
		(0.030)
Hybrid Rice	0.052	0.004
		(0.014)
Other Modern Seeds	0.065	0.024
		(0.026)
Local Varieties	0.304	0.052
		(0.040)

The table shows separate regressions for adoption of the rice varieties in each row on the treatment and district fixed effects. All regressions use the data from the follow-up survey with farmers in August/September of 2017. The first column shows mean adoption in the control group while the second column shows the coefficient estimate and its standard error (in parentheses). The standard errors in each regression are clustered at the block level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Table B.1.8: Comparing Treatment Effects with an SMS messaging intervention

	(1) Adoption	(2) Acres	(3) Adoption	(4) Acres
SMS	-0.007	-0.012	-0.007	0.012
	(0.016)	(0.028)	(0.019)	(0.031)
Treatment			0.035	0.089*
			(0.026)	(0.046)
Treatment * SMS			-0.000	-0.049
			(0.032)	(0.055)
Dependent Variable Control Mean	0.063	0.093	0.063	0.093
R-Squared	0.024	0.023	0.028	0.027
District Fixed Effects	X	X	X	X
Observations	6653	6653	6653	6653

The table shows the treatment effects of the dealer-based extension, SMS message, and their combined effect. All regressions use the data from the follow-up survey with farmers in August/September of 2017. The dependent variables are whether the farmer was currently using Swarna-Sub1 (columns 1 and 3), and the acreage cultivated with Swarna-Sub1 (columns 2 and 4). The standard errors in each regression are clustered at the block level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Table B.1.9: Average risk level of adopters by treatment group

	(1) Above-Median Risk	(2) Days Flood
Treatment	0.259*	6.742
	(0.144)	(4.322)
Mean in Control	0.239	6.273
Number of Observations	441	441
R squared	0.068	0.046

The regressions show average exposure to flood risk between Swarna-Sub1 adopters in treatment and control blocks. The dependent variable in column 1 is the binary indicator for above-median risk (exposure to flooding for four or more days). The dependent variable in column 2 is the days of exposure across all monsoon seasons (June-October) from 2011 to 2017. Standard errors are clustered at the block level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Table B.1.10: Heterogeneous Treatment Effects on Adoption

	(1)	(2)	(3)	(4)	(5)
Treatment	0.032	0.027	0.033*	0.023	0.013
	(0.020)	(0.018)	(0.019)	(0.018)	(0.017)
Scheduled Tribe	0.023*		0.018		0.018
or Caste	(0.012)		(0.016)		(0.016)
Below Poverty	0.030***			0.013	0.019
Line Card	(0.011)			(0.012)	(0.013)
Area Cultivated	0.015***	0.013***			0.013***
(Acres)	(0.003)	(0.004)			(0.003)
Treatment * Area		0.002			0.003
Cultivated		(0.006)			(0.006)
Treatment *			0.013		0.011
Scheduled Tribe or Caste			(0.025)		(0.026)
Treatment *				0.023	0.021
Below Poverty Line Card				(0.019)	(0.021)
Dependent Variable Control Mean	0.069	0.069	0.063	0.063	0.069
R-Squared	0.046	0.042	0.029	0.031	0.047
District Fixed Effects	X	X	X	X	X
Observations	6177	6193	6642	6628	6177

The table shows heterogeneous effects of the dealer treatment by farm size, caste, and poverty status (columns 2-5). Column 1 shows the correlations between these characteristics and adoption, across both treatment and control blocks. All regressions use the data from the follow-up survey with farmers in August/September of 2017. The dependent variable in all regressions is whether the farmer was currently using Swarna-Sub1. The standard errors in each regression are clustered at the block level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Table B.1.11: Share of dealers in estimation sample that received seeds and information

	(1) All	(2) In Business
Dealer-Based Extension	0.423***	0.404***
	(0.052)	(0.057)
Dependent Variable Control Mean	0.000	0.000
R-Squared	0.329	0.324
District Fixed Effects	X	X
Observations	473	274

The table shows the "first-stage impact" of a dealer being located in a treatment block on the probability that they were provided Swarna-Sub1 seeds and information. Column 1 is for all dealers that were reached during the year 2 survey, while column 2 is only for the dealers that were still selling rice seeds. The dependent variable in both regressions is an indicator for whether the dealer received seeds and information. The standard errors are clustered at the block level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Table B.1.12: Dealer-level correlation between receiving intervention and selling Swarna-Sub1

	(1) Year 2	(2) Year 3	(3) Year 5
Dealer Received Intervention	0.194*	0.052	0.167
	(0.106)	(0.074)	(0.102)
Mean Outcome No Intervention	0.385	0.250	0.349
R-Squared	0.098	0.121	0.257
District Fixed Effects	X	X	X
Observations	133	135	113

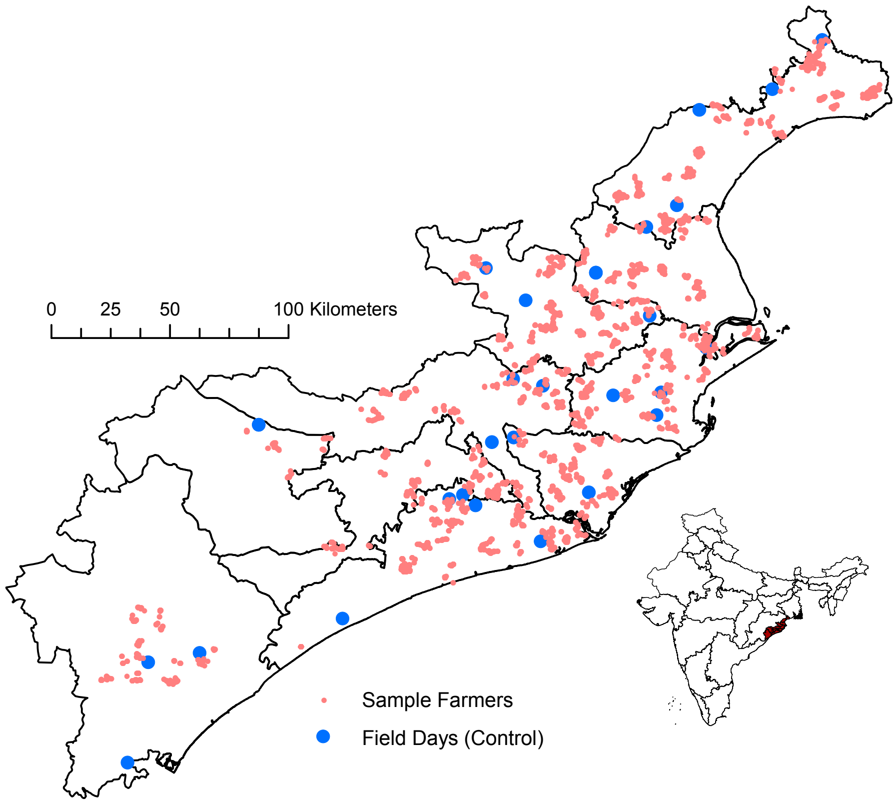
The table shows the correlation between being provided Swarna-Sub1 seeds and information (receiving the intervention in treatment blocks) and selling Swarna-Sub1 seeds during the following four years. The data in all columns are limited to treatment blocks. The dependent variable in all regressions is an indicator for the dealer selling the seeds that season. Column 1 is for year 2 (2017), while columns 2 and 3 are for years 3 (2018) and 5 (2020), respectively. The standard errors are clustered at the block level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels. See Table 2.5 in the main text for the results using the random block-level variation in treatment.

Table B.1.13: Correlation between farmer-level adoption of Swarna-Sub1 in 2017 and local seed production

	(1)	(2)
2014-2016 Seed Production	0.007***	0.010***
	(0.002)	(0.003)
Dependent Variable Mean	0.081	0.081
R-Squared	0.030	0.059
District Fixed Effects	X	X
Control Variables		X
Observations	6653	6599

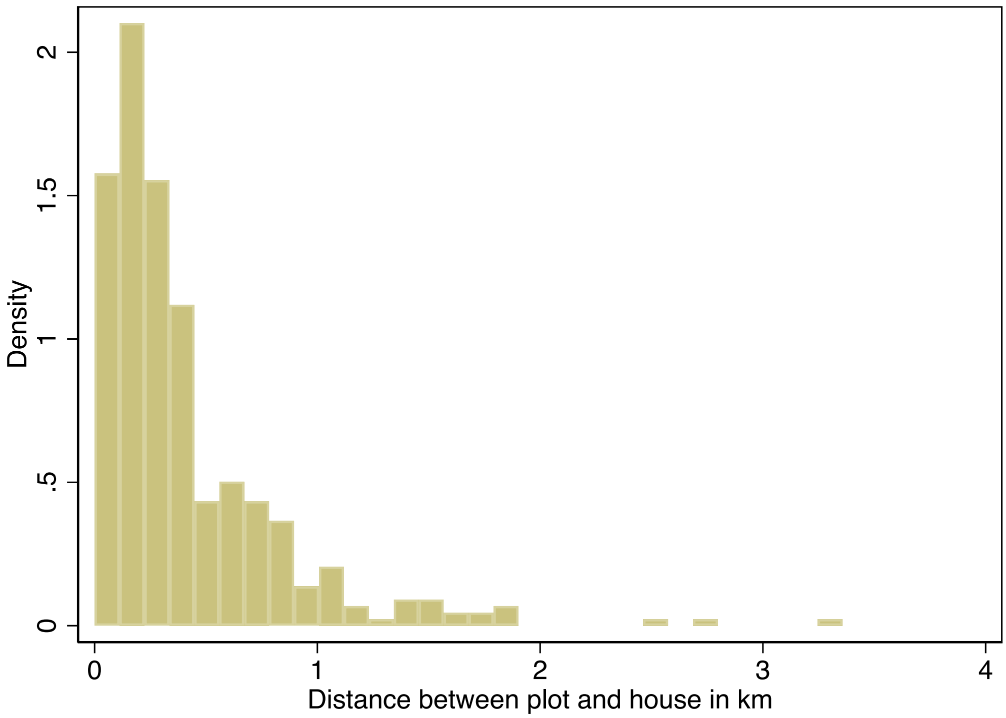
The table shows the within-district correlation between Swarna-Sub1 seed adoption by farmers and the amount of seed produced locally by growers. The estimates come from the 2017 survey with farmers where Swarna-Sub1 adoption is regressed on the average annual Swarna-Sub1 seed production in the block from 2014-2016. Seed production is measured in hundreds of quintals (1 quintal=100 kg). The dependent variable in both regressions is an indicator variable for adopting Swarna-Sub1. The control variables in column 2 are all of the covariates in Table 2.1 of the main text. The standard errors in each regression are clustered at the block level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Figure B.1.1: Location of the sample



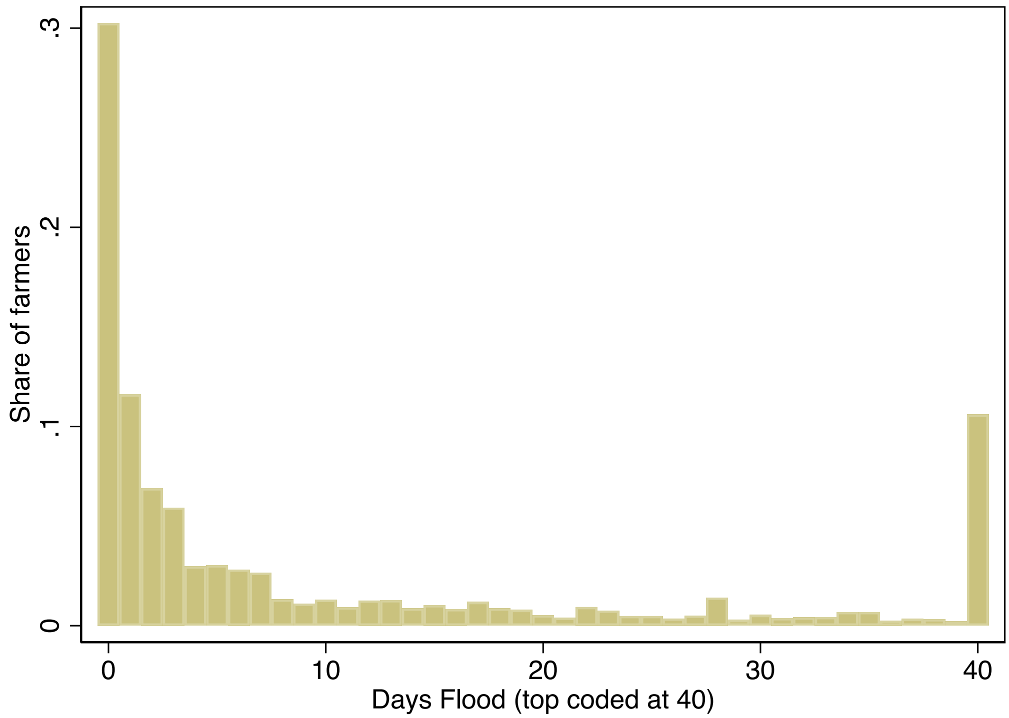
Notes: The figure shows the location for 5,536 of the 7,200 same farmers where we obtained GPS coordinates (light red dots) and the location of the farmer field days in the control blocks (blue dots). The map of India in the lower right shows the location of the sample area in the coastal belt of Odisha state. The district boundaries were obtained from the GADM database of Global Administrative Areas (Global Administrative Areas, 2018).

Figure B.1.2: Distance between plots and houses



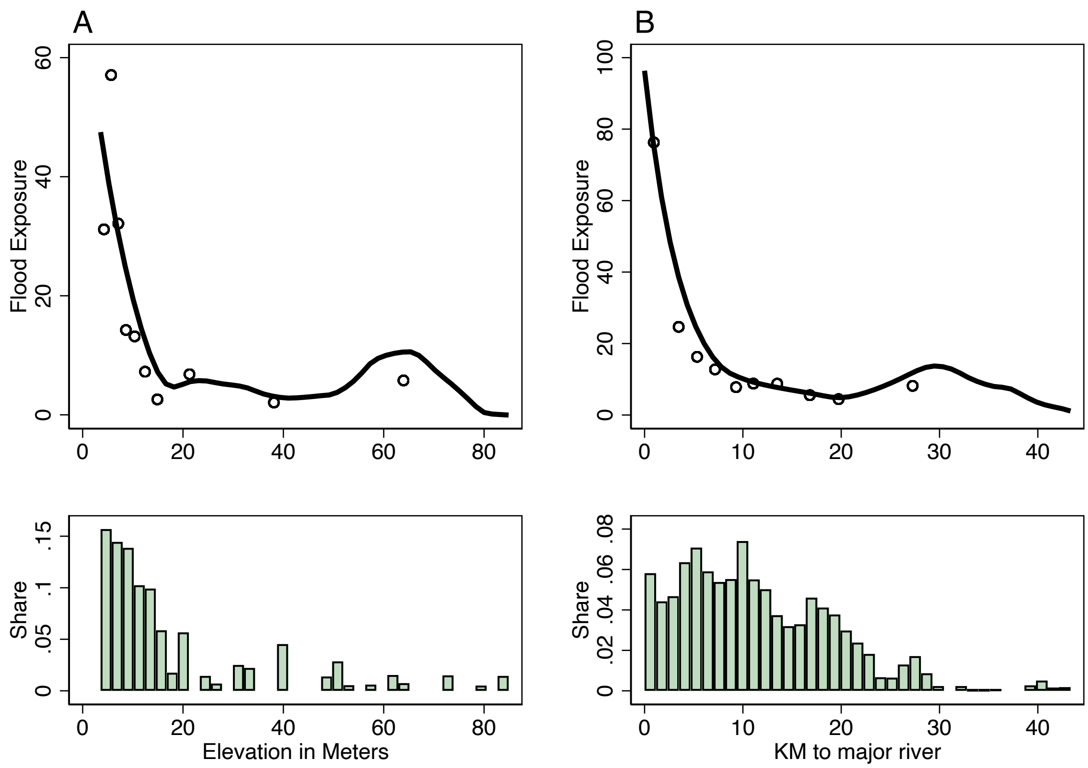
Notes: Figure shows the distribution of distances between houses and the rice plots (in km) for farmers in Emerick and Dar (2021). The district in this study is one of the 10 districts in the current paper. 92 percent of fields are within 1 km of the house.

Figure B.1.3: Distribution of measure of flood exposure



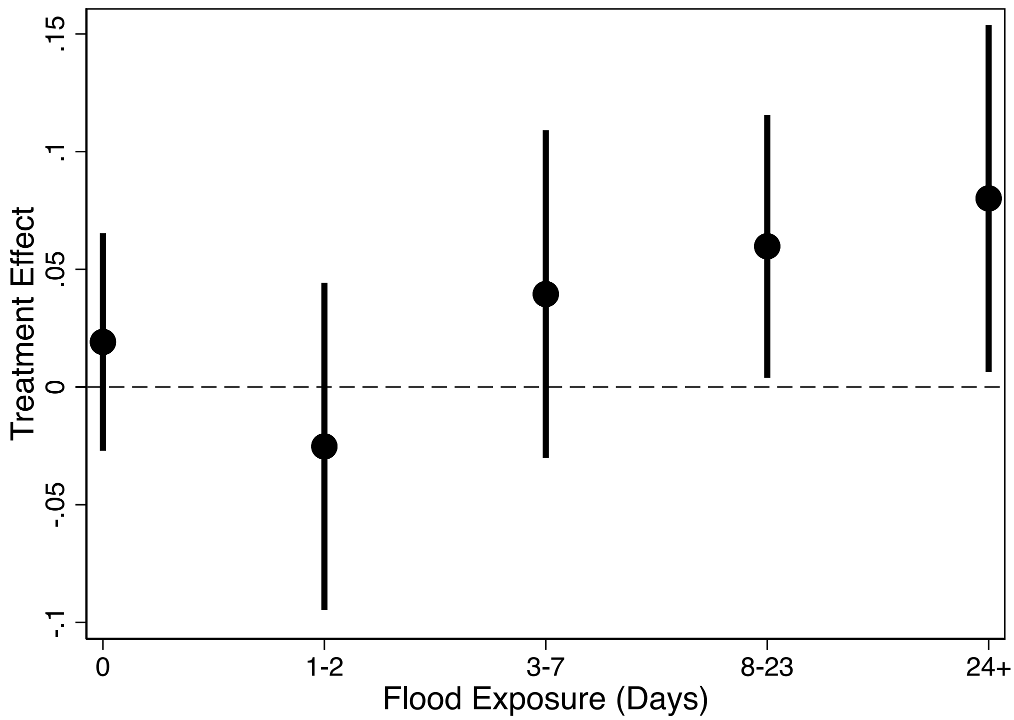
Notes: Figure shows the distribution of the days flooded from 2011 to 2017 for 5,536 households. The height of each bar displays the share of farmers with the corresponding number of days of exposure. All farmers with more than 40 days of exposure are included in the last bin at 40 days.

Figure B.1.4: Correlation between 2011-2017 flood exposure, elevation, and proximity to rivers



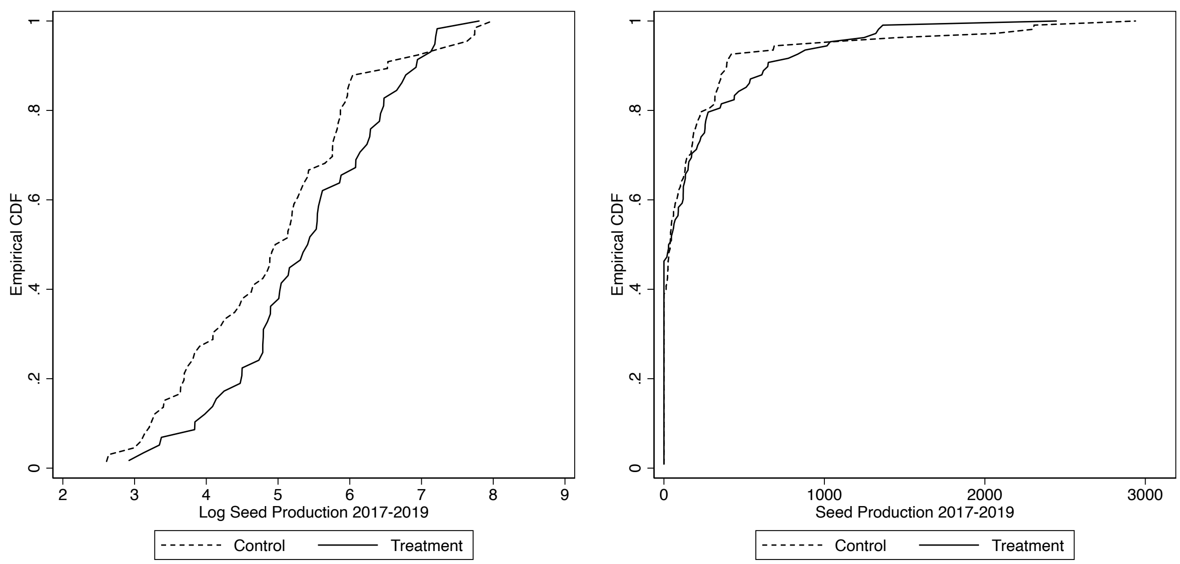
Notes: Panel A shows a non-parametric fan regression of flood exposure on elevation (heavy black line) and the average exposure levels for 10 equal-sized bins of elevation. The distribution of elevation is displayed at the bottom of the panel. Panel B shows a similar figure where flood exposure is regressed on proximity to major rivers.

Figure B.1.5: Treatment effects by flood exposure with imputing locations for households with missing GPS coordinates



Notes: The figure shows treatment effects from a single regression of adoption on separate treatment indicators for different levels of flood exposure and district fixed effects. It is identical to Figure 2.2 in the main text with the one exception being that household locations are imputed from village locations for 926 observations with missing GPS coordinates. The 5 bins of flood exposure correspond to households with no exposure from 2011-2017 and then an approximately equal division of households with at least one day of exposure. The dots are the treatment effects of dealer-based extension and the vertical lines denote 95 percent confidence intervals.

Figure B.1.6: Cumulative Distribution Functions of seed production by treatment, 2017-2019



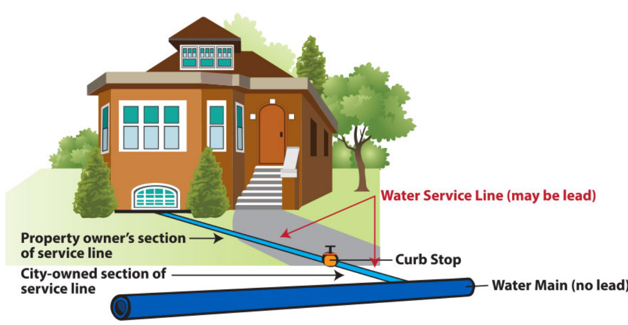
Notes: The figure shows the cumulative distribution functions of block-year level seed production for the years 2017, 2018, and 2019. The left panel uses the log of seed production while the right panel uses the level (measured in quintals where 1 quintal = 100kg).

Appendix C

Poisoned by Policy: The Impact of the Flint Water Crisis on Political Participation

C.1 Flint and the Water Crisis

Figure C.1.1: Lead Poisoning at Point of Use



Source: City of Milwaukee, milkwaukee.gov/water.

Figure C.1.2: Distribution of Lead Tests by Precinct

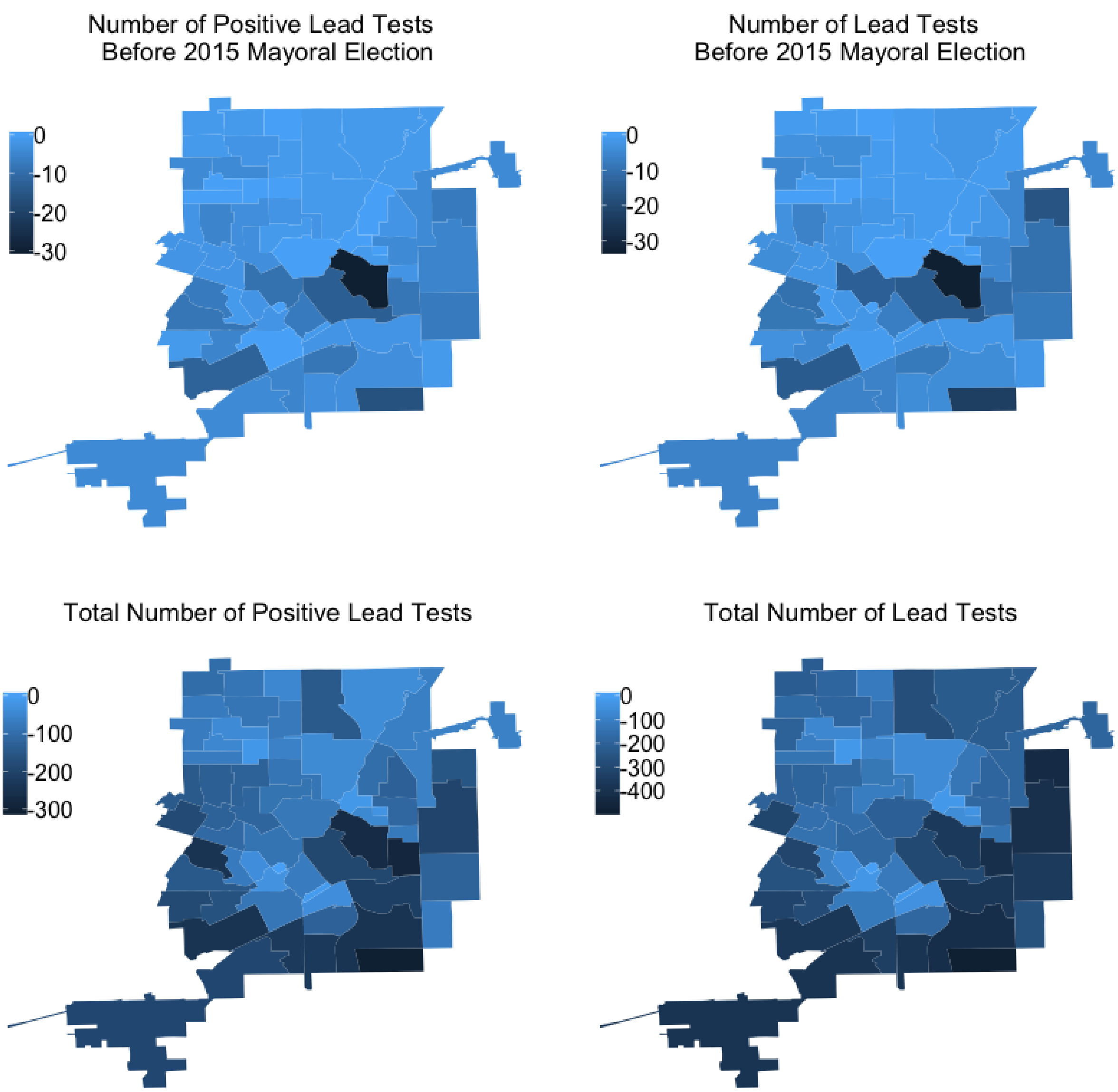


Figure C.1.3: Demographics by Precinct

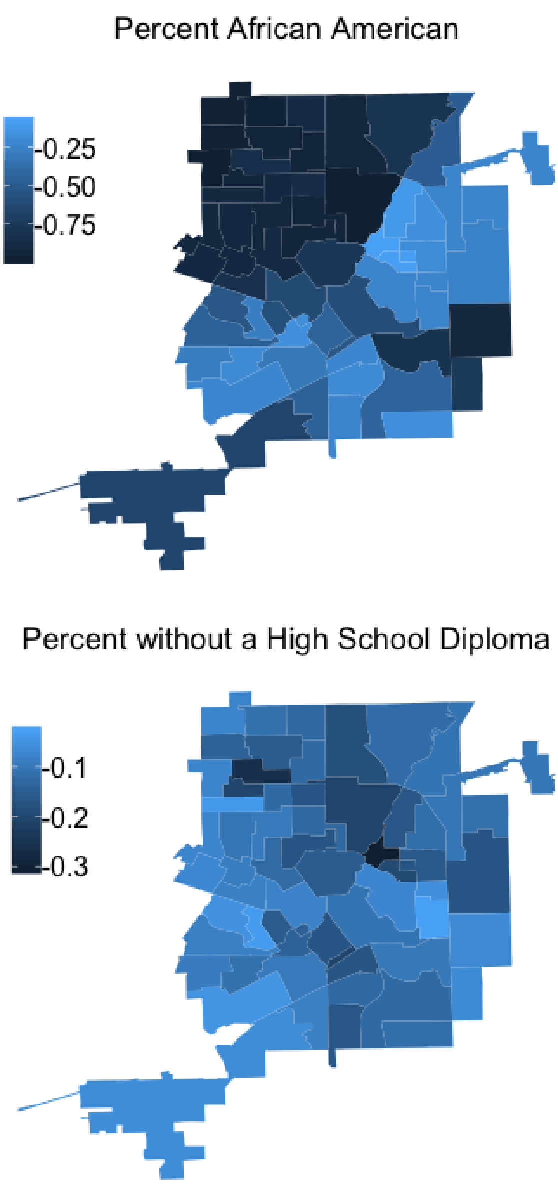
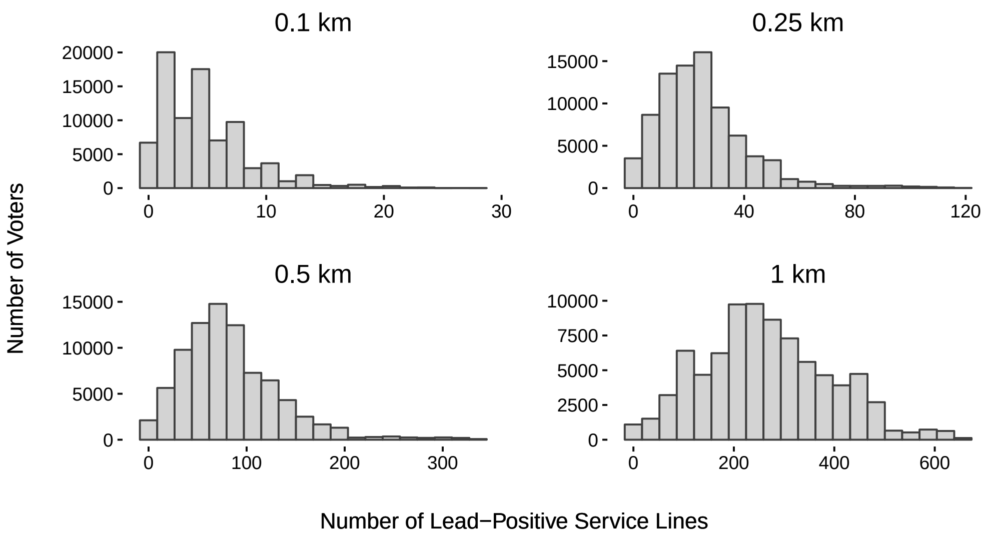


Table C.1.1: Timeline of Water Crisis and Elections

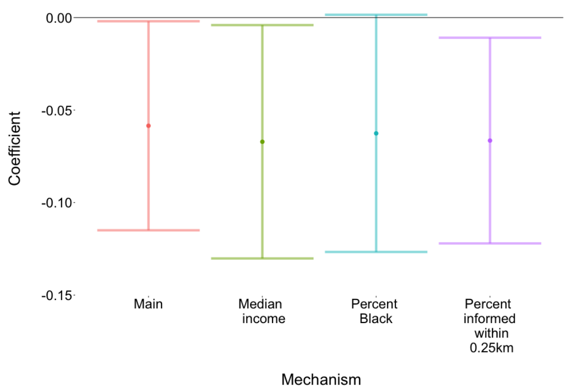
Aug 8, 2011	*Mayoral primary*
Nov 8, 2011	*Mayoral election*
Nov 29, 2011	Gov. Snyder appoints emergency manager
Feb 28, 2012	*Presidential primary*
Aug 8, 2012	*Congressional primary*
Nov 6, 2012	*Presidential election*
April 25, 2014	Emergency manager leads switch to river water
	Immediate, noticeable fall in water quality citywide
Aug 5, 2014	*Gubernatorial primary*
Oct 13, 2014	General Motors announces water is corroding engines at its plant
Nov 11, 2014	*Gubernatorial election*
Feb 18, 2015	City tests find lead in residential water
April 30, 2015	Emergency management is lifted
Aug 4, 2015	*Mayoral primary*
Sept, 2015	Residents, pediatrician, scientist report elevated lead levels in water and children's blood
Sept 25, 2015	City issues public warning
Oct 16, 2015	Declaration of emergency; activation of National Guard; emergency federal assistance
Nov 3, 2015	*Mayoral election*
Dec 2015 - Jan 2016	Michigan Attorney General brings criminal charges against Dept. of Environmental Quality and City of Flint employees
March 8, 2016	*Presidential primary*
Aug 2, 2016	*Congressional primary*
Nov 8, 2016	*Presidential election*

Figure C.1.4: Individual Variation in Neighborhood Lead Exposure

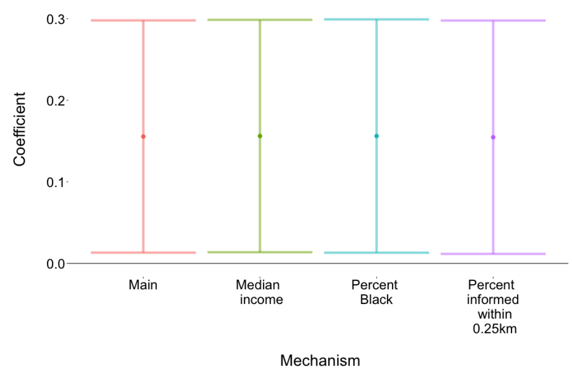


C.2 Additional analysis

Figure C.2.1: Mechanisms



(a) Positive



(b) Negative

Table C.2.1: Other possible mechanisms

	Tested Positive				Tested Negative			
	Main (1)	Median Income (2)	Percent Black (3)	Percent informed (4)	Main (5)	Median Income (6)	Percent Black (7)	Percent informed (8)
Informed	0.04***	0.06***	0.02	0.04***	0.06*	−0.02	0.09*	0.07
	(0.01)	(0.02)	(0.02)	(0.01)	(0.04)	(0.07)	(0.05)	(0.04)
Post	−0.00	−0.01*	−0.01	0.00	0.00	−0.00	−0.03***	0.01
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Informed · Post	0.03	0.04	0.04	0.01	−0.02	0.04	0.05	−0.07
	(0.02)	(0.03)	(0.04)	(0.02)	(0.05)	(0.08)	(0.07)	(0.07)
Black	0.01*	0.01*	0.01	0.01*	−0.00	−0.00	−0.01	−0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Black · Post	0.01*	0.01*	0.01*	0.01**	0.01**	0.01*	0.01	0.02**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Informed · Black	−0.02	−0.02	−0.02	−0.02	−0.11**	−0.11***	−0.11**	−0.11**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.05)	(0.04)	(0.04)	(0.04)
Informed · Post · Black	−0.06**	−0.07**	−0.06*	−0.07**	0.16**	0.16**	0.16**	0.15**
	(0.03)	(0.03)	(0.03)	(0.03)	(0.07)	(0.07)	(0.07)	(0.07)
Pct informed lead within 0.5k	0.00	0.00	0.01**		0.00	−0.00	0.01*	
	(0.00)	(0.00)	(0.00)		(0.00)	(0.00)	(0.00)	
Mechanism		0.00	0.01*	0.00*		0.00	0.02***	0.00
		(0.00)	(0.01)	(0.00)		(0.00)	(0.01)	(0.00)
Mechanism·Post		0.00*	0.02*	−0.01		0.00	0.06***	−0.01**
		(0.00)	(0.01)	(0.00)		(0.00)	(0.01)	(0.01)
Informed · Mechanism		−0.00	0.03	0.00		0.00	−0.06	−0.02
		(0.00)	(0.03)	(0.01)		(0.00)	(0.07)	(0.04)
Informed · Mechanism·Post		−0.00	−0.01	0.02*		−0.00	−0.14	0.09
		(0.00)	(0.05)	(0.01)		(0.00)	(0.13)	(0.06)
Dep. var. means	0.306	0.306	0.306	0.306	0.292	0.292	0.292	0.292
Num. obs.	41517	41224	41224	41517	32645	32310	32310	32645
R^2 (full model)	0.51	0.51	0.51	0.51	0.48	0.48	0.49	0.48
R^2 (proj model)	0.50	0.50	0.50	0.50	0.47	0.47	0.47	0.47
Adj. R^2 (full model)	0.51	0.51	0.51	0.51	0.48	0.48	0.48	0.48
Adj. R^2 (proj model)	0.49	0.49	0.49	0.49	0.46	0.46	0.46	0.46
Num. groups: block	209	208	208	209	204	203	203	204

Note: These figures plot the results from running 3.2 for a set of three potential alternative mechanisms: median income, neighborhood racial composition, and neighborhood lead test information. The variable *Mechanism* refers to the potential mechanism named in the column header. The variable $Informed$ is equal to 1 for residents who received lead test results before the 2015 mayoral election. The variable *Post* is equal to one for the 2015 mayoral election, which occurred after the water crisis began. Results for residents informed of positive lead tests are shown in columns 1-4, and results for residents informed of negative lead tests are shown in columns 5-8. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

Table C.2.2: Heterogeneity by median income by race

	Tested Positive			Tested Negative		
	Main	Black	White	Main	Black	White
	(1)	(2)	(3)	(4)	(5)	(6)
	(0.02)	(0.03)	(0.04)	(0.06)	(0.07)	(0.12)
Post	−0.01	−0.00	−0.01	0.00	0.01	0.00
	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)
Informed · post	0.03	−0.01	−0.01	0.11	0.20***	−0.24
	(0.03)	(0.05)	(0.06)	(0.08)	(0.07)	(0.26)
Median income	0.00	0.00	−0.00	0.00	−0.00	−0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Informed · Median income	−0.00	0.00	−0.00	0.00	0.00	−0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Post ·Median income	0.00**	0.00*	0.00	0.00	0.00	−0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Informed· Post ·Median income	−0.00	−0.00	0.00	−0.00	−0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Dep. var. mean	0.306	0.324	0.307	0.292	0.305	0.314
Num. obs.	41224	22961	11411	32310	17357	8764
R^2 (full model)	0.51	0.53	0.51	0.48	0.49	0.50
R^2 (proj model)	0.50	0.51	0.48	0.47	0.46	0.47
Adj. R^2 (full model)	0.51	0.52	0.50	0.48	0.48	0.49
Adj. R^2 (proj model)	0.49	0.50	0.47	0.46	0.46	0.45
Num. groups: block	208	201	185	203	193	183

Note: This table reports the results from running 3.2 for the median income mechanism. Each column shows results for the subsample named in the column title. The variable $Informed$ is equal to 1 for residents who received lead test results before the 2015 mayoral election. The variable *Post* is equal to one for the 2015 mayoral election, which occurred after the water crisis began. Results for residents informed of positive lead tests are shown in columns 1-4, and results for residents informed of negative lead tests are shown in columns 5-8. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

Table C.2.3: Heterogeneity by Race and Gender: Tested Positive Sample

	Pooled		Black		White	
	Female	Male	Female	Male	Female	Male
Informed	0.03**	0.03**	0.04*	−0.01	−0.00	0.02
	(0.01)	(0.01)	(0.03)	(0.02)	(0.03)	(0.03)
Post	0.01**	−0.00	0.02***	0.00	−0.00	−0.00
	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
Informed · Post	0.00	0.01	−0.06*	−0.01	0.05	−0.01
	(0.02)	(0.02)	(0.03)	(0.04)	(0.04)	(0.04)
Num. obs.	21902	19615	12241	10720	5773	5641
R^2 (full model)	0.52	0.51	0.53	0.54	0.52	0.51
R^2 (proj model)	0.49	0.49	0.49	0.51	0.48	0.46
Adj. R^2 (full model)	0.51	0.51	0.52	0.53	0.50	0.49
Adj. R^2 (proj model)	0.49	0.49	0.49	0.50	0.46	0.44
Num. groups: block	208	202	198	194	182	180
Dep var mean	0.327	0.281	0.347	0.297	0.332	0.280

Note: This table explores heterogeneity by race and gender by running specification 3.1 on the subsamples named in the column descriptions. The variable $Informed$ is equal to 1 for residents who received lead test results before the 2015 mayoral election. The variable $Post$ is equal to one for the 2015 mayoral election, which occurred after the water crisis began. Results for residents informed of positive lead tests are shown in columns 1-3, and results for residents informed of negative lead tests are shown in columns 4-6. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

Table C.2.4: Heterogeneity by Race and Gender: Tested Negative Sample

	Pooled		Black		White	
	Female	Male	Female	Male	Female	Male
Informed	0.04	-0.03	0.03	-0.09^{**}	0.15^{***}	0.05
	(0.04)	(0.03)	(0.05)	(0.04)	(0.05)	(0.05)
Post	0.02^{***}	0.00	0.02^{***}	0.01	0.01	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Informed · Post	0.04	0.08	0.10^{*}	0.17^{**}	-0.03	-0.11
	(0.05)	(0.05)	(0.06)	(0.07)	(0.09)	(0.09)
Num. obs.	18032	14613	9555	7802	4546	4221
R^2 (full model)	0.48	0.49	0.49	0.49	0.52	0.50
R^2 (proj model)	0.46	0.46	0.46	0.46	0.47	0.45
Adj. R^2 (full model)	0.48	0.48	0.48	0.48	0.49	0.47
Adj. R^2 (proj model)	0.46	0.46	0.45	0.44	0.45	0.42
Num. groups: block	199	200	187	190	177	169
Dep var mean	0.307	0.274	0.328	0.276	0.334	0.293

Note: This table explores heterogeneity by race and gender by running specification 3.1 on the subsamples named in the column descriptions. The variable $Informed$ is equal to 1 for residents who received lead test results before the 2015 mayoral election. The variable *Post* is equal to one for the 2015 mayoral election, which occurred after the water crisis began. Results for residents informed of positive lead tests are shown in columns 1-3, and results for residents informed of negative lead tests are shown in columns 4-6. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

Table C.2.5: Robustness: Limiting the Sample to Voters Who Received Lead Test Results Within 2 Months of the 2015 Election

	Tested Positive			Tested Negative		
	All (1)	Black (2)	White (3)	All (4)	Black (5)	White (6)
Informed	0.04^{***}	0.04	0.00	-0.00	-0.05	0.09^{*}
	(0.01)	(0.02)	(0.02)	(0.02)	(0.03)	(0.04)
Post	0.02^{**}	0.02^{**}	0.01	0.01	0.03^{***}	-0.02^{*}
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Informed · Post	-0.01	-0.04	0.01	0.06	0.13^{**}	-0.04
	(0.01)	(0.02)	(0.03)	(0.04)	(0.05)	(0.06)
Dep. var. means	0.310	0.327	0.327	0.332	0.33	0.354
Num. obs.	18811	10337	4755	11360	6259	3348
R^2 (full model)	0.54	0.56	0.55	0.53	0.54	0.55
R^2 (proj model)	0.50	0.51	0.48	0.50	0.50	0.48
Adj. R^2 (full model)	0.54	0.55	0.53	0.52	0.53	0.53
Adj. R^2 (proj model)	0.50	0.50	0.46	0.49	0.48	0.45
Num. groups: block	188	181	161	190	169	155

Note: This table reports results from limiting the sample to voters who received test results within 2 months of the 2015 mayoral election before running specification 3.1 on the subsamples named in the column descriptions. The variable $Informed$ is equal to 1 for residents who received lead test results before the 2015 mayoral election. The variable $Post$ is equal to one for the 2015 mayoral election, which occurred after the water crisis began. Results for residents informed of positive lead tests are shown in columns 1-3, and results for residents informed of negative lead tests are shown in columns 4-6. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

Table C.2.6: Robustness: Expanding the Sample to All Voters - Including Never-Tested

	Tested Positive			Tested Negative		
	All (1)	Black (2)	White (3)	All (4)	Black (5)	White (6)
Informed	0.03***	0.03**	0.02	−0.00	−0.01	0.03*
	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)
Post	−0.00***	0.00	−0.01***	−0.00***	0.00	−0.01***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Informed · Post	0.01	−0.02	0.03	0.04***	0.12***	−0.02
	(0.01)	(0.02)	(0.03)	(0.01)	(0.03)	(0.02)
Dep. var. means	0.306	0.324	0.307	0.194	0.202	0.197
Num. obs.	195771	104318	57622	195771	104318	57622
R^2 (full model)	0.47	0.48	0.47	0.47	0.48	0.47
R^2 (proj model)	0.47	0.47	0.46	0.47	0.47	0.46
Adj. R^2 (full model)	0.47	0.48	0.47	0.47	0.48	0.47
Adj. R^2 (proj model)	0.47	0.47	0.46	0.47	0.47	0.46
Num. groups: block	222	217	213	222	217	213

Note: This table reports results from expanding the sample to include all voters including those who never opted into lead testing 3.1 on the subsamples named in the column descriptions. The variable $Informed$ is equal to 1 for residents who received lead test results before the 2015 mayoral election. The variable *Post* is equal to one for the 2015 mayoral election, which occurred after the water crisis began. Results for residents informed of positive lead tests are shown in columns 1-3, and results for residents informed of negative lead tests are shown in columns 4-6. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.