NBER WORKING PAPER SERIES

SEARCH COSTS, INTERMEDIATION, AND TRADE:
EXPERIMENTAL EVIDENCE FROM UGANDAN AGRICULTURAL MARKETS

Lauren F. Bergquist
Craig McIntosh
Meredith Startz

Working Paper 33221
http://www.nber.org/papers/w33221

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2024

We gratefully acknowledge funding from USAID/BASIS, USAID/DIL, the Bill & Melinda Gates Foundation and the UK Foreign, Commonwealth & Development Office, awarded through the Agricultural Technology Adoption Initiative (ATAI), the International Growth Center, the Policy Design and Evaluation Lab, and an anonymous donor. We thank IPA, Stephanie Annyas and Laza Razafimbelo for excellent project management and James Allen, Russell Morton, and Jared Stolove for excellent research assistance. We thank Paul Nyende from AgriNet and the Kudu team, including Richard Ssekibuule, Nicole Immorlica, Kevin Leyton-Brown, Brendan Lucier, Neil Newman, and John Quinn, for their partnership. This trial was registered as AEARCTR-0000749, and is covered by UCSD IRB #140744. All errors are our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Search Costs, Intermediation, and Trade: Experimental Evidence from Ugandan Agricultural Markets
Lauren F. Bergquist, Craig McIntosh, and Meredith Startz
NBER Working Paper No. 33221
December 2024
JEL No. F10, O10

ABSTRACT

We study the large-scale experimental rollout of a platform that reduced search and matching frictions in Ugandan agricultural markets by connecting buyers and sellers. Market integration improved substantially: trade increased and price gaps fell. Interpreting the experiment through a trade model, we estimate treatment effects accounting for equilibrium changes that impact control markets. The intervention reduced fixed trade costs by 21% and increased trade flows between treated markets by 6% and across all markets by 1%. Scale economies shaped engagement: few farmers used the platform, but equilibrium price convergence from improved arbitrage by larger traders passed through to farm revenue.

Lauren F. Bergquist
Department of Economics
and Jackson School of Global Affairs
Yale University
87 Trumbull Street
New Haven, CT 06511
and NBER
lauren.bergquist@yale.edu

Meredith Startz
Department of Economics
Dartmouth College
Hanover, NH 03755
and NBER
meredith.l.startz@dartmouth.edu

Craig McIntosh
School of Global Policy and Strategy
University of California, San Diego
9500 Gilman Drive
La Jolla, CA 92093-0519
ctmcintosh@ucsd.edu

A randomized controlled trials registry entry is available at
https://www.socialscienceregistry.org/trials/749

1 Introduction

Trade barriers in sub-Saharan Africa are among the highest in the world (Teravaninthorn and Raballand, 2009; Porteous, 2019). This impedes integration of agricultural surplus and deficit areas, putting a wedge between consumer and producer prices, with implications for food security, farmer income, and aggregate welfare (Barrett, 2008). While this is due in part to high transportation costs, spatial price variation is much larger than can be explained by transportation alone (Atkin and Donaldson, 2015). Evidence from the spread of mobile phones suggests that information frictions may be an important contributor – as phones reduced the cost of accessing information about distant places, market integration improved and price dispersion fell (Aker, 2010; Jensen, 2010; Allen, 2014). However, integration remains highly incomplete even in places with high mobile phone penetration (Moser et al., 2009). Furthermore, we know very little about *which* information matters – much of the literature has focused on prices, but interventions that disseminate price information have had limited impact (Fafchamps and Minten, 2012).

We investigate the role of a different information friction: search for transaction partners. Digital platforms aimed at facilitating matches are an increasingly popular tool in markets in low and middle income countries (LMICs). We conduct a large-scale randomized control trial of a mobile marketplace in Uganda, which links potential buyers and sellers of agricultural goods through a simple SMS-based platform. The design of the platform and experiment allow us to isolate the causal effect of reducing buyer-seller search costs, distinct from other types of information or trade barriers. The at-scale randomization enables us to measure impacts on aggregate outcomes like trade flows and prices, accounting for general equilibrium forces that affect both treated and control markets.

The mobile marketplace we study, called Kudu, acts as a clearinghouse. Those buying and selling agricultural commodities submit bids and asks, and the platform directly connects potential matches, who can then choose to transact or not off-platform. Kudu is free and promoted to both farmers and traders, with in-village support services to ensure ease of access. As part of the study, introduction was randomized across 110 Ugandan subcounties, covering a population of almost three million. We track outcomes over six harvest seasons, collecting high frequency data on crop prices in 236 markets, as well as multiple rounds of surveys with traders and farmers.

We begin with a simple treatment-control comparison, and find that there is more trade and price convergence among markets with access to Kudu. Trade increases on both extensive and intensive margins: treated market pairs are more likely to engage in any trade with each other, have a larger number of traders who trade between them, and see higher volumes traded. Prices are higher in treated markets relative to control markets in crop surplus areas, but lower in deficit areas. As a result, price gaps between treated markets are smaller, even though average price levels do not change. A separate treatment arm that provided price information rather than connections to new transaction partners had no impact, confirming that the channel for these impacts is likely buyer-seller search.

Randomization circumvents many of the inference problems that arise when studying potentially endogenous variation in trade costs, and randomizing at the subcounty level – rather than the individual, or even village level – enables us to see measurable changes in equilibrium outcomes like prices. However, a remaining challenge is that although control markets do not benefit directly from the reduction in search costs, they are affected by changes in these equilibrium outcomes through trade links. This means that simple treatment-control comparisons do not yield valid estimates of treatment effects, a challenge inherent to a trade setting. To correctly measure the impact of access to the platform, we therefore interpret the experimental variation through the lens of a trade model, estimating treatment effects in the presence of general equilibrium forces that create spillovers between treatment and control units.

We build a model of domestic agricultural trade with fixed and variable trade costs between markets, and traders who are heterogeneous in both size and idiosyncratic match quality with specific transaction partners. We imbed a search friction by requiring sellers to pay a fixed cost to find buyers in a particular market before selling there. This represents the costly search activities we observe in the survey data, such as time spent calling contacts or traveling to another market to look for buyers. Kudu potentially lowers this search cost by making it easier to connect with new partners. We estimate the direct effect of the platform on trade costs using experimental variation in bilateral treatment status between markets, holding market-level factors (such as prices) constant. We find that Kudu drove a 21% reduction in fixed trade costs between treated markets.

To quantify the total impact of Kudu, we estimate additional parameters of the model – such as the elasticity of substitution across sellers and the distribution of traders' operating costs – that govern the size of indirect effects operating through general equilibrium forces. To do this, we make use of random variation in both own treatment status and in exposure to treatment in other markets. These estimates allow us to simulate a scenario in which no market has access to Kudu; in the presence of spillovers, this is the counterfactual needed to estimate treatment effects. With this in hand, we can identify the impact of Kudu on treated routes *and* on control routes. We find that Kudu increased trade flows on treated routes by 6% and in the aggregate across all markets (including negative spillovers on control routes) by 1%. While still substantial, this is only about a third of the treatment effect on aggregate trade flows implied by a simple treatment-control comparison that does not account for equilibrium effects.

Kudu's impact demonstrates that there are still large information frictions impeding agricultural trade in LMICs, and that these can be meaningfully alleviated by a low-cost clearinghouse technology that makes it easier to match with buyers. Many practitioners hope that platforms like Kudu will increase farmers' revenue, perhaps by enabling them to sell directly to buyers in higher priced markets. While we find that many farmers do benefit, the mechanism is more nuanced.

First, farmers very rarely engage in cross-market trade, either with or without Kudu. Instead, almost all activity on the platform and the resulting improvement in arbitrage is driven by professional traders. Our structural estimation sheds light on why: the size of fixed costs is such that all but the very largest farmers operate on too small a scale to make cross-market trade profitable. The intervention lowers these fixed costs, allowing smaller traders to enter more trade routes on the margin – which is precisely what we observe in the survey data – but the average level of scale economies remains too large for most farmers.

Despite their lack of direct engagement, farmers are still affected by changes in equilibrium prices. However, the reduction in search costs leads to price convergence, rather than a change in average prices. This is good for farmers and bad for consumers in surplus markets, where initially low prices rise, and the opposite in deficit markets, where initially high prices fall. There are net gains from trade, such that overall welfare in the study area increases. We see empirically that farmers in relative surplus areas enjoy significant increases in crop revenue

Finally, we use our structural model to explore how the total impact of the platform changes as access is counterfactually rolled out to more markets, holding the direct technological impact on trade costs across treated markets constant. At low levels of saturation, aggregate gains are small, and trade diversion effects are high, with large changes in the volume of trade on treated routes for a given change in fixed trade costs. When the platform is rolled out to all markets, treatment effects on average trade volumes are relatively small, but the change in the allocation of trade yields welfare gains that are roughly three times as large.

This paper contributes to a literature on interventions to increase smallholder farmers' revenues in LMICs. Much of the existing work in this space has focused on supply side interventions to increase farmer productivity, such as agricultural extension (Beaman et al., 2023), credit (Karlan et al., 2014), and insurance (Carter et al., 2017). We focus on market access and the role of search costs. The few other studies in this space have mostly focused on price search,[1], with several randomized evaluations of programs to disseminate price information to farmers via mobile phones finding mixed, and overall limited, impacts (Camacho and Conover, 2011; Fafchamps and Minten, 2012; Nakasone, 2014; Mitra et al., 2018; Hildebrant et al., 2020).[2] In contrast, we study buyer-seller search, which has received much less attention, especially in the context of agricultural markets.

Most of the previously mentioned experiments with smallholder farmers are randomized at the individual or village level, while prices are often determined in equilibrium across a wider area, complicating inference.[3]. We randomize across much larger units and interpret the experimental

[1] A partial exception is Goyal (2010) which finds that internet kiosks facilitating direct access to a single large private buyer of soy in India increases farmer prices. In contrast, the Kudu platform facilitates matches across a large number of buyers and sellers countrywide.

[2] Wiseman (2023) does find that price information may matter for small-scale traders engaged in cross-border trade.

[3] For example, Svensson and Yanagizawa (2009) find evidence that broadcasting prices via radio leads to higher farmgate prices in Uganda; however, a follow-up paper suggests that once accounting for general equilibrium effects, average farmer revenues impacts are minimal (Svensson and Yanagizawa-Drott, 2012) Hildebrant et al. (2020) find evidence of spillovers from a price alert system in Ghana.

variation through the lens of a model that accounts for indirect effects on control markets. In doing so, we contribute to a growing development literature on accounting for GE forces within experiments (Muralidharan and Niehaus, 2017; Fink et al., 2020; Egger et al., 2022; Bergquist et al., 2023; Cai and Szeidl, 2024). We also connect to a non-experimental literature on the influence of trade costs on agricultural market integration and outcomes for small farmers in spatial general equilibrium ((Gollin and Rogerson, 2014; Shamdasani, 2021; Allen and Atkin, 2022; Kebede, 2024)). By randomizing the introduction of Kudu, we join a small set of papers that randomize a component of route-specific trade costs (e.g., Brooks and Donovan (2020)) or combine experimental variation with spatial models (e.g., Kreindler (2024)).

We also contribute to a broader trade literature on search frictions in LMICs. Following the early work on the spread of mobile phone access (Aker, 2010; Jensen, 2010; Allen, 2014), a recent literature has explored non-price search frictions. Vitali (2024) looks at how consumer search affects the locations of garment firms in Uganda, Cai et al. (2024) study search for input suppliers in Chinese writing brush supply chains, and Startz (2024) measures product search frictions facing Nigerian importers of differentiated goods. Dillon et al. (2024) and Rudder and Dillon (2024) study the experimental dissemination of "yellow pages" telephone directories, and find that they increase SMEs' sales but do not necessarily lead to new links with other firms. Bai et al. (2021) document the importance of congestion and incomplete information about quality in search by buyers on an e-commerce platform.

Finally, we build on a nascent literature that investigates how scale economies shape market participation and outcomes in LMICs. Bassi et al. (2022) and Kaboski et al. (2024) respectively study Ugandan SMEs and entrepreneurs, highlighting the importance of fixed cost investments, which when divisible can shape inter-firm relationships and when indivisible can affect firm risk-taking. Foster and Rosenzweig (2022) explore the relationship between farm size and machine capacity in determining productivity in LMIC agriculture. Grant and Startz (2024) document the role that fixed costs of trade play in determining the length of distribution chains in Nigeria. We similarly find that economies of scale affect which agents participate directly in cross-market trading opportunities. However, we also show that general equilibrium effects can still generate gains for smaller market participants.

The rest of the paper is structured as follows: Section 2 discusses the setting and study design. Section 3 presents reduced-form effects on market-level outcomes. Sections 4, 5, and 6 present the structure, estimation, and results of the model respectively. Section 7 discusses implications for engagement by traders and farmers, and for full scale up outside a research context. Section 8 concludes.

2 Setting and Study Design

2.1 Sample and Data

The study took place in 110 subcounties in Uganda that were selected by our implementing partners as promising for platform roll-out.[4] Maize is the most commonly grown and consumed crop in Uganda, as well as the one most traded on the platform, and so we focus our analysis and discussion on it throughout the paper. Our study areas are spread across regions of the country and include a mix of major surplus and minor deficit areas for maize production (Figure A.1).

The study ran for three years, starting in 2015 and concluding in 2018, and spans six major agricultural seasons. Figure A.2 presents a timeline for the project. During this period, we gathered three core types of data, illustrated in Figure A.3: high-frequency market data, trader survey data, and farmer survey data. All impact analysis is based on these three data sources, although we also analyze administrative data from the platform to understand take-up and usage.

We began with a census of all markets within our study subcounties that were permanent (i.e., not meeting only on specific days of the week) and featured buying and selling of maize (as opposed to wet markets where only fruits and vegetables are sold). The 236 markets identified in this census were further classified as “hubs” (major local commercial centers) or “spokes” (more remote local markets). We gathered information biweekly (every two weeks) by calling an informant in each market, typically a trader whose store was based there. The resulting high-frequency market data captures the buying and selling prices, availability, and average quality of four major food crops (maize, nambale beans, matooke bananas, and tomatoes). Reporting was consistent, with 88% of the attempted (market by survey round) observations successfully recorded.

Our second dataset comes from surveys with traders. We first conducted a census of traders based in study markets who bought and sold at least one study crop. For markets with fewer than 10 such traders, we surveyed all traders; for markets with more than this, we randomly sampled 10 traders. Traders were administered a baseline survey prior to treatment in 2015, a midline survey in 2016 after one year of treatment, and an endline in 2018 after three years of treatment. For the trader midline, we were able to survey 1,358 of the 1,457 baseline traders (93.2%). The trader endline originally located 1,248 traders (85.7%), after which we randomly sampled 20% of attritors (41 individuals) for intensive tracking. We successfully located 37 of these (92.7%), bringing the weighted tracking rate for the trader endline to 98.6%. The analysis is weighted to make it representative of all traders in study markets.[5]

Finally, we surveyed a sample of agricultural households. We first listed all villages located in study subcounties. For each market in our study, we selected the village containing the market (which is typically urban) and randomly sampled one of the remaining villages within the same parish (which tend to be more rural). For these two villages, we then listed all the households based

[4] In the Ugandan administrative hierarchy, villages are organized into parishes, and parishes into subcounties.

[5] We also conducted a census of traders present in the markets at endline to study entry and exit (see Table 5.

on administrative records held by the village chairperson, and randomly sampled households from these lists. We sampled 8-9 farming households in each market village and another 4 in each rural village. We imposed two eligibility criteria: the household had to (i) be engaged in agriculture, and (ii) have sold some quantity of any of the four crops included in the study in the previous year (maize, beans, bananas, and tomatoes). In practice, these criteria excluded few households, with over 90% qualifying for study inclusion. Study households completed a baseline survey in 2015 and an endline survey in 2018, covering agricultural production, marketing, and sales. The household endline originally located 2,744 of the 2,971 baseline respondents (92.4%), and we then randomly sampled 17% or 39 households for intensive tracking. 31 of these households were successfully intensively tracked (79.5%), giving us a weighted household tracking rate of 98.7%. Farmer analysis is weighted to make it representative of all farming households in the sampled study villages.

2.2 Production, Marketing, and Prices

As is common in the region, the farmers in our study area are smallholders and do very limited marketing of their crops. They are small net producers on average, though our sample contains net consumers as well. Farmers grow an average of 835kg/year of maize in two harvest seasons. They sell 58% of this on average, typically making two sales per year (one per season). Farmers sell very close to home, with 66% selling exclusively at the farm-gate. The remainder sell in nearby markets, averaging just 2km away from their farm-gate. Farmers typically sell to a single trader a year, with whom they have transacted in the past.

In contrast, traders operate on a much larger scale and do a great deal of active marketing. The median study trader buys and sells 25 tons of maize per year, more than 50 times the size of the median farmer. About two-thirds of traders (62%) use 5-10 ton trucks to move their inventory; the remainder use motorcycles or bicycles. Traders typically operate from a home base in a local market, making many small purchases within their subcounty, aggregating, and selling downstream in fewer, larger transactions. The median market in our study area is home to six traders. Churn is low, with the median market seeing one new and one exiting trader over the three year study period. At baseline, traders bought maize at an average of 12.7 cents/kg and sold at 16.4 cents/kg, a buy-sell margin of 29%. From baseline monthly revenues of $2,243, traders report an average monthly profit of $297.[6] By comparison, average total monthly household expenditure in our farmer sample is $65.

Mobile phones are an important technology used by traders, who report calling a seller, buyer, or other contact ahead of time to get information in 54% of their transactions. More traditional search-by-visiting is also common, with the remainder of transactions occurring when a trader goes to a market without calling ahead for any prior information. At baseline, about 50% of traders also

[6] Consistent with these figures, Bergquist and Dinerstein (2020) estimate that the median trader in their sample in Kenya retains 12% of total revenues in profits.

report using radio broadcasts as a market discovery tool, and 10% some kind of SMS service. These services typically only offer simple price alerts; none to our knowledge provides direct connections to buyers or sellers. Use of these services is lower among farmers, with only 7% of farmers reporting listening to radio broadcasts for market information and 2% receiving information via SMS.

Market prices for maize show strong variation both over space and over time in our study area (see Table B.1). Figure 1 presents the absolute value of the gap in prices between each pair of markets in our sample (black solid line), based on our market surveys. The average price gap for maize is about 110 Ugandan shillings (UGX) per kg, or roughly 15% of average price. Unsurprisingly, price gaps rise with the distance between the markets. A major driver of these price gaps is transportation costs, which are among the highest in the world (Teravaninthorn and Raballand, 2009) in this region. Transportation costs alone, however, cannot explain the full gap in prices. Figure 1 also plots transport cost as a function of distance (dotted line), as reported in our trader surveys.[7] We observe price gaps that are roughly 50% higher than transport costs, suggesting pervasive violations of the law of one price.

2.3 Kudu: a Market-Matching Platform

The introduction and rapid spread of mobile phones across sub-Saharan Africa offered the promise of dramatically reducing search costs. Indeed, the rollout of cell-phone towers in the early 2000s substantially reduced price dispersion in grain markets (Aker, 2010; Aker and Mbiti, 2010). Building off this success, recent efforts have attempted to move beyond the passive reduction in search costs facilitated by easier bilateral communication via mobile phones, and into more active facilitation of search on mobile platforms designed specifically for agriculture. The first generation of these initiatives focused on the dissemination of price information to farmers via mobile phone (Camacho and Conover, 2011; Fafchamps and Minten, 2012; Nakasone, 2014; Mitra et al., 2018; Hildebrant et al., 2020).

In contrast to these price information interventions, Kudu facilitates matches between individual market participants. The system, which was developed at Kampala's Makerere University in partnership with researchers at Microsoft Research, is designed to make it easier for buyers and sellers of agricultural commodities to connect. Users can post asks (sale offers) and bids (purchase offers) either using a smartphone or by registering their location and then using a basic feature phone to send messages to the platform via free-form text message or a USSD drop-down menu.[8] A call

[7]Traders reported round trip transport costs along each of their five most commonly travelled routes, and the vehicle size typically used. From this data, we construct an estimate of per kg transport costs, which we then estimate as a non-parametric function of the km traveled.

[8]Kudu instructs users to post their reservation prices. The (non-binding) price recommended by the platform in the event of a match is the seller's price, so this is incentive compatible for the buyer. Perhaps because of this, seller ask prices were often higher than buyer bid and market prices. Qualitative interviews suggest that sellers often post strategic offers, much in the way they would in more traditional, in-person negotiations. Buyers' bid prices, on the other hand, track market prices very well (see Figure A.6 and Figure A.7). In practice, Kudu accommodates this sort of negotiating behavior by allowing for a small negative overlap of prices, allowing those that were close but not

center also collects asks and bids by phone. Based on the price, quantity, and location of the buyer and seller, the system then matches particular buyers and sellers as proposed trading partners.[9] Users are contacted by SMS to inform them that the match has occurred and provide them with the contact information of the other party. Buyers and sellers can then directly connect and arrange for the sale, although frequently a Kudu employee will help facilitate this communication, reaching out by phone to both parties to gauge their interest in the deal and coordinate next steps. If interested, buyers and sellers meet in person to exchange the goods and make payment. Kudu has no mechanism for processing payments, nor for enforcing that the terms of the original bid or ask are honored if a transaction takes place. We therefore interpret the function of Kudu as purely making a connection between promising potential transaction partners.

2.4 Randomization

The introduction of Kudu was randomly assigned at the subcounty level. At the time of our study, the typical subcounty in Uganda had about 30,000 residents. This is therefore an "at-scale" randomization, conducted with the goal of being able to observe treatment-control differences in outcomes such as prices and trade flows that are determined in equilibrium at the market level. We blocked the randomization on whether the subcounty contained a hub market (17%) or not (83%), and then stratified by a subcounty-level price index (mean of the z-scores of the prices of each of the four focus crops at the markets in each subcounty).

The goal of the intervention was to make Kudu accessible to all farmers and traders operating within treatment subcounties. This included not only the individuals in our trader and farming household samples, but also non-sampled individuals in sample villages, as well as those in non-sample villages. This was accomplished through intensive promotion of the platform in treatment subcounties, both on the ground and via regular text messages. While nothing prevented traders or farmers outside of treated subcounties from registering to use Kudu, in practice we find that this rarely happened. Staff were sent to treatment markets to serve as the on-the-ground agents promoting the platform to local farmers and traders.[10] In addition, we promoted Kudu via text

quite overlapping to match. Kudu also does not enforce the price recommendation, and in practice price could be renegotiated flexibly by the buyer and seller.

[9] There were two processes by which buyers and sellers could be matched during the study period. The first was an algorithm that cleared the market each day, attempting to maximize what it calculated to be the global "gains from trade", using a penalty function decreasing in the price difference between the bid and ask and increasing in distance. The second was a hand-matching process conducted by employees who could view a dashboard of the business on the platform and attempt to match trades manually. During the study the large majority of successful matches were made manually. For more technical details on the Kudu platform, see Newman et al. (2018).

[10] We initially partnered with a private sector agribusiness firm to employ and train 210 Commission Agents (CAs) to promote Kudu. However, CAs were not reliable promoters, perhaps because they were recruited from among local traders and therefore lacked the incentive to promote the platform to others, although they were regular users of Kudu for their own trading operations. Ultimately, the study team hired staff to do this promotion, which was much more successful in generating broad awareness of Kudu.

messages sent every two weeks.[11] These messages included an advertisement and information on how to trade on the platform, either by registering directly or by contacting a local Kudu staffer.[12]

Table B.4 examines the balance of the market survey for the core variables measured (price, number of traders, and quality rating), while Table B.5 uses the market survey data in dyadic form and examines the baseline balance of the experiment on price gaps across markets. The experiment is well balanced at the market level. For the trader and household analysis, balance is analyzed using the sample still present at endline and is weighted using the attrition weights so as to mirror the structure of the outcome analysis. Table B.6 analyzes the baseline attributes of traders across seventeen different attributes and finds no evidence of baseline imbalance. Table B.7 conducts the same exercise for households, finding two out of seventeen outcomes significantly different at the 10% level and one at the 5% level, in line with what we would expect by chance.[13]

2.5 Platform Usage

Over the three years of the study, Kudu received 29,308 unique asks and 31,177 unique bids. Maize accounts for 67% of asks on the platform, though 19 total crops were successfully traded, with the next most common being soya, rice, and beans.[14] Bids (offers to buy) and asks (offers to sell) posted on the platform each day climbed through the first year to reach a steady maximum of about 400 tons of per day in bids and 200 tons in asks.[15]

Figure A.5 shows the spatial distribution of asks, indicating study markets across the country posting upwards of 1,000 asks each. Among the individuals posting asks to sell on Kudu, 45% were sampled study traders, 14% were agents of an agribusiness firm contracted to promote

[11]This message was sent to all treatment traders, as well as a randomly selected three-quarters of treatment farmer households. This farmer-level randomization, which was conducted at the household level (blocked on subcounty), was set up in order to generate exogenous variation in direct use of Kudu. The goal was to separately estimate the direct impact of using Kudu vs. the indirect impact of living in an area in which others were using Kudu and therefore being exposed to GE effects of others' use, but not directly using the platform oneself. In practice, this second randomization offers a negligible first stage, as farmer take-up is close to zero. See Section 7 for further discussion.

[12]This message also included price information for maize, beans, matooke, and tomatoes in users' local market and main downstream hub market, as well as five randomly sampled treatment markets that rotated in each two-week round. However, as we describe in Section 3.3, we also conducted a separate experiment in which we sent only price information, and see no impact. We therefore interpret this price information component of the intervention as having negligible impact.

[13]Appendix Figure A.4 and Tables B.2 and B.3 present tests comparing attrition in the treatment to the control across the three data types, and documents that attrition is balanced across treatment. Among all the tests that we conduct, only the intensive tracking rate in the trader survey appears differential, with more control traders found than treatment, but given that this arises from finding 14 of 14 control traders versus 24 of 27 treatment traders who were intensively tracked, this has relatively little influence on study-level effects. Overall, weighted attrition rates are very low and the overall unweighted attrition rate from the combination standard and intensive tracking is similar across treatment arms for all data types.

[14]This information comes from administrative data on platform usage. All transactions take place off-platform, but Kudu also implemented a robust system of phone surveys for tracking completed transactions.

[15]Standing up supply and demand simultaneously was an issue at inception of the project; an initial surge of asks in the first season overwhelmed demand, but then a drive to encourage buying on the platform was highly successful and for the remainder of the project the total demand on the platform exceeded supply.

Kudu, and 6% were sampled study farmers. For those posting bids to buy, the corresponding percentages are 48% and 11% for sampled traders and agents. Less than 1% of sellers are sampled farmers. The remainder of bids and ask came mostly from within the treated study areas, but from individual users who were not sampled in our data collection. Overall, Kudu saw 7,300 tons of grain successfully transacted during the study period, worth a total value of about $2.3 million USD. Figure A.8 shows the cumulative sales over the platform during the duration of the study.

Take-up of Kudu among treated traders was high, with 80% posting to the platform at least once and 22% successfully trading on the exchange. However, take-up was much lower among farmers, with only 26% of treated households ever posting to the platform and fewer than 2% successfully transacting on the platform. Therefore, Kudu was a system primarily used by intermediaries rather than for farmers to sell their output directly. Appendix Table B.8 presents predictors of take-up for both samples. We see that scale, in terms of quantity sold or traded, is an important predictor of adoption, a point to which we will return later.[16]

3 Reduced Form Impacts

We now turn to the reduced form effects of the platform on trade flows and market prices. We use the phrase "reduced form effects" to refer to the standard experimental measure of post-treatment differences in outcomes between treatment and control markets or trade routes (i.e. between pairs of markets). As we will address explicitly later in the paper, this may not capture the full treatment effect of the intervention, as prices and trade flows in control markets may be affected through general equilibrium forces. However, this reduced form variation will be key to estimating our model and ultimately quantifying full treatment effects.

3.1 Route-level Effects

Figure 2 presents reduced form effects on several measures of trade between subcounties: whether any trade is occurring, the number of traders engaged in trade along the route, and the volume of trade flowing along the route. These outcomes are drawn from our trader survey data. Figure 3 presents reduced form effects on price gaps between markets, using data from our market-level price surveys. The top rows of Figures 2-3 present non-parametric local Fan regressions of each outcome on the distance between the subcounties, separately for treatment and control routes. Treatment routes are those between pairs of subcounties or markets in which both destination and origin are in the treatment group and can therefore possibly be connected via Kudu; control routes are all other routes. Distance is measured as the road distance of the shortest route connecting the two.

[16] Interestingly, the only other significant predictor of adoption is gender. We find female farmers are less likely to use Kudu. We see weaker, although similarly signed effects among traders, though there are very few female traders overall.

We begin by noting some important patterns observed on control routes. In the top left panel of Figure 2, we see that while the probability of any trade is high for short routes (i.e. those connecting subcounties that are close to one another), this diminishes rapidly with distance. The probability of any trade occurring is close to zero beyond 200km distance. Consistent with this, the number of traders serving a route (second column) also falls quickly with distance. Total trade volumes (third column) also decline in distance. Finally, price gaps are larger between markets located at further distances, as shown in Figure 3.

When we look at these outcomes along Kudu-treated routes, we observe a higher probability of any trade (first column of Figure 2) and a larger number of traders engaged in trade (second column), suggesting greater trade on the extensive margin. We also see larger volumes of trade flows (third column). These increases in trade flows come with reduced price gaps, as shown in Figure 3. Notably, these effects are concentrated among relatively short routes. Beyond about 200km, markets do not trade directly, and the introduction of Kudu does not appear to alter patterns of trade.

To gauge the statistical significance of these differences, the bottom rows of Figures 2-3 present results from a randomization inference procedure, in which 1,000 placebo randomized treatment assignments are drawn. The "reduced form effect" – i.e., the difference between treatment and control line for the realized randomization in the top row – is shown in black. The gray lines show the same "effect" for each of the 1,000 placebo treatment assignments. Long and short-dashed lines indicate the 90th and 95th confidence intervals of these placebo "effects." Appendix Figures A.9 and A.10 present the resulting p-values from this randomization procedure. Consistent with the top row, which suggests treatment effects are concentrated among relatively nearby markets, we see statistically significant effects for routes at short distances. Beyond 200km, we see that effects taper off to precise zeros.

We present the same results on trade flows and price gaps in regression form in Tables 1 and 2, respectively. We run the following specification:

$$Y_{dr} = \alpha + \beta_1 T_d + \beta_2 D_d + \varepsilon_{dr} \tag{1}$$

where Y_{dr} is the outcome of interest along route (market dyad) d in survey round r. Outcomes are regressed on T_d, a dummy for whether the route is treated and D_d, a measure of the shortest road distance between the pair of markets. Standard errors are clustered two-way by each subcounty (the unit of randomization).

The top panels (Panel A) of each table present results for the full sample of routes, while the middle (Panel B) and bottom (Panel C) panels break down results for short and long distance routes, respectively. Panels A of both tables suggest impacts across the full sample of routes that are consistent with those observed in Figures 2-3, with effects in the expected direction but which are only marginally or not significant. However, the middle panels document that effects are concentrated along short routes, with significantly higher probability of trade, a larger number of traders, and higher trade volumes (Panel B Table 1), as well as lower price gaps (Panel B Table 2), along routes less then 200km. In contrast, we see fairly precise zero impacts of the platform on trading outcomes (Panel C Table 1) and no significant effect on price gaps (Panel C Table 2) for far distance routes above 200km, as shown in the bottom panels. While many practitioners hope that platforms like Kudu will help increase long distance trade (e.g., connecting farmers to urban markets), in reality we find that the platform generates meaningful differences in trade flows and price gaps only among nearby markets.

3.2 Market-level Effects

We have shown that trade increased and the difference in prices fell between treated pairs of markets. We next turn to analyzing effects on "monadic" outcomes – those that are market-specific rather than "dyadic" or route-specific.

We investigate effects on market prices as well as outcomes for traders and farmers based in treated markets. Note that impacts are likely be heterogenous, and in particular, to differ between surplus markets (where we might expect treatment to generate increases in net outflows) and deficit markets (where we might expect the opposite) as trade increases and arbitrage improves. In the analysis that follows, we define the surplus-deficit status of each area based on average marketed surplus per farmer at baseline.

3.2.1 Price Impacts

Figure 4 presents reduced form effects on price levels in relative surplus versus relative deficit areas. First, we note in the left panel that, as expected, prices are higher in relative deficit areas and lower in relative surplus areas in the control group. However, we see a less steep relationship between prices and surplus-deficit status in the treatment group. Prices are relatively lower in treated deficit markets compared to control markets, and higher in treated surplus markets. The right panel presents this reduced form effect, along with the 90% and 95% bootstrapped confidence intervals. We see that prices are weakly lower in deficit areas and statistically significantly higher in surplus areas.

Table 3 presents similar results in regression form. We see in Column 1 that the overall effect on average price levels is a statistical zero. This is consistent with the netting out of two competing effects seen in the previous figure (the density in the right-hand panel of Figure 4 shows that for the median trading center, the average price effect is roughly zero). Column 2 presents heterogeneity by baseline average marketed surplus, where we again see that prices are weakly lower in relative deficit areas and higher in relative surplus areas. With an average baseline marketed surplus of about one ton, these effects almost exactly offset each other for the median market.

3.2.2 Trader Impacts

We saw in Section 3.1 that there are a greater number of traders active along treated routes. Who are these additional traders? Table 4 documents that traders operating along treated routes are smaller on average than those operating on control routes, as measured by baseline profits, volume traded, and value traded (Columns, 1, 3, and 5). (It is critical that these measures of size are pre-treatment, so that this analysis sheds light on the *selection* of traders into treated trading routes, and not the effect of treatment on a given set of traders.) Columns 2, 4, and 6 show that, in particular, treatment allows a drop in the *minimum* size among traders active along a route. This is consistent with Kudu lowering the fixed cost of serving treated routes and drawing in smaller entrants, a point to which we return in Section 4.

In Table 5, we turn to Kudu's effect on trader business outcomes, regressing firm-level post-treatment outcomes on the trader's home market treatment status. We find that while treated traders' total quantity traded goes up (Column 1), the buy-sell margin at which they trade goes down (Column 2). While these results are imprecisely measured, they are consistent with Kudu enabling traders to find new markets and expand their trading volumes, but – as other traders are also exposed to Kudu – the reduction in price gaps between markets cuts into their buy-sell margins. Indeed, self-reported trader profit falls as a result of the intervention.

These changes however, are not sufficient to alter exit or entry into the trading industry as a whole, as opposed to entry into serving particular routes. We see in Column 3 that there is no effect on the probability that treated traders are still in business post-treatment. We also re-conduct our trader census at endline, to order to measure whether treatment affects entry. In Column 4, we see no effects on the number of new traders in business in treated markets.

3.2.3 Farmer Impacts

We next turn to looking at outcomes for farming households. Consistent with the zero average effect on market prices documented in Section 3.2.1 we see in Table 6 that there is no statistically significant effect on total revenues, maize revenues, maize volumes sold, or price received for maize sold, for farmers living in treated subcounties as a whole. Point estimates are positive and, in some cases, quite large (for example, the point estimate on total revenues is 9.7% of average revenue), but estimates are imprecise.[17]

Although some practitioners might hope that Kudu would benefit farmers across the board, it seems more likely that, following impacts on price levels, outcomes should vary by the surplus-deficit status of the market. Recall that Kudu increased prices in surplus areas and reduced them in deficit areas. Figure 5 therefore displays treatment effects on farmer revenue by the relative surplus status of their home market. In the left panel we see that farmer revenue increases with

[17] Note that the farmer sales price data is available only for those who report sales (either at farmgate or at market) but we do not find a treatment effect of Kudu on reporting a price or on the location of sale).

surplus status in both treatment and control. However, the right panel shows that treatment significantly increases farmer revenues in surplus areas (and decreases revenues in deficit, although this effect is not statistically significant). This suggests that shifts in market prices can impact farming households, even though most did not directly use Kudu or engage in cross-market trade.

We also test whether Kudu affected production, either on average or differentially by surplus status. One might imagine that as farmers receive higher prices in surplus areas, this could drive increases in production, for example, through increased input usage. The converse might be true in deficit areas. However, as shown in Appendix Table B.9, we find no significant impact of Kudu on harvest levels. Although point estimates go in the expected direction – positive in surplus areas and negative in deficit – they are far from significant. This suggests that the increase in trade we observe arises mainly from changes in the allocation of existing production, rather than through increases in productive specialization. It is of course possible that such production changes require more time to adjust, beyond the three years encompassed by our study. If that is the case, the gains from Kudu we document during the study period are a lower bound on those that could emerge in the long-run.

3.3 Unpacking Matching Frictions

Both the route- and market-level reduced form effects presented suggest that the intervention lowered bilateral trade costs and therefore improved arbitrage. In the following section, we will write down a model that is consistent with all the observed effects, in which Kudu reduced the fixed cost of searching for potential transaction partners on treated routes by providing users with a quick, easy, and potentially profitable match.

The unique feature of Kudu is the direct creation of these new buyer-seller linkages. We also, however, provided randomized price information in a number of different ways during the three years of the study, and so have the ability to speak to the literature that has previously focused on the role of price information. Using these sub-experiments, we show that price information alone has very limited impact on market integration in our setting (perhaps because cell phones are already ubiquitous) and further positive evidence in favor of the role of buyer-seller matching frictions. These findings influence the way we model and interpret the treatment effects of Kudu.

We have two sub-experiments designed to test the role of price information alone. First, we randomly rolled out SMS-based price information alerts to a subset of control markets.[18] We rolled in three control markets in each of the 12 market survey rounds between October 2016-March 2017 (roughly the second half of the study period), and then, subsequent to the household and trader endline surveys, we rolled in an additional 36 control trading markets and observe a final four rounds of market surveys with this system in place. Because this roll-in did not include promotion

[18] Recall from Section 2.4 that these price alerts included information on maize, beans, matooke, and tomatoes at the recipient's local market, downstream hub market, and closest major market, as well as five randomly sampled treatment markets each week.

of Kudu or any support in using the platform, its impact isolates the effect of price information alone. To analyze this price information-only experiment, we can then match the price gaps from the biweekly market data to the timing of the introduction of price information to the market. In Table B.10, we present the impact of this sub-experiment, comparing price gaps among markets rolled into the price-only intervention to pure controls, before and after the intervention using panel data with round fixed effects. Unlike the introduction of Kudu, which drove strong price convergence, here we find no evidence of a reduction in gaps in prices across markets that receive only price information (even for those at short distances).

A second sub-experiment shedding light on the role of price information comes from within-treatment variation. As described in Section 2, as part of the price information sent to treated traders and farmers in the main experiment, we randomly selected five treated markets in each round, and broadcasted information about prices in those markets to the entire treated network. Though more short-run, this intervention was quite powerful, as it sent the market's price information to thousands of traders and farmers simultaneously. However, this price-only intervention also had no effect on prices or dispersion in the featured markets (see Figure A.11).

These two sub-experiments suggest that price information did not drive the main treatment effects observed in the study. This is further corroborated by the fact that Kudu increased the flow of trade from net surplus to net deficit regions of the country, a form of heterogeneity that is relatively time-invariant and already well-understood by professional maize traders, the main users of Kudu. We also find no evidence that Kudu reduced high-frequency deviations from expected average prices, as one would expect if the type of search cost alleviated were price information (see Table B.11).

Instead, our evidence suggest that Kudu worked by reducing buyer-seller matching frictions, introducing users to specific other users who were available and willing to trade, and thereby reducing the cost of searching for potential transaction partners.[19] Kudu appears to support the formation of new trading relationships, by reducing matching frictions and introducing new trading partners. We now present a model with this logic at its core.

4 Model

The reduced form empirical patterns presented in the previous section show that the intervention generated experimental variation in aggregate outcomes between treatment and control markets. However, these comparisons are not sufficient to correctly estimate the magnitude of treatment

[19] In fact, our survey data suggests that many of the new trading relationships formed by Kudu were quite durable, outlasting the initial deal formed on Kudu. We find that 43% of the traders who initially matched on Kudu report transacting again with that same individual off-Kudu – and at large volumes, with these repeat transactions accounting for 7x the volume of the deal initially conducted on-Kudu. This is consistent with the fact that the total volume of trade generated by the intervention (based on trader surveys of trading patterns) is 4x the volume of trade between treated markets that is observed in the administrative data from Kudu.

effects, because control markets are likely to have been affected indirectly via their trade connections to treated markets. We can still make use of the experimental variation, but need to interpret it carefully. In this section, we describe a model that will inform how we separate direct effects on treated units from indirect equilibrium effects. This will enable us to use the experimental variation to correctly estimate the effects of access to Kudu.

In order to capture the key empirical patterns presented in the previous section, the model must have two main features. First, it must allow for an extensive margin of trade between markets. At baseline, there are many market pairs that do not trade with one another at all, and we see that access to Kudu increases both the probability of any trade and the number of traders serving a given route. Second, there must be a role for frictions in matching between buyers and sellers, which are alleviated by Kudu. We outline a model that has these two key realistic features, while remaining as simple and comparable to standard quantitative trade models in the literature as possible.

4.1 Model setup

4.1.1 Geography

There are locations $i, j \in \{1, ..., J\}$, each with a continuum of consumers of measure Z_i. Consumers in i are endowed with quantity L_i of a homogeneous crop and income D_i from other activities, which we can think of as an "outside" good.[20]

Trading the crop between locations incurs a multiplicative variable cost, τ_{ij} and a fixed cost F_{ij}, both denominated in terms of the outside good. The outside good is freely and costlessly traded, and its price is normalized to one.

4.1.2 Traders

Each location is home to a continuum of traders, with measure N_i.[21] They purchase the crop in their home market at price p_i, and can resell anywhere by paying the trade costs between their home market and the destination market.

Traders must also pay a variable cost of operation, and are heterogeneous in the level of this cost. Trader ω has operating cost $a(\omega)$, drawn from a Pareto distribution with shape parameter k and minimum a_L, so that the CDF is:

$$G(a) = 1 - \frac{a_L^k}{a^k}$$

[20] We see no treatment effects on harvest levels, either on average or by surplus status (see Appendix Table B.9). For simplicity and to match this empirical reality, we therefore model the environment as an endowment economy.

[21] We take N_i as exogenous, as we see no treatment effects on entry or exit of traders into the trading industry. See Table 5 for details.

4.1.3 Demand

Consumers purchase the crop from traders selling in their home market, and have idiosyncratic match values with individual traders, $\varepsilon(\omega)$, which are drawn i.i.d from a Gumbel distribution with mean zero and unit variance. These represent factors like language, ethnicity, availability, personal characteristics of traders, and location within the market. Thus, while the crop itself is homogeneous, we allow for the possibility that sellers are imperfect substitutes from the perspective of buyers. Consumers make a discrete choice over sellers, buying the crop from the one that maximizes their utility.

Utility is quasi-linear in the numeraire, so that consumers have indirect utility:

$$u(\omega) = y - e + e\left(\mathbf{ln}e - \mathbf{ln}p(\omega) + \mu\varepsilon(\omega)\right)$$

where y is per capita income, e is per capita expenditure on the crop, $p(\omega)$ is the price charged by trader ω, μ is a positive constant parameterizing the degree of substitutability across sellers, and $\varepsilon(\omega)$ is the match value with trader ω.

4.1.4 Search and matching frictions

All agents freely observe prices and trade costs in all locations. However, in order to sell outside their home market, sellers must find specific buyers. We attribute part of the fixed cost of selling in a destination market, F_{ij}, to a search and matching cost, S_{ij}, that a seller pays to match with potential buyers there. This represents the costs of the real search strategies traders report using, such as traveling to the market to find buyers or time spent making calls to find out who is buying on a given day. Once this cost has been paid, the seller is "available" to buyers in that market. The measure of sellers available to sell in any market j is Ω_j. Buyers then draw idiosyncratic match values with each available seller, as described above.

4.2 Equilibrium

4.2.1 Trader optimization

Facing these costs and demand, traders decide which destination markets to sell in, and what prices to charge. In any destination market j, a trader ω whose home market is i earns operating profits:

$$\pi_{ij}(\omega) = \left(\frac{1}{\sigma}\right)\left(\frac{\left(\frac{\sigma}{\sigma-1}\right)\tau_{ij}p_i a(\omega)}{P_j}\right)^{1-\sigma} E_j - F_{ij}$$

where E_j is aggregate expenditure on the crop in j, P_j is the price index in j (defined below), and σ is the own-price elasticity of demand $\sigma = \frac{\mu-1}{\mu}$.

A trader will serve any market where they can earn non-negative operating profits. Therefore, there is a cutoff operating cost a^*_{ij} such that any trader in i with costs less than or equal to this level, $a(\omega) \leq a^*_{ij}$, will serve destination j, and otherwise will not:

$$a^*_{ij} = \left(\frac{P_j}{\left(\frac{\sigma}{\sigma-1}\right) \tau_{ij} p_i} \right) \left(\frac{\sigma F_{ij}}{E_j} \right)^{\frac{1}{1-\sigma}}$$

Given these individual trader choices, we can define an aggregate price index in destination j as $P_j^{1-\sigma} = \sum_i N_i \int_{a_L}^{a^*_{ij}} p_{ij}(a)^{1-\sigma}\, dG(a)$ where $p_{ij}(a)$ is the price charged by a trader with operating cost a buying in i and selling in j.[22]

4.2.2 Bilateral trade flows

We can characterize equilibrium trade flows in terms of three margins: whether there is any trade on a route, the number of traders serving the route, and the value of trade on the route.

There will be some trade on a given ij route if the trader with the lowest operating cost (a_L) can earn non-negative profits:

$$\left(\frac{1}{\sigma}\right) \left(\frac{\left(\frac{\sigma}{\sigma-1}\right) \tau_{ij} p_i a_L}{P_j} \right)^{1-\sigma} E_j > F_{ij} \tag{2}$$

Note that this does not depend on any trader-specific factors – only equilibrium features of the sending and receiving markets, bilateral trade costs, and the minimum of the trader operating cost distribution.

We can find the number of traders and the value of trade on a route in equilibrium by integrating over the set of traders who have low enough costs to serve the route profitably.

The number of traders who will serve a route ij is:

$$N_{ij} = \begin{cases} N_i \int_{a_L}^{a^*_{ij}} dG(a) & \text{for } a^*_{ij} \geq a_L \\ 0 & \text{otherwise} \end{cases}$$

The value of trade on a route ij is:

[22] The profit maximizing price is a constant markup over marginal cost: $p_{ij}(\omega) = \left(\frac{\sigma}{\sigma-1}\right) \tau_{ij} p_i a(\omega)$, where the marginal cost depends on p_i, the price at which the crop was purchased in the home market, $a(\omega)$, the trader-specific operating cost, and τ_{ij}, the variable cost of trading from i to j.

$$M_{ij} = \begin{cases} N_i \int\limits_{a_L}^{a_{ij}^*} r_{ij}(a)\, dG(a) & \text{for } a_{ij}^* \geq a_L \\ 0 & \text{otherwise} \end{cases}$$

where r_{ij} is the revenue or value of trade per trader.

This implies that, for $a_{ij}^* \geq a_L$:

$$N_{ij} \quad = \quad N_i - N_i a_L^k P_j^{-k} \left(\left(\frac{\sigma}{\sigma-1} \right) \tau_{ij} p_i \right)^k (\sigma F_{ij})^{\frac{-k}{1-\sigma}} E_j^{\frac{k}{1-\sigma}} \tag{3}$$

$$M_{ij} \quad = \quad N_i \left(\frac{\left(\frac{\sigma}{\sigma-1}\right) \tau_{ij} p_i}{P_j} \right)^{1-\sigma} E_j \frac{k a_L^{1-\sigma}}{1-k-\sigma} \left(\left(\frac{P_j}{a_L \left(\frac{\sigma}{\sigma-1}\right) \tau_{ij} p_i} \right)^{1-k-\sigma} \left(\frac{\sigma F_{ij}}{E_j} \right)^{-k} - 1 \right) \tag{4}$$

4.2.3 Welfare

Welfare is given by:

$$W_j = Y_j - E_j + E_j \mathbf{ln} \left(\frac{E_j}{P_j} \right)$$

where Y_j is aggregate income. In equilibrium, income Y_j and price indexes P_j are determined by prices p_i and cost thresholds a_{ij}^* such that the equations in Section 4.2.1 and 4.2.2 hold.

5 Estimation

In order to use the model to guide the estimation of treatment effects, we first make more specific assumptions about the form of trade costs and how they may depend on access to Kudu. We then derive equations that correspond to the main outcomes in our data, and use them to estimate the parameters of the model using the experimental variation in trade costs.

5.1 Mapping the experiment to the model

Kudu enables a seller to match with buyers in a distant market by simply posting on the platform, which may substitute for other costly search strategies, such as traveling and making calls. We therefore interpret access to Kudu as affecting the search and matching component of the fixed cost of trade, if agents in both the origin and destination markets can use the platform.

We specify fixed trade costs as: $F_{ij} \equiv \exp\left(\phi + \theta \mathbf{ln} d_{ij} + \beta \mathbf{1}_{ij}^K - v_{ij}\right)$ where $\mathbf{1}_{ij}^K$ is an indicator variable equal to one if both i and j are treated. Fixed costs may also depend on distance between markets d_{ij} (e.g. due to the time needed to drive a truck from one place to another), as parameterized by θ, and an idiosyncratic route specific component $v_{ij} \sim N\left(0, \sigma_v^2\right)$. Our primary parameter of interest is β, which describes the impact of Kudu on bilateral fixed costs.

We specify variable trade costs as $\tau_{ij}^{\sigma-1} \equiv d_{ij}^{\gamma} e^{-u_{ij}}$, where γ parameterizes the dependence on distance between markets, and $u_{ij} \sim N\left(0, \sigma_u^2\right)$ is an idiosyncratic route-specific component.

5.2 Experimenting in Equilibrium

The advantage of our setting is that the experimental roll out of Kudu generates route-specific, exogenous variation in the search and matching component of fixed trade costs. Nonetheless, we face two types of challenges in using this experimental variation.

The first challenge is that, although Kudu only changes trade costs on treated routes, we expect this to lead to general equilibrium changes that affect control markets and routes as well. To see why, recall that prices in each destination market depend on the number of traders from all other markets that sell in that destination. This is determined by route-specific cutoff operating costs, a_{ij}^*, which in turn depend on not only route-specific trade costs, but also on prices in source markets. This means that experimentally-driven changes in treated source market prices will change trade on routes linking treatment and control markets (even though trade costs don't change), and therefore change control market prices.

These equilibrium impacts on prices imply a Stable Unit Treatment Value Assumption (SUTVA) violation – we cannot obtain a correct estimate of of the treatment effects of Kudu simply by comparing treated and control markets or routes. Importantly, this is not the type of spillover that could be avoided with tighter control of compliance with treatment assignment, or by randomizing over larger units. Rather, it is a challenge inherent to experimentation with trade costs, which by nature typically involves linkages between treated and control units. To address this, we combine two strategies, detailed below: we estimate model-implied equations with market fixed effects, and we make use of experimental variation in exposure to treated markets.

The second challenge for our estimation is the importance of the extensive margin of trade in our setting. There are many zeros in our data on the number of traders and value of trade at the route level. This means that the log-linear estimating equations typically derived from trade models may suffer from both a selection bias in the set of routes with non-zero trade, and a bias due to heterogeneity in the size composition of traders serving each route. We address this with Heckman-type structural corrections to our estimating equations, as proposed by Helpman et al. (2008).

5.3 Estimating Equations

Following the model and specification of trade cost, we derive three estimating equations. Each corresponds to an outcome observed in Figure 2 and therefore maps to the reduced form variation driven by the experiment.

5.3.1 Any trade

Equation 5 is derived from Equation 2, and describes the probability that there is any trade observed between market i and market j:

$$\mathbf{Pr}\left(T_{ij}=1\right)=\Phi\left(\hat{\zeta}_0+\hat{\zeta}_j+\frac{(1-\sigma)}{\sigma_\eta^2}\text{ln}p_i-\frac{(\gamma+\theta)}{\sigma_\eta^2}\text{ln}d_{ij}-\frac{\beta}{\sigma_\eta^2}\mathbf{1}_{ij}^K\right) \tag{5}$$

The likelihood of trade is lower when the wholesale price in the sending market is higher (mediated by how price sensitive buyers are, via σ), when markets are further apart in distance (mediated by the elasticities of fixed and variable trade costs with respect to distance, θ and γ), and by whether both markets have access to Kudu ($\mathbf{1}_{ij}=1$) (mediated by the treatment effect of Kudu on fixed costs, β). The combined variance of the idiosyncratic shocks to trade costs, σ_η^2, determines how many routes are on the margin of profitability, such that treatment induces them to begin trading. Finally, $\hat{\zeta}_j$ is a destination market fixed effect.

5.3.2 Number of traders

Equation 6 is derived from Equation 3, and describes the number of traders serving a route. For ease of analysis, we redefine the outcome to consider the fraction of traders based in market i that do *not* serve market j:[23]

$$\ln\left(1-\frac{N_{ij}}{N_i}\right)=\varphi_0+\varphi_j+k\text{ln}p_i-\frac{k\left(\gamma+\theta\right)}{1-\sigma}\text{ln}d_{ij}-\frac{k\beta}{1-\sigma}\mathbf{1}_{ij}^K+\varphi_{ij} \tag{6}$$

This states that the share of traders in market i that do not sell in market j is lower when the source market is cheaper (ln p_i), the distance is shorter, and when the route is treated. The shape of the trader operating cost distribution, κ, governs how many new traders are pulled across the threshold of profitability when trade costs fall. Again, φ_j is a destination market fixed effect.

5.3.3 Value of trade

Finally, Equation 7 is derived from Equation 4, and describes the value of trade per trader flowing from market i to market j.

$$\ln\left(\frac{M_{ij}}{N_i}\right)=\psi_0+\psi_j+(1-\sigma)\,\text{ln}p_i-\gamma\text{ln}d_{ij}+B\frac{\phi\left(\hat{z}_{ij}^*\right)}{\Phi\left(\hat{z}_{ij}^*\right)}+\ln\left\{exp\left[\delta\left(\hat{z}_{ij}^*+\hat{\bar{\eta}}_{ij}^*\right)\right]-1\right\} \tag{7}$$

The last two terms address selection into this equation due to treatment-induced extensive

[23]This rearrangement both enables us to estimate the equation using a standard log-linear form, and avoids the need to account for selection on the extensive margin, since there are no zeros in the outcome variable.

margin effects (via a Heckman-style correction following Helpman et al. (2008)), accounting, respectively, for the fact that as treatment reduces trade costs, the set of routes with any trade changes, as does the composition of traders serving those routes. The selection term is a function of $\hat{z}^*_{ij} = \Phi^{-1}\left(\hat{\rho}_{ij}\right)$, where $\hat{\rho}_{ij}$ is the predicted value from Equation 9, and B, which is a coefficient to be estimated. The composition term depends on $\delta \equiv \frac{\sigma_\eta(k-\sigma+1)}{\sigma-1}$, $\hat{z}_{ij}$, and the Mills ratio $\hat{\bar{\eta}} \equiv \frac{\phi\left(\hat{z}^*_{ij}\right)}{\Phi\left(\hat{z}^*_{ij}\right)}$.

The remainder of the equation states that, once we have controlled for those extensive margins, the intensive margin of trade only depends on variables costs, which are themselves a function of sender price and distance, and destination market fixed effects captured by ψ_j.

5.4 Identification

Our goal is to use Equations 5-7 to obtain unbiased estimates of the model parameters. In doing so, we face the two challenges previously mentioned: first, the SUTVA violation implied by the equilibrium effects of the intervention; and second, the biases created by selection on the route- and trader-level extensive margins of trade. The latter is address through rearranging the outcome variable in Equation 6 and the inclusion of additional terms in Equation 7 following the strategy of Helpman et al. (2008), as detailed in the previous section. The former requires further discussion, which we turn to now.

To deal with the equilibrium effects of the intervention, we control for market-level fixed effects, following the standard "structural gravity" approach in the trade literature. Comparing estimating Equations 5-7 to their model counterparts Equations 2-4, note that outcomes determined in equilibrium at the destination market level – in particular, the price index P_j – are now subsumed in destination market fixed effects. Otherwise, these would be in the error term and correlated with both trade outcomes and with the route-level treatment status and would bias our main coefficients of interest. With the inclusion of these fixed effects, our specification uses within-market, route-level variation in treatment to separate the direct effect of Kudu on bilateral trade costs from the indirect effect on market-level prices.

If our only goal was to measure unbiased impacts of route-level treatment on route-level outcomes, we could include sending market as well as destination market fixed effects, which would also subsume the origin price, p_i. However, we would like to separately identify the price elasticity of demand, σ, so that we can not only estimate direct effects of treatment, but also quantify the indirect effects in counterfactual scenarios.[24] However, sending market prices may be correlated with unobserved idiosyncratic components of trade costs. Therefore, we need a source of exogenous variation in price.

Fortunately, the experiment offers a second source of random variation in addition to the route-level treatment status: subcounty-level exposure to other treated subcounties. Formally, we define:

[24] While there is no particular reason to believe that the estimates of γ and θ will be biased in our framework, we are also less concerned about these, as distance is policy invariant.

$$\mu_i \equiv \sum_j \mathbb{1}_{\mathbb{j}}^{\mathbb{T}} \frac{X_j}{d_{ij}}$$

where μ_i is the exposure of subcounty i, defined as the sum of exposure to all other treated subcounties j ($\mathbb{1}_{\mathbb{j}}^{\mathbb{T}} = 1$), weighted by subcounty j's market size (as proxied by j's total bilateral trade flows with all other markets, X_j) and by the inverse of distance between i and j (d_{ij}). As noted by Miguel and Kremer (2004), and formalized by Borusyak and Hull (2023), even if treatment is random, *exposure* is only random conditional on economic geography and other features of the baseline economic environment. Therefore, we follow Borusyak and Hull (2023) and run 1,000 placebo randomization draws. For each placebo draw, we construct the exposure measure for each subcounty. We then demean our realized exposure measure μ_i by the average of the exposure measure under the 1,000 placebo randomization draws, $\bar{z}_i$, such that $\tilde{\mu}_i \equiv \mu_i - \bar{z}_i$. The resulting measure $\tilde{\mu}_i$ captures variation in exposure that arises purely from the realized randomization draw, which is exogenous.

In Table 7, we estimate the impact of exposure to treatment in other subcounties on own price using:

$$lnp_i = \alpha_1 + \alpha_2 \tilde{\mu}_i + \varepsilon_i$$

Column 1 shows a significant impact on own price of randomized exposure to treatment in other subcounties.[25] In Column 2, we construct this randomized variation in exposure separately based on whether the treated subcounties to which one is exposed are themselves in surplus or deficit at baseline. We see that (random) exposure to treated surplus subcounties reduces own price, as demand is diverted to treated surplus subcounties and away from one's own subcounty. Conversely, (random) exposure to treated deficit subcounties increases own price, as supply is diverted to treated deficit subcounties, albeit this latter effect is not significant. In Column 3, we interact own treatment status with these randomized exposure measures. We see the diversion effects are most pronounced among controls, as evidence by the significant coefficients on exposure to treated surplus subcounties and exposure to treated deficit subcounties.

These results are useful in two ways. First, they are evidence of the existence of spillovers, providing empirical validation for our theoretical concern about SUTVA violations. Second, they offer exogenous variation in price that we can use in our main structural estimation. We use the specification from Column 3 in Table 7 as an instrument for sending market price. Intuitively, a route's own treatment status identifies the direct treatment effect on trade costs (β), while

[25]Standard errors are clustered by subcounty, the level of randomization. However, Borusyak and Hull (2023) suggest that this may be insufficiently conservative, as variation in exposure may be correlated at a geographic level greater than that of randomization. We therefore follow Borusyak and Hull (2023) and present randomization inference p-values in the notes of Table 7. Although the coefficient in Column 1 (for which theory has no predicted sign) is no longer significant under randomization inference, results from Columns 2 and 3 (for which theory does predict effects) remain statistically significant under this more rigorous standard.

randomized exposure identifies indirect effects that operate through impacts on market-level changes in prices.

5.5 Results and Parameter Estimates

We estimate the model parameters jointly via generalized method of moments (GMM). The moment conditions are standard, specifying that the residuals from Equations 5-7 are orthogonal to the exogenous variables and the price instruments.[26][27] We calculate GMM standard errors with two-way clustering on sender and receiver sub-counties.

Table 8 presents results. We see sensible estimates for our key parameters. Recall that β governs the direct treatment effect on trade costs. We estimate a β of 0.21. This suggests that the direct impact of Kudu is to reduce the size of fixed costs by about one-fifth, on average, across treated markets. These estimates imply that Kudu was effective in driving sizable reductions in fixed trade costs.

6 Aggregate impact of Kudu

6.1 Model Simulation

To understand the impact of Kudu in general equilibrium, we simulate the full model. The simulation uses the parameter estimates from the previous section,[28] data from our study area at baseline (measuring production, expenditure, and the number of traders in each market, plus the distance between each pair of markets),[29] and random draws of the idiosyncratic parts of variable and fixed trade costs. The final component is an assignment of treatment status to each market, which we vary across simulations as described below to capture actual and counterfactual scenarios. These pieces let us calculate endowments and trade costs, which tell us what trade flows

[26] Parameters that appear in multiple estimating equations have more than one corresponding moment condition. Thus, we are over-identified.

[27] To increase computational efficiency, we use a nested-loop approach. In the outer loop, we search over values for structural primitives of the model (β, σ, γ, θ, κ, σ_η^2, B). In the inner loop, we solve for the set of fixed effects that exactly match the corresponding moments, conditional on the primitives. To address the incidental parameters problem arising from the large number of fixed effects, we use Bayesian shrinkage. We "shrink" the fixed effects by making each a weighted average of the mean estimated fixed effect and the individual estimate. The optimal weight is determined in a cross-validation exercise in which we partition the data into 5 folds. Holding the weight constant, we estimate the model with shrinkage 5 times, leaving out 1 fold each time, and calculate the out-of-sample predictions on the excluded fold. We conduct a grid search to find the weighting factor that maximizes out-of-sample fit, which is alpha = 0.09.

[28] We also calibrate three parameters to approximately match the average level of trade outcomes: the minimum operation cost a_L, the intercept of fixed costs, ϕ, and the relative variance of the fixed versus variable parts of the idiosyncratic components of trade costs, σ_u^2 versus σ_v^2.).

[29] Because we do not see treatment effects on harvests or number of traders per market, we hold these fixed across simulated scenarios, consistent with our modeling assumption, rather than allowing them to be determined endogenously.

will be for all pairs of markets as a function of prices. We then iterate to find a fixed point for prices in all markets in order to solve for the equilibrium.

This set up allows us to see what trade flows and prices look like in equilibrium under a variety of scenarios about access to Kudu. The first we consider is an "actual Kudu" scenario, in which the treatment status of each market is the same as in the real RCT. Using this simulation, we construct a model-based version of our reduced form estimates in Table 1, by comparing treated to control routes. Table 9 presents results. We see a close alignment of between the model and data-based reduced form treatment effects, both overall and by distance.

This exercise also underlines the fundamental limitation of the reduced form estimates: they compare treated routes to control routes, in a world with treatment. However, if control routes are affected by treatment, they no longer proxy for how treated routes would have looked in the absence of treatment, as is typically intended in standard experimental designs. To correctly estimate treatment effects, we make use of a second scenario – critically, one that we cannot observe in the data – in which *none* of the markets in the simulation has access to Kudu. In the following section, we describe how we use this counterfactual to calculate the true treatment effects of Kudu, accounting properly for GE forces and impacts on control markets.

6.2 Treatment effects on trade flows

In the "no Kudu" scenario, everything is the same as in the "actual Kudu" scenario, except that none of the markets are assigned to treatment, and so no route has the reduction in fixed costs induced by Kudu.[30] We solve for trade flows and prices under this scenario as we did with the first.

We can then estimate treatment effects using the ideal comparison: comparing each route in a world with Kudu to itself in a counterfactual world without Kudu. Notably, we can make this comparison both for routes that are treated in the "actual Kudu" scenario and for those that are control. This allows us to calculate the impact on treated routes "(IoT)" and the impact on control routes "(IoC)." We implement this analysis by running a similar specification to our reduced form equation in Equation 1, but replace "Treatment" with an indicator for the "actual Kudu" scenario.

Table 10 presents these impacts on our three main outcomes: whether there is any trade along a route on the extensive margin (Column 1), the number of traders operating along a route (Column 2), and the volume of trade flowing across the route (Column 3). Panel A presents the impact on treated routes (IoT). We see treatment effects are positive and substantial. Treated routes see a 16% higher probability of any trade, a 17% increase in the number of traders operating along the route, and a 6% increase in the volume of trade flows, relative to their mean.[31] Yet, these impacts are smaller than the estimates one would get by taking the simple difference between the treatment and control routes, which we refer to as the ITT (Intention to Treat), shown again in the

[30] In addition to using the same parameters and data, we also hold the random draws for the idiosyncratic components of trade costs fixed, so that we do not mix treatment effects and simulation error.

[31] Table notes present the mean outcome under the "without Kudu" counterfactual.

table notes for ease of comparison.[32] Although the actual impact on the extensive margin of any trade is not very different from the ITT estimate (both are around 16%), effects on the number of traders and trade volumes are substantially smaller than would be suggested by a straightforward ITT analysis, with the impact on number of traders falling from almost 21% of its mean to 17% and the impact on trade volumes falling from over 7% of its mean to 6%.

These overestimates of impacts on treated routes are driven by spillovers to control routes. For the standard reduced form analysis of the experiment to be valid, one typically assumes that impacts on control units are zero. However, we see in Panel B that the impacts on control routes (IoC) are in fact negative. While the impact on the probability of any trade along control routes is small, at a (not significant) 0.5% reduction, we see significant negative effects on the number of traders and volume of trade flows along control routes. Control routes experience a roughly 1% drop in both number of traders and in trade volumes relative to their means, suggesting trade diversion away from control routes.

Panel C presents average impacts across all routes (IoA), which is simply the weighted average of the positive impact on treated routes (IoT) and negative impact on control routes (IoC).[33] Across all routes, Kudu increases the probability of trade by 4%, the number of traders by 4%, and the volume of trade flows by 1% relative to the mean. These estimates suggest that Kudu produced meaningful increases in aggregate trade flows by reducing the fixed cost of search and matching.

However, these effects are smaller than would be suggested by the "naive" IoA, presented in the table notes of Panel C, which assumes the impact on treated routes is the ITT and the IoC is zero. Tables notes also present the magnitude of this mismeasurement, comparing the "naive" IoA to the true IoA, as a percent of the true IoA. We see that naive average impacts are 11% larger than actual impacts on the probability of any trade, 41% larger than actual impacts on the number of traders, and 154% larger than the actual impact on trade volumes. Though small per route, the negative spillovers on control units make up a substantial part of the bias in the IoA as three-quarters of routes are untreated in the experiment as implemented. We return to this point when we discuss implications for the impact of this intervention at full-scale in Section 7.2.

6.3 Welfare and Distributional Impacts

Correctly estimating treatment effects on trade costs and therefore on bilateral trade is an important first step in understanding the role of search costs. However, the objects of interest to policymakers are not trade or trade costs per se, but rather how those translate into market-level outcomes

[32]The ITT uses the same *method* as our reduced form estimates from Section 3. However, we estimate them using the simulated data rather than the survey data, to ensure that differences between these and our main estimates reflect only the the fundamental SUTVA violation and not simulation error.

[33]Recall treated routes are those that are connected by Kudu. As Kudu was randomly introduced in half of the study subcounties, only one-quarter (0.5*0.5) of the origin-destination pairs are connected by Kudu. Therefore, IoA = 0.25*IoT + 0.75*IoC.

relevant to welfare. Our structural approach allows us to calculate treatment effects on these outcomes, accounting properly for GE forces.

Figure 6 presents treatment effects on wholesale prices, producer surplus, consumer surplus, and total welfare in treated markets. In deficit areas (to the left in the figure), Kudu reduces prices on average. This improves consumer surplus, but lowers producer surplus. Conversely, in surplus areas (to the right in the figure), Kudu increase prices on average. This improves producer surplus, but lowers consumer surplus. By reducing trade costs, Kudu therefore generates classic winners and losers from trade, with net producers in surplus areas and net consumers in deficit areas gaining, and net consumers in surplus areas and net producers in deficit areas losing. However, because surplus areas contain more net producers and deficit areas more net consumers, average welfare improves in both locations. On net, we estimate that aggregate welfare improves by 0.033% in treated markets (and 0.019% across all markets, including control).[34]

7 Policy Implications

7.1 Implications of scale economies for take-up

Platforms like Kudu are often motivated by policymakers as a way to 'cut out the middleman' and improve farmer welfare by connecting the rural poor directly to high-priced urban markets. We find, however, that usage of the platform was very limited among farmers, suggesting that a rethink of this logic is in order. In Section 4, we showed that if Kudu reduces the search and matching component of fixed trade costs, we should expect to see smaller agents able to serve more routes profitably. This is precisely what we see among traders in Section 3.2.2. So, why are farmers not similarly being enabled to engage in more trade directly? Our model quantification sheds light on this question.

Figure 7 shows the distribution of transaction sizes for farmers (black line) and traders (grey line) from the survey data. We see that traders are substantially larger than farmers on average, although the two distributions overlap. Using the model simulations from Section 6, we can calculate the minimum "threshold" size necessary to make trade along each route profitable, given the fixed costs. We superimpose these thresholds on the farmer and trader size distributions in Figure 7. The long dash shows the average threshold size in the absence of Kudu, while the short dash presents the – now lower – average threshold size with Kudu.

We note two implications of this figure. First, the decline in the average threshold driven by Kudu shifts the minimum size for transacting to the left. This moves it further into the trader size distribution, suggesting that Kudu should lead to greater entry among traders at the route level, which is exactly what we see in Table 10. Further, Figure 7 implies that these new entrants should

[34]The aggregate welfare effects are small because maize, while extremely important as a single good, still makes up only a small fraction of total consumption.

be smaller than incumbent traders serving the route. This is indeed what we observe, as shown in Table 4.

The second notable feature of Figure 7 is that the Kudu-induced reduction in the threshold size still lies far to the right of the farmer size distribution. Even with a 21% reduction in fixed costs, almost all farmers are still simply too small to make direct engagement in cross-market trade profitable on most routes.[35] This suggests that – despite the commonly cited motivation of using technology to allow smallholders to directly connect to markets – scale economies make this a tall order. Fixed costs would need to fall much further to make this possible. Importantly, search for buyers is only one component of these fixed costs – they may also be driven by transportation or other factors – and so there may not be any information-related intervention that reduces scale economies sufficiently to engage farmers directly in cross-market trade.

In sum, most farmers are on average too small to have any motivation to adopt Kudu, even given a substantial reduction in the minimum threshold size required for cross-market trade. Consistent with this, only 2% of farmers trade on Kudu. In Appendix Table B.12, we also show that Kudu does not affect the number of traders to which farmers sold in the 12 months prior to endline, the number of *new* traders to which farmers sold, or whether farmers made any sales at market (rather than farmgate). Overall, Kudu does not appear to alter farmers' traditional marketing channels.

However, given the substantial general equilibrium effects of the platform, this does not mean that farmers are not impacted by Kudu. As traders take-up Kudu and cross-market arbitrage improves, shifting market prices can impact farming households. Indeed, we see in both the reduced form results and our model quantification that prices rise in surplus areas, and we showed in Figure 5 that this translated to increases in farmer revenue. Although this does not operate through the mechanism that policymakers may have had in mind – that mobile marketplaces like Kudu will enable the smallest, poorest farmers to directly access a wider market and bypass intermediaries – it appears that farmers can benefit from the passed through effects of arbitrage by those intermediaries.[36]

7.2 Impacts under scaled implementation

Network interventions like Kudu would typically be implemented universally; markets function best when deepest and there may be a natural monopoly dimension to this type of trading platform. Hence the ideal policy comparison is the impact at 100% saturation relative to the pure control at 0% saturation, neither of which has any empirical counterfactual, as there are no control units in

[35] We show here the *average* minimum transaction size across all routes. We do see about 2% of farmers take-up. These are the largest farmers, and they tend to engage in very short distance cross-market trade, for which the threshold is lowest.

[36] It is worth noting that some pass-through of market price shifts to the farmgate will occur under any model of competition between traders, ranging from perfect competition to monopsony. Of course, the *degree* of pass-through – a key feature governing the *magnitude* of farmer gains from a primarily trader-used intervention – will be mediated by the degree of competition (and the shape of farmer supply). This is something we aim to explore in future work.

the former and no treated units in the latter. We can therefore only use model-based inference to examine these questions.

We conclude by examining this question in more depth, as well as illustrating how standard experimental inference would diverge from the true treatment effects across the distribution of treatment intensity. The standard experimental estimate at any saturation s is just the difference between the treatment and control at that saturation, which we refer to as ITT_s. In situations in which SUTVA holds, then $IoT_s = ITT_s$, $IoC_s = 0 \forall s$, the IoA_s at any saturation is simply $s * ITT_s$, and the average projected impact under universal treatment is the observed ITT. In this context, scaling simply means 'treating more units'. With GE effects, the evaluation of program scaling is more complex. Now, the $IoC_s \neq 0$, the $IoT_s = ITT_s + IoC_s$, and the $IoA_s = s * IoT_s + (1-s) * IoC_s$, or $s * ITT_s + IoC_s$. This latter formulation is informative because it shows that the spillovers to the control lead to incorrect experimental inference about everyone in the study (the IoC_s is not multiplied times s). The 'naive' IoA_s is the one we would calculate if maintaining SUTVA, namely $s * ITT_s$, which will not equal the correct IoA_s in the presence of spillovers. With impacts on both treatment and control units in GE, the IoA_s becomes the obvious basis for evaluation of the total effect of the program.

Figure 8 shows what happens to our three main outcomes – any trade, number of traders, and trade volumes – as we shift the intensity of treatment in counterfactual scenarios from 0% to 100% of markets. The panels in the top row present the per-route impact on treated routes (IoT) in red, the per-route impact on control routes (IoC) in blue, and average impact across all routes (IoA) in purple. We see that the gains per treated route decline with the fraction of routes that are treated, while the (absolute) magnitude of the negative spillovers to each control route increases. This is because the IoT is driven by the direct reduction in trade costs and counteracted by the indirect GE effect on prices, while the control group has no reduction in trade costs and is only affected by the price change. As saturation increases, the size of the change in trade costs enjoyed by each treated route stays the same, while the GE effect on prices increases. The impact across all routes (IoA) increases with treatment intensity, reflecting a composition shift in which more routes are treated and therefore enjoy the Kudu-induced reduction in trade costs.[37] With universal treatment we see that Kudu induces a 10% increase (over the mean) in the probability of trade, a 14% increase in the number of traders, and a 3% increase in trade volumes (see Appendix Table B.13).

The second row of Figure 8 compares the naive experimental estimate of the IoA_s (dashed purple line) with the true IoA_s (solid purple line). Since the negative spillovers to the control (IoC_s) increase with saturation, the error in the naive IoA likewise increases with the fraction treated. While this bias always arises from the IoC_s, the saturation also mechanically drives the extent to which the total bias is driven by mis-estimation of the treated versus control outcome, as illustrated by the red and blue shading, respectively.

[37]Because this figure shows a route-level outcome, the average effect is weighted by the probability of a treatment-treatment pairing, which is .25 at a 50% market-level saturation).

In the third row of Figure 8, we directly represent the extent of mismeasurement in the two key estimands, the IoT_s and the IoA_s. The dotted red line shows mismeasurement in the naive ITT as a percentage of the true IoT, while the dotted purple line shows the mismeasurement in naive IoA as a percent of the true IoA. IoT mismeasurement is low at low saturation, and generally slopes upward. This validates the informal reasoning that often motivates arguments to disregard the potential for bias due to SUTVA violations in small-scale RCTs: if the prevalence of treated units in the population is low, then spillovers will be small per-control unit that is designed to serve as a counterfactual for treated units. Because the per-unit IoC is negligible, so is the bias in the IoT, which is what we see in Figure 8. If the goal is to estimate an accurate IoT for an intervention that will not be scaled up to higher saturation levels, then this reasoning is sound.

Importantly, however, this logic does not follow for estimating aggregate impacts, even at a low saturation level. While spillovers on control units are small and diffuse at low treatment saturation, by definition there are many control units. The purple dotted line shows that adding these up can actually lead to mismeasurement of the total impact of a program or intervention that is larger at low saturation. Even when the IoT is quite accurate and there is no intention to extrapolate to a program of a different scale or saturation level, evaluations of low saturation experiments can lead to very misleading conclusions about aggregate impact *even at the scale studied*. In contrast, in our setting, higher saturation experiments yield very biased IoT, but, in fact, more accurate evaluations of total impact (IoA).

Finally, the bottom row of Figure 8 demonstrates the scaling bias directly. For each saturation, we conduct the naive exercise of extrapolating the IoA under universal implementation from the experimental comparison at that saturation, which is simply the ITT_s. This can then be compared to the true model-based IoA_{100}. The quantity represented on the Y axis is the percent over- or under-estimation in the naive estimate of universal implementation that we would derive from an experiment run at each saturation, namely $100 * (\frac{ITT_s}{IoA_{100}} - 1)$. The naive scaled estimates are generally too large, but the degree of bias in the naive estimates falls as the saturation goes towards 100%. Nonetheless, these biases are very large relative to any others we find; with trade volumes for example even the smallest bias suggests a 100% overestimation of treatment effects at universal implementation based on an experiment run at 90% saturation. This bias will be a product both of two factors: the fact that the true IoT changes with s and so experiments run at smaller scale lack external validity for treatment effects observable only at high saturations, as well as the bias to internal validity in the ITT_s arising from spillovers to the control at any saturation. The fact that the overall bias in the scaled estimate decreases at high saturations implies that the former effect dominates: experiments run at close to universal treatment do a better job of capturing the high-saturation treatment effect, and this effect dominates the larger spillover bias present in high-saturation estimates of the ITT.

8 Conclusion

This study shows that search and matching frictions continue to inhibit trade in African agricultural markets. The introduction of a trading platform to facilitate connections between buyers and seller resulted in greater trade flows on both the extensive and intensive margins. These increases in trade reduced price dispersion, increasing prices in surplus areas and decreasing prices in deficit areas. Results are consistent with the platform having reduced the fixed cost of trade by 21%. This led to price changes that yielded benefits for net producers in surplus areas, net consumers in deficit areas, and increased overall welfare in the study area.

Accounting for equilibrium effects is key to correct estimation of the impacts of trade cost interventions, even those randomized "at-scale." Control markets and routes are affected by trade and price effects in general equilibrium, and therefore no longer serve as valid counterfactuals for the estimation of treatment effects. However, experimental variation – when interpreted through the lens of an equilibrium model - can provide estimates of accurate treatment effects, accounting for GE forces. In our context, aggregate impacts on trade flows are 38% of those suggested by naive "reduced form" comparisons of treatment and control routes.

Finally the platform was used almost exclusively by intermediaries. This is consistent with scale economies in trade, which make interventions targeted at engaging farmers directly in cross-market trade unlikely to succeed, on the margin. However, farmers and consumers can benefit indirectly from market price changes due to trader adoption of such platforms. We see evidence that revenues increases significantly for farmers in surplus areas. This suggests that future market access-based interventions need to think carefully about incidence, and how the most effective programs and policies might target groups that are not necessarily the main intended beneficiaries. This is something we aim to explore in future work.

References

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

_ **and Isaac M Mbiti**, "Mobile phones and economic development in Africa," *Journal of economic Perspectives*, 2010, *24* (3), 207–32.

Allen, Treb, "Information Frictions in Trade," *Econometrica*, 2014, *82* (6), 2041–2083.

_ **and David Atkin**, "Volatility and the Gains from Trade," *Econometrica*, 2022, *90* (5).

Atkin, David and Dave Donaldson, "Who's Getting Globalized? The Size and Implications of Intra-national Trade Costs," *NBER Working Paper 21439*, 2015.

Bai, Jie, Maggie Chen, Jin Liu, Xiaosheng Mu, and Daniel Yi Xu, "Search and information frictions on global e-commerce platforms: Evidence from Aliexpress," *NBER Working Paper 28100*, 2021.

Barrett, Christopher B, "Smallholder market participation: Concepts and evidence from eastern and southern Africa," *Food policy*, 2008, *33* (4), 299–317.

Bassi, Vittorio, Ritwika Sen Raffaele Muoio Tommaso Porzio, and Esau Tugume, "Achieving Scale Collectively," *Econometrica*, 2022, *90* (6), 2937–2978.

Beaman, Lori, Dean Karlan, Bram Thuysbeart, and Christopher Udry, "Selection into credit markts: evidence from agriculture in Mali," *Econometrica*, 2023, *91* (5).

Bergquist, Lauren Falcao and Michael Dinerstein, "Competition and Entry in Agricultural Markets: Experimental Evidence from Kenya," *American Economic Review*, 2020, *110* (12), 3705–3747.

_ , Benjamin Faber, Thibault Fally, Matthias Hoelzlein, Edward Miguel, and Andres Rodriguez-Clare, "Scaling Up Agricultural Policy Interventions: Theory and Evidence from Uganda," *Working Paper*, 2023.

Borusyak, Kirill and Peter Hull, "Non-Random Exposure to Exogenous Shocks," *Econometrica*, 2023.

Brooks, Wyatt and Kevin Donovan, "Eliminating Uncertainty in Market Access: The Impact of New Bridges in Rural Nicaragua," *Econometrica*, 2020, *88* (5), 1965 – 1997.

Cai, Jing and Adam Szeidl, "Indirect Effects of Access to Finance," *American Economic Review*, 2024, *114* (8), 2308–51.

_ , Wei Lin, and Adam Szeidl, "Firm-to-Firm Referrals," *Working Paper*, 2024.

Camacho, Adriana and Emily Conover, "The Impact of Receiving Price and Climate Information in the Agricultural Sector," *IDB Working Paper Series, No. IDB-WP-220*, 2011.

Carter, Michael, Alain de Janvry, Elisabeth Sadoulet, and Alexandros Sarris, "Index insurance for developing country agriculture: a reassessment," *Annual Review of Resource Economics*, 2017, *9.*

Dillon, Brian, Jenny Aker, and Joshua Blumenstock, "Yellow Pages: Information, Connections, and Firm Performanace," *Working Paper*, 2024.

Egger, Dennis, Johannes Haushofer, Edward Miguel, Paul Niehaus, and Michael Walker, "General Equilibrium Effects of Cash Transfers: Experimental Evidence from Kenya," *Econometrica*, 2022, *90* (6), 2603–2643.

Fafchamps, Marcel and Bart Minten, "Impact of SMS-based agricultural information on Indian farmers," *The World Bank Economic Review*, 2012, *26* (3), 383–414.

Fink, Gunther, B. Kelsey Jack, and Felix Masiye, "Seasonal liquidity, rural labor markets and agricultural production," *American Economic Review*, 2020, *100* (11), 3351–3392.

Foster, Andrew and Mark Rosenzweig, "Are There Too Many Farms in the World? Labor Market Transaction Costs, Machine Capacities, and Optimal Farm Size," *Journal of Political Economy*, 2022, *130* (3).

Gollin, Douglas and Richard Rogerson, "Productivity, transport costs and subsistence agriculture," *Journal of Development Economics*, 2014, *107*, 38–48.

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

Grant, Matthew and Meredith Startz, "Cutting Out the Middleman: The Structure of Chains of Intermediation," *Working Paper*, 2024.

Helpman, Elhanan, Marc Melitz, and Yona Rubinstein, "Estimating Trade Flows: Trading Partners and Trading Volumes," *The Quarterly Journal of Economics*, 2008, *123* (2), 441–487.

Hildebrant, Nicole, Yaw Nyarko, Giorgia Romangnoli, and Emilia Soldani, "Price Information, Inter-Village Networks, and "Bargaining Spillovers": Experimental Evidence from Ghana," *Working Paper*, 2020.

Jensen, Robert T, "Information, efficiency, and welfare in agricultural markets," *Agricultural Economics*, 2010, *41*, 203–216.

Kaboski, Joseph, Molly Lipscomb, Virgiliu Midrigan, and Carolyn Pelnik, "How Important are Indivisibilities for Development? Experimental Evidence from Uganda," *Working Paper*, 2024.

Karlan, Dean, Robert Osei, Isaac Osei-Akoto, and Christopher Udry, "Agricultural Decisions after Relaxing Credit and Risk Constraints," *Quarterly Journal of Economics*, 2014, *129* (2), 597–652.

Kebede, Hundanol A., "Gains from market integration: Welfare effects of new rural roads in Ethiopia," *Journal of Development Economics*, 2024, *168.*

Kreindler, Gabriel, "Peak-Hour Road Congestion Pricing: Experimental Evidence and Equilibrium Implications," *Econometrica*, 2024, *92* (4).

Miguel, Edward and Michael Kremer, "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities," *Econometrica*, 2004.

Mitra, Sandip, Dilip Mookherjee, Maximo Torero, and Sujata Visaria, "Asymmetric information and middleman margins: An experiment with Indian potato farmers," *Review of Economics and Statistics*, 2018, *100* (1), 1–13.

Moser, Christine, Christopher Barrett, and Bart Minten, "Spatial Integration at Multiple Scales: Rice Markets in Madagascar," *Agricultural Economics*, 2009, *40* (3), 281–94.

Muralidharan, Karthik and Paul Niehaus, "Experimentation at Scale," *Journal of Economic Perspectives*, 2017, *31* (4), 103–24.

Nakasone, Eduardo, "The Role of Price Information in Agricultural Markets: Evidence from Rural Peru," *Working Paper*, 2014.

Newman, Neil, Lauren Falcao Bergquist, Nicole Immorlica, Kevin Leyton-Brown, Brendan Lucier, Craig McIntosh, John Quinn, and Richard Ssekibuule, "Designing and evolving an electronic agricultural marketplace in Uganda," in "Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies" 2018, pp. 1–11.

Porteous, Obie, "High Trade Costs and Their Consequences: An Estimated Dynamic Model of African Agricultural Storage and Trade.," *American Economic Journal: Applied Economics*, 2019, *11* (4), 327–366.

Rudder, Jess and Brian Dillon, "Search Cost and Relational Contracting," *Working Paper*, 2024.

Shamdasani, Yogita, "Rural road infrastructure and agricultural production: Evidence from India," *Journal of Development Economics*, 2021, *152.*

Startz, Meredith, "The Value of Face-to-Face: Search and Contracting Problems in Nigerian Trade," *Working Paper*, 2024.

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

_ **and David Yanagizawa-Drott**, "Estimating Impact in Partial vs. General Equilibrium: A Cautionary Tale from a Natural Experiment in Uganda," *Working Paper*, 2012.

Teravaninthorn, Supee and Gael Raballand, "Transport Prices and Costs in Africa," *World Bank*, 2009.

Vitali, Anna, "Consumer Search and Firm Location: Theory and Evidence from the Garment Sector in Uganda," *Working Paper*, 2024.

Wiseman, Eleanor, "Border Trade and Information Frictions: Evidence from Informal Traders in Kenya," 2023.

Figures

Figure 1: **Price gaps and transport costs.** The y-axis presents the absolute difference in prices across each market dyad (pair) in the sample. The solid black line presents the gap observed in prices across each pair of markets in our sample. The dotted line presents estimated transport costs. To generate this prediction, we asked surveyed traders to report the costs of traveling roundtrip along each of their five most commonly travelled routes and the vehicle size typically used. From this data, we construct an estimate of per kg transport costs, which we then estimate as a function of the km traveled. These transport costs represent an upper bound on the price dispersion that should be observed if transport costs are the only trade costs. The gray area represents excess price dispersion, the portion of price dispersion that cannot be explained by transportation costs.

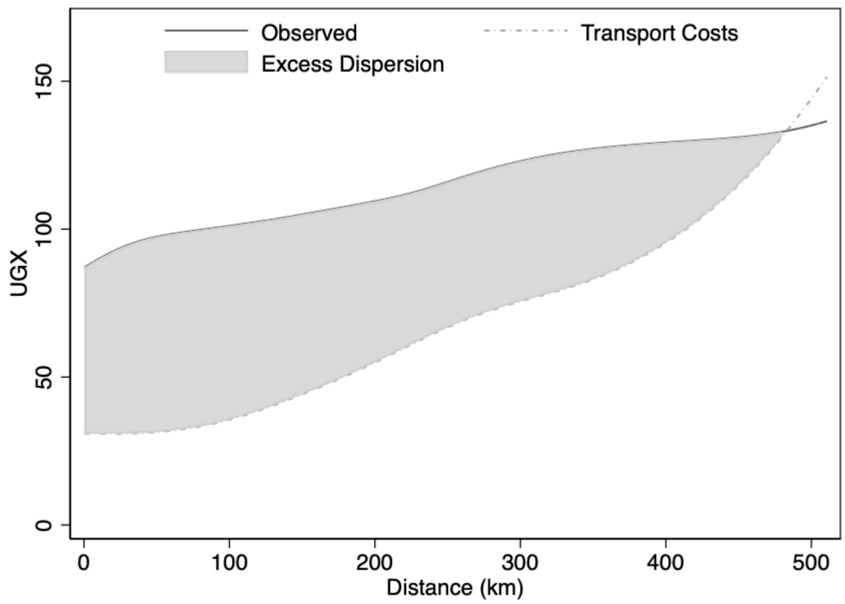


Figure 2: **Reduced form effects on trade flows.** The top lefthand figure presents a non-parametric local Fan regression of the probability of trade along any treated route (dotted line), in which both markets are treated and therefore are connected by Kudu, vs. control routes (solid line), as a function of distance. The bottom row presents results from a randomization inference procedure, in which 1,000 placebo randomized treatment assignments are drawn. The "reduced form effect," i.e., the difference between treatment and control line in the lefthand figure, from the realized randomization, is shown in black. The grey lines show the same effect for each of the 1,000 placebo treatment assignments. Long and short-dashed lines indicate the 90th and 95th confidence intervals of these placebo "reduced form effects." Subsequent columns present similar results for the number of traders trading along a route (second column) and trade volumes along the route (third column). See Figure A.9 for randomization p-values.

Any Trade **Number Traders** **Trade Volumes**

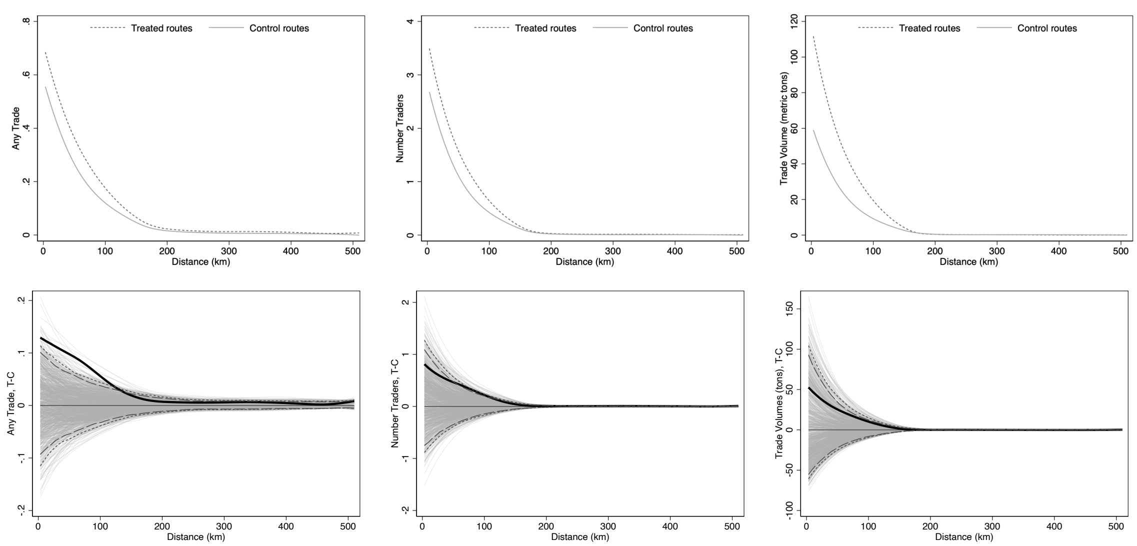
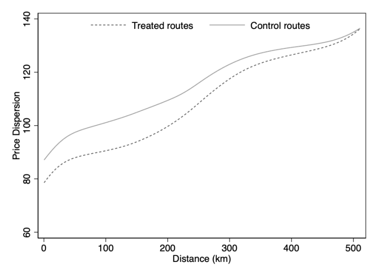


Figure 3: **Reduced form effects on price gaps.** The top panel presents a non-parametric local Fan regression of the price gaps across markets along any treated route (dotted line), in which both markets are treated and therefore are connected by Kudu, vs. control routes (solid line), as a function of distance. The bottom row presents results from a randomization inference procedure, in which 1,000 placebo randomized treatment assignments are drawn. The "reduced form effect," i.e. the difference between treatment and control line in the lefthand figure, from the realized randomization is shown in black. The grey lines show the same effect for each of the 1,000 placebo treatment assignments. Long and short-dashed lines indicate the 90th and 95th confidence intervals of these placebo "reduced form effects." See Figure A.10 for randomization p-values.



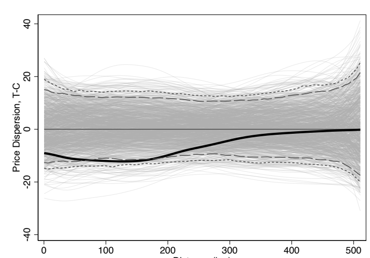


Figure 4: **Market price effects in surplus vs. deficit areas.** The left panel shows the level of market survey prices in treatment vs. control markets, with respect to the average market surplus per farmer, as measured in tons at baseline. The right panel shows the difference between the two (the treatment effect), along with the 90% and 95% confidence intervals from a bootstrap estimation and the density of observations.

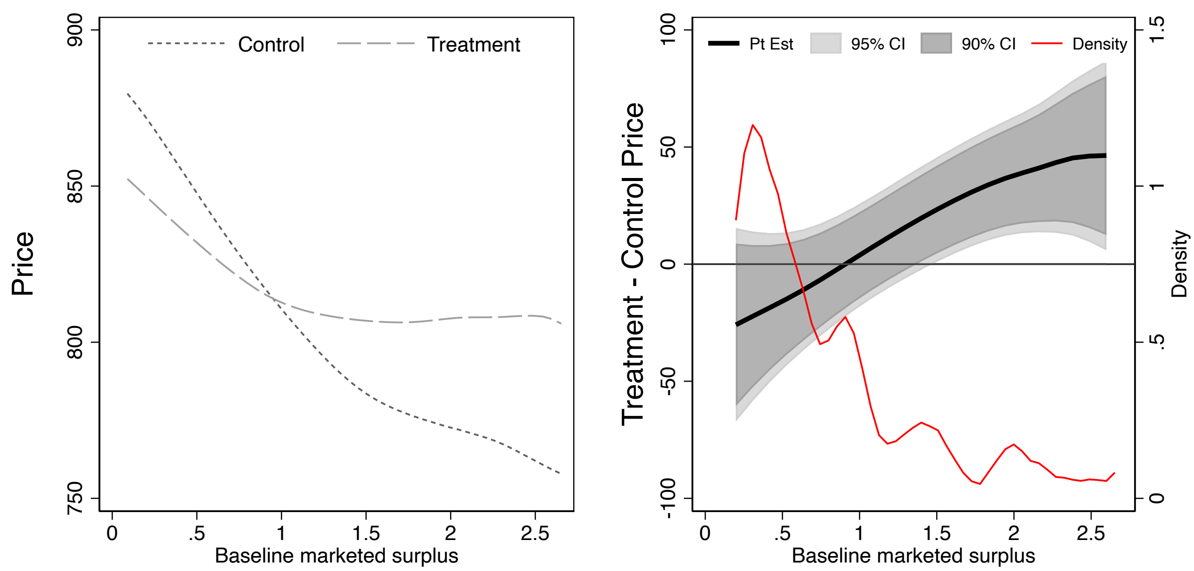


Figure 5: **Farmer revenue effects in surplus vs. deficit areas.** The left panel shows the level of farmer revenue from the household survey in treatment vs. control markets, with respect to the average market surplus per farmer, as measured in tons at baseline. The right panel shows the difference between the two (the treatment effect), along with the 90% and 95% confidence intervals from a bootstrap estimation and the density of observations.

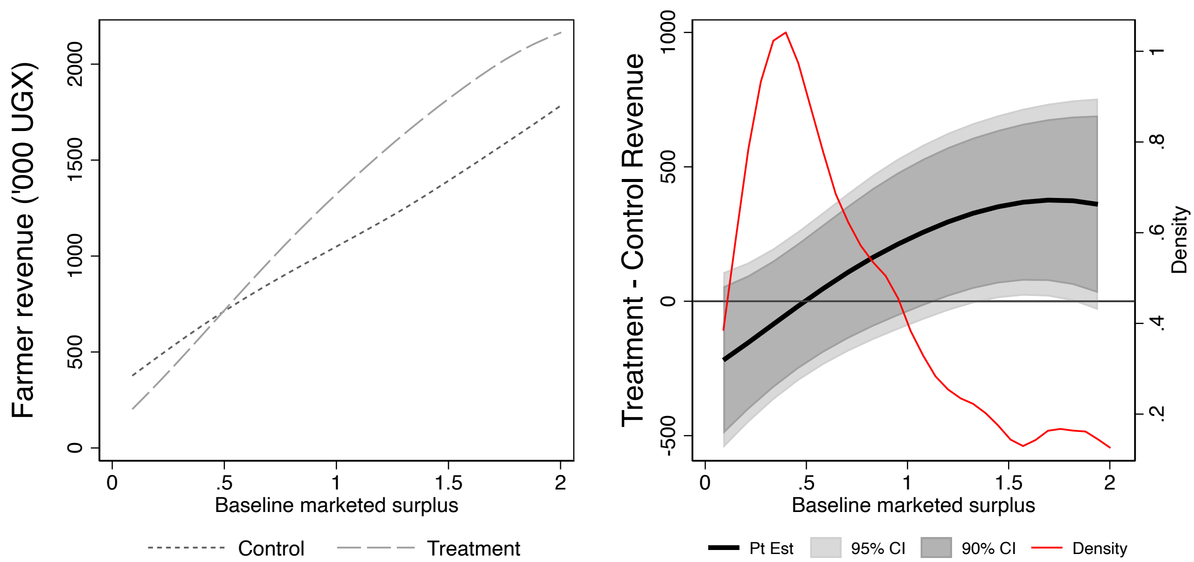


Figure 6: **Welfare effects.** Figure plots the total model-derived impacts of the intervention on producer surplus (short-dash line), consumer surplus (long-dash line), overall welfare (solid line) and prices (dots), as a function of standardized marketed surplus.

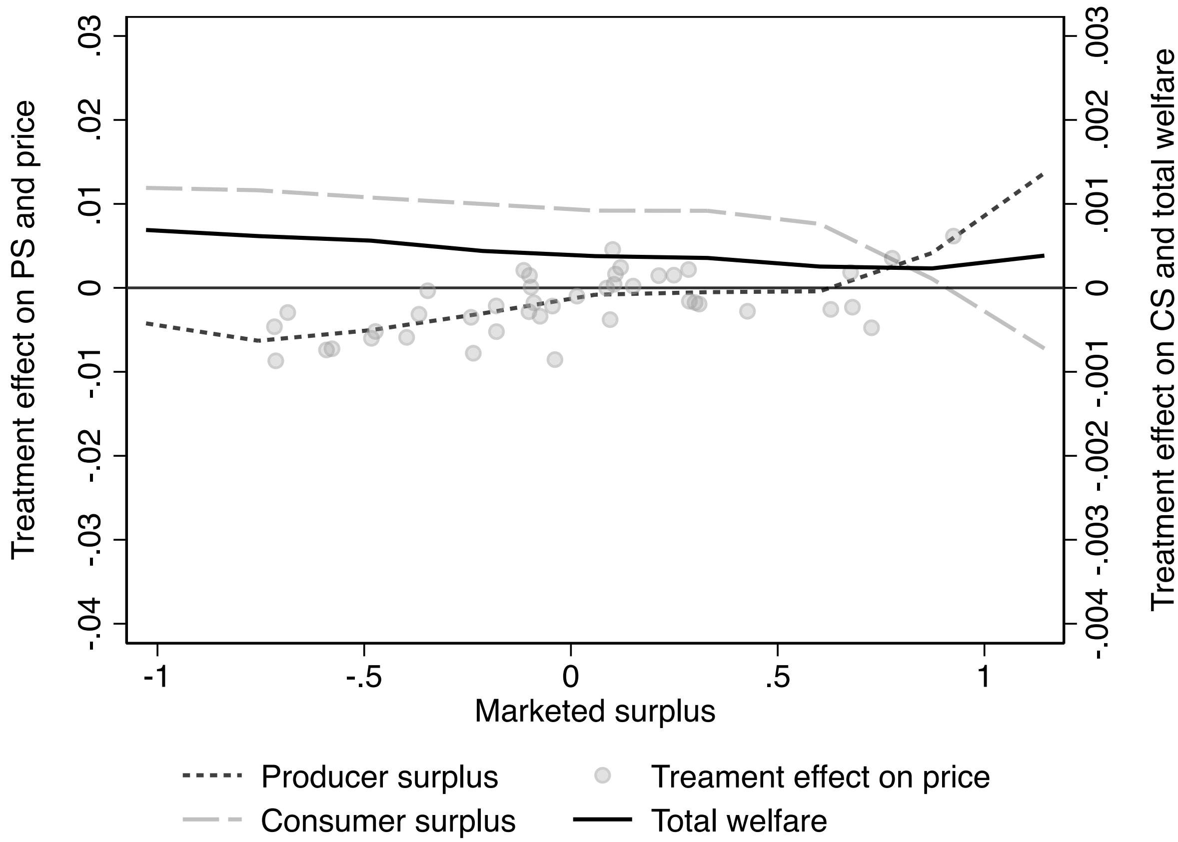


Figure 7: **Threshold size.** The densities represented in this picture are the log transaction size for farmers (left distribution) and traders (right distribution). The vertical lines represent the model-derived threshold size for trade in the absence of the intervention (dashed vertical line) and in the presence of the intervention (dotted vertical line).

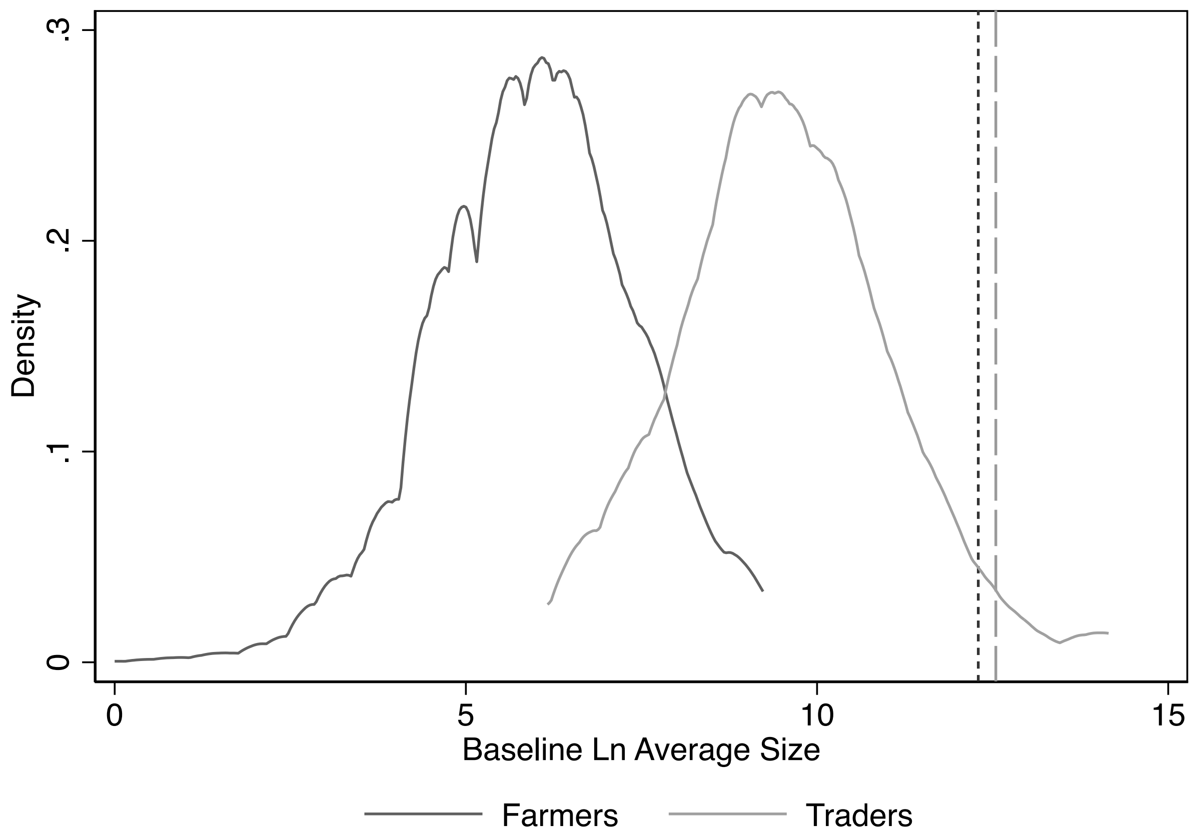
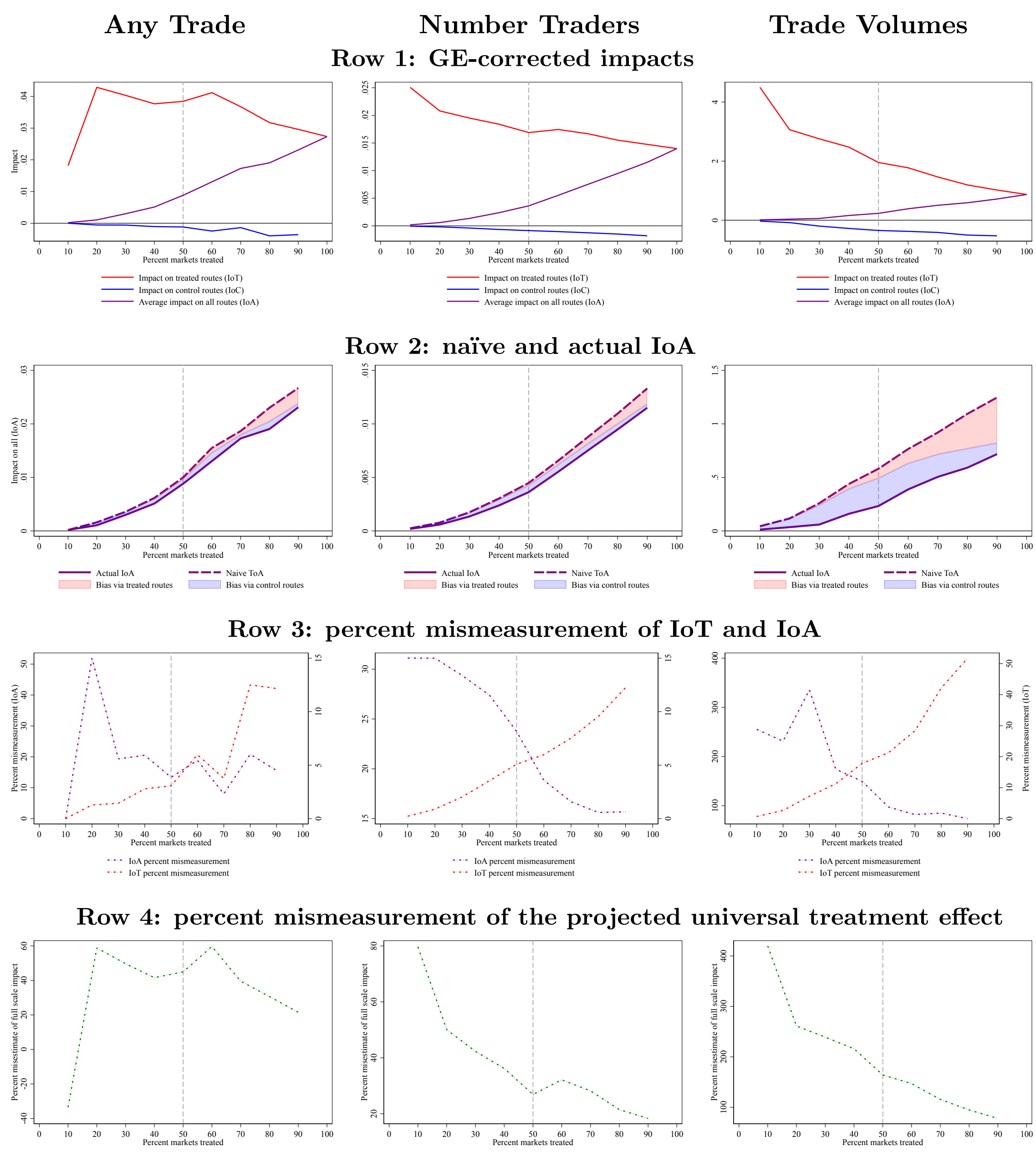


Figure 8: **Impacts by saturation rate.** The panels in the top row present the per-route impact on treated routes (IoT) in red, the per-route impact on control routes (IoC) in blue, and average impact across all routes (IoA) in purple under counterfactuals representing various intensities of treatment, ranging from 0% to 100% (in steps of 10%). Outcomes are any trade along the route (left column), the number of traders trading along the route (middle column), and trade volumes along the route (right column). The second row of Figure 8 compares the naive experimental estimate of the IoAs (dashed purple line) with the true IoAs (solid purple line) under each counterfactual saturation rate. In red we show the IoA bias due to miscalculation of the impact on all treated routes and in blue the IoA bias due to miscalculation of the impact on all control routes. The third row presents the percent mismeasurement in the IoT in dotted red ($100 * \frac{naiveIoT - actualIoT}{actualIoT}$ and the percent mismeasurement in the IoA in dotted purple ($100 * \frac{naiveIoA - actualIoA}{actualIoA}$. The bottom row calculates the naive estimate of the IoA at universal implementation that one would wrongly infer under the assumption that the ITT at a given saturation is the IoA at 100% universal implementation. The is presented as a percentage of the true model-based IoA at 100% saturation.



Tables

Table 1: **Reduced form effects on dyadic trade flow outcomes.** Reduced form results on any trade along the route (first column), the number of traders trading along the route (second column), and trade volumes along the route (third column). Outcome is measured at a route level and regressed on a dummy for treatment (= 1 if both the markets are treated and therefore could possibly be connected by Kudu), distance between the pair of markets, and round fixed effects. Standard errors are clustered two-way by each subcounty (the unit of randomization). Panel A presents the results from the full sample, Panel B for nearby markets (routes less than 200km) and Panel C for far markets (routes more than 200km).

	Any Trade	Number Traders	Volume (tons)
	Panel A: All routes		
Treated route	0.02	0.05	3.03
	(0.01)	(0.05)	(1.90)
Dist (10km)	-0.00***	-0.02***	-0.51***
	(0.00)	(0.00)	(0.14)
Observations	11236	11236	11236
Mean DV	0.05	0.19	4.71
	Panel B: Routes less than 200km		
Treated route	0.06**	0.26*	12.24**
	(0.03)	(0.14)	(6.09)
Dist (10km)	-0.03***	-0.13***	-3.19***
	(0.00)	(0.02)	(0.93)
Observations	3429	3429	3429
Mean DV	0.16	0.61	15.00
	Panel C: Routes greater than 200km		
Treated route	0.00	0.00	-0.05
	(0.00)	(0.01)	(0.10)
Dist (10km)	-0.00*	-0.00*	-0.01
	(0.00)	(0.00)	(0.00)
Observations	7807	7807	7807
Mean DV	0.01	0.01	0.18

Table 2: **Reduced form effects on dyadic price gaps.** Reduced form results on price gaps across markets. Outcome is regressed on a dummy for treatment (= 1 if both the markets are treated and therefore could possibly be connected by Kudu), the distance between the pair of markets, and round fixed effects. Standard errors are clustered two-way by each subcounty (the unit of randomization). Panel A presents the results from the full sample, Panel B for nearby markets (routes less than 200km) and Panel C for far markets (routes more than 200km).

	Price Gaps
Panel A: All routes	
Treated route	-6.36
	(5.09)
Dist (10km)	1.00***
	(0.12)
Observations	1443397
Mean DV	116.30
Panel B: Routes less than 200km	
Treated route	-10.30*
	(5.31)
Dist (10km)	0.76*
	(0.42)
Observations	456811
Mean DV	97.43
Panel C: Routes greater than 200km	
Treated route	-4.29
	(5.62)
Dist (10km)	0.85***
	(0.21)
Observations	986586
Mean DV	125.03

Table 3: **Reduced form price effects in surplus and deficit areas.** In Column 1, monadic data on market price regressed on a dummy for treatment and round fixed effects. Column 2 interacts treatment with the baseline marketed surplus in the subcounty. Standard errors clustered by subcounty, the unit of randomization.

	(1)	(2)
Treatment	-4.801	-20.78
	(11.56)	(16.40)
Baseline marketed surplus		-28.66***
		(7.716)
Treat*Baseline marketed surplus		19.23*
		(9.995)
Observations	15211	15161
Mean DV	831.6	831.6
Mean Baseline Marketed Surplus		0.907
R2	0.844	0.847

Table 4: **New entrants on average smaller.** Columns 1, 3, and 5 show the average size across all traders active on the route post-treatment, while Columns 2, 4, and 6 show the minimum size. Importantly, size is measured by pre-treatment proxies at baseline. This table therefore shows how treatment induces into treated routes new types of traders (specifically, smaller traders). In Columns 1-2, size is measured by baseline profits of the trading enterprise. In Columns 3-4, it is measured by total tons traded (across all routes, at baseline) and in Columns 5-6 by total value traded (across all routes, at baseline). In all columns, standard errors are clustered two-way for each subcounty.

	Profits (mil UGX)		Tons traded		Value traded (mil UGX)	
	Mean	Min	Mean	Min	Mean	Min
Treated route	-1.91**	-2.05***	-76.04	-98.83	-39.37	-48.92
	(0.94)	(0.77)	(81.42)	(77.49)	(34.69)	(32.80)
Log dist (10km)	1.19**	2.05***	170.98*	186.26**	73.59*	81.21**
	(0.48)	(0.46)	(89.11)	(88.49)	(38.50)	(38.21)
Observations	780	780	780	780	754	754
Mean DV	7.39	5.34	255.38	189.14	126.35	91.94
P-value	0.04	0.01	0.35	0.21	0.26	0.14

Table 5: **Effects on trader business outcomes.** The first three columns present reduced form impacts on total trading volume, margins between buying and selling prices, and self-reported profits, regressing each outcome on treatment, its baseline value, and controls selected from baseline covariates set using double lasso, as pre-specified. The fourth column runs an almost identical specification (omitting baseline value controls) for whether the trader is still in business. Finally, the fifth column uses market-level data on the number of new traders who have entered the market as measured in the endline trader census. In all columns, standard errors are clustered at the subcounty, the level of randomization.

	Trading Volumes (tons)	Buy-Sell Margins (UGX)	Self-Reported Profit (UGX)	Still in Business	Number New Traders
Treatment	129.76	-10.96	-1010.77*	-0.02	-0.01
	(106.01)	(9.28)	(554.97)	(0.02)	(0.14)
Observations	2370	2268	2592	2863	236
Mean of DV	242.94	135.01	7279.21	0.86	1.05
R squared	0.04	0.02	0.12	0.01	0.00

Table 6: **Effects on farmer revenues, volumes sold, and prices.** Columns present reduced form impacts on total revenues (in thousands of UGX), maize revenues (in thousands of UGX), quantity of maize sold (in kgs), and price per kg of maize sold (in UGX). Outcomes are regressed on treatment, its baseline value, and controls selected from baseline covariates set using double lasso, as pre-specified. Standard errors are clustered at the subcounty, the level of randomization.

	Revenues Total ('000)	Revenues Maize ('000)	Qnt Sold Maize	Price Maize
Treat	99.2	72.0	61.3	18.0
	(90.9)	(68.5)	(118.2)	(14.0)
Observations	2775	2775	2769	1959
Mean DV	1019	672	1040	631
R2	0.32	0.31	0.33	0.02
Controls	Yes	Yes	Yes	Yes

Table 7: **Impact of experimental exposure.** Exposure is defined as $\mu_i \equiv \sum_j \mathbb{1}_{\mathbb{j}}^{\mathbb{T}} \frac{X_j}{d_{ij}}$, where μ_i is the exposure of subcounty i, defined as the sum of exposure to all other treated subcounties j ($\mathbb{1}_{\mathbb{j}}^{\mathbb{T}} = 1$), weighted by subcounty j's market size (as proxied by j's total bilateral trade flows with all other markets, X_j) and by the inverse of distance between i and j (d_{ij}). We follow Borusyak and Hull (2021) and run 1,000 placebo randomization draws. For each placebo draw, we construct the exposure measure for each subcounty. We then demean our realized exposure measure μ_i by the average of the exposure measure under the 1,000 placebo randomization draws, $\bar{z}_i$, such that $\tilde{\mu_i} \equiv \mu_i - \bar{z}_i$. The resulting measure $\tilde{\mu_i}$ therefore captures variation in exposure that arises purely from the realized randomization draw, which is exogenous. Column 1 regresses ln price on this exposure measure. Column 2 constructs this same procedure but separately for exposure to surplus and deficit market. Column 3 includes treatment interaction terms, allowing the impact of exposure to vary by own treatment status. The table presents standard errors clustered by subcounty, the unit of randomization. Table notes present p-values from randomization inference.

	Ln Price	Ln Price	Ln Price
Exposure to treated markets	-0.15***		
	(0.04)		
Exposure to treated surplus markets		-0.47***	-0.64***
		(0.16)	(0.22)
Exposure to treated deficit markets		0.21	0.71***
		(0.13)	(0.26)
Own market treated x Exposure to treated surplus markets			0.26
			(0.32)
Own market treated x Exposure to treated deficit markets			-0.70**
			(0.29)
Own market treated			0.01
			(0.02)
Observations	22688	22688	22688
RI p-value: Exposure to treated markets	0.66		
RI p-value: Exposure to treated surplus markets		0.04	0.10
RI p-value: Exposure to treated deficit markets		0.59	0.07
RI p-value: Own market treated*Exposure to treated surplus markets			0.39
RI p-value: Own market treated*Exposure to treated deficit markets			0.11
RI p-value: Own market treated			0.88

Table 8: **Parameter estimates.** Parameter estimates from a joint estimation of Equations 5-7 using generalized method of moments (GMM). We calculate GMM standard errors (in parentheses) using two- way clustering on sender and receiver sub-counties.

	Estimate (SE)
β	-0.21 (0.12)
σ	4.65 (0.44)
γ	-0.50 (1.01)
θ	0.84 (1.55)
κ	0.05 (0.00)
σ_η^2	0.94 (0.15)
B	0.95 (1.05)

Table 9: **Model fit.** We use the simulated model of outcomes under an "actual Kudu" scenario, in which the treatment status of each market is the same as in the real RCT. We then construct a model-based version of our reduced form estimates in Table 1, regressing each outcome on a dummy for treatment (= 1 if both the markets are treated and therefore could possibly be connected by Kudu) and distance between the pair of markets. Standard errors are clustered two-way by each subcounty, the unit of randomization. Outcomes are: any trade along the route (first column), the number of traders trading along the route (second column), and trade volumes along the route (third column). Panel A presents the results from the full sample, Panel B for nearby markets (routes less than 200km) and Panel C for far markets (routes more than 200km).

	Any Trade	Number Traders	Volume (tons)
		Panel A: All routes	
Treated route	0.04	0.02	2.28
	(0.03)	(0.01)	(6.24)
Dist (10km)	-0.01***	-0.01***	-2.46***
	(0.00)	(0.00)	(0.38)
Observations	5671	5671	5671
Mean DV	0.27	0.10	32.05
		Panel B: Routes less than 200km	
Treated route	0.05	0.07**	13.31
	(0.04)	(0.04)	(16.66)
Dist (10km)	-0.04***	-0.05***	-11.84***
	(0.00)	(0.00)	(1.78)
Observations	1731	1731	1731
Mean DV	0.54	0.28	81.41
		Panel C: Routes greater than 200km	
Treated route	0.04	0.00	-0.70
	(0.03)	(0.00)	(2.89)
Dist (10km)	-0.01***	-0.00***	-0.54***
	(0.00)	(0.00)	(0.15)
Observations	3940	3940	3940
Mean DV	0.15	0.02	10.37

Table 10: **GE-corrected impacts.** To construct the GE-corrected impacts, we run a similar specification to those in Tables 1 and 9, however, instead of regressing outcomes on a dummy for treatment, we now regress them on a dummy for being in the "actual Kudu" scenario (compared to being in the "no Kudu" scenario). This allows us to calculate the impact on treated routes "(IoT)," shown in Panel A (which restricts the regression to treated routes) and the impact on control routes "(IoC)," shown in Panel B (which restricts the regression to control routes). Panel C runs the same specification with the full sample of routes to calculate the average impact on all routes "(IoA)." We control for distance between the pair of markets and cluster standard errors two-ways by each subcounty, the unit of randomization. Table notes present the "naive" impacts. The "IoT Naive" is the reduced form treatment effect from regressing outcomes on treatment under the "actual Kudu" scenario (this is the same as the estimates shown in Table 9). The "IoC Naive" is simply zero, as naive estimates assume control routes are unaffected by the intervention. Finally, the "IoA Naive" is the weighted average of the two.

	Any Trade	Number Traders	Volume (tons)
Panel A: Impact on treated routes (IoT)			
Kudu	0.0411***	0.0173***	1.9266***
	(0.0065)	(0.0016)	(0.2795)
Dist (10km)	-0.01***	-0.01***	-2.71***
	(0.00)	(0.00)	(0.50)
Observations	2970	2970	2970
IoT Naive (ITT)	0.0418	0.0209	2.2823
Panel B: Impact on control routes (IoC)			
Kudu	-0.0012	-0.0009***	-0.3652***
	(0.0008)	(0.0001)	(0.0739)
Dist (10km)	-0.01***	-0.01***	-2.38***
	(0.00)	(0.00)	(0.42)
Observations	8372	8372	8372
IoC Naive	0	0	0
Panel C: Average impact on all routes (IoA)			
Kudu	0.0099***	0.0039***	0.23***
	(0.0024)	(0.00077)	(0.083)
Dist (10km)	-0.01***	-0.01***	-2.46***
	(0.00)	(0.00)	(0.38)
Observations	11342	11342	11342
IoA Naive	0.011	0.0055	0.60
Naive mismeasurement	0.11	0.41	1.54
Mean DV	0.26	0.10	31.82

Appendix A Figures

Figure A.1: **Maps of the Study Area.** The left-hand panel is USAID's FEWS-Net map of Surplus Maize Areas of Uganda, and the right-hand panel shows the 110 study subcounties.

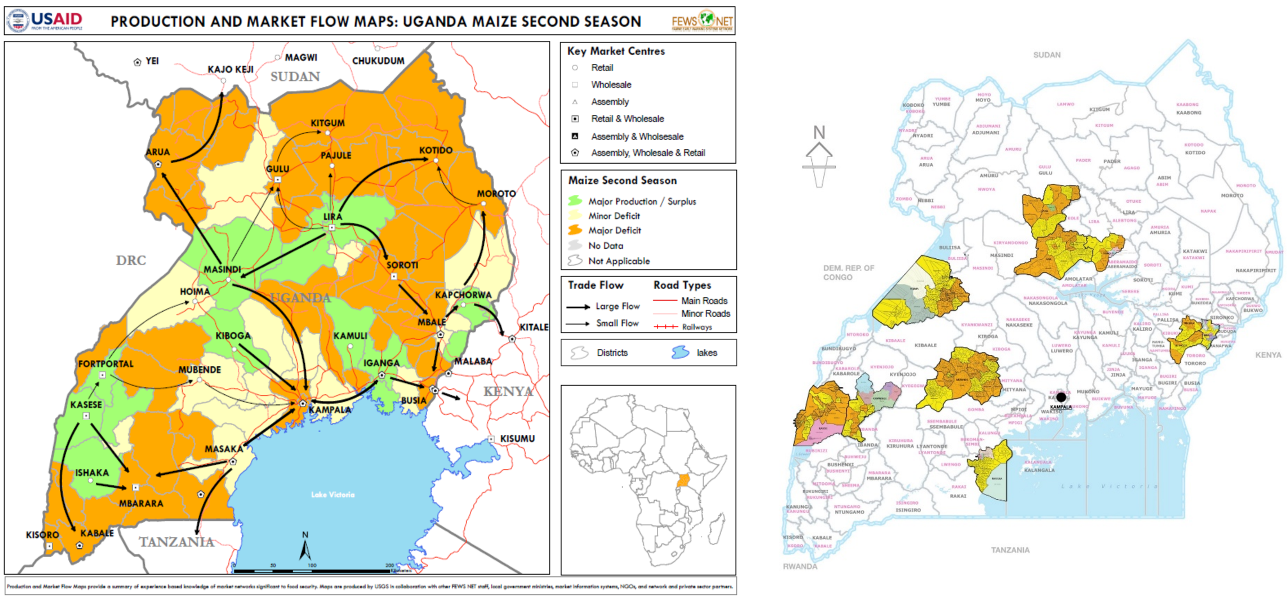


Figure A.2: **Study Timeline.** This figure presents a study timeline. The study ran for three years, starting in 2015 and concluding in 2018, and spans six major agricultural seasons (denoted in orange). During this period, we gathered three core types of data. First, we collected high-frequency market data, which ran throughout the study period at the 236 markets throughout our sample. Second, we surveyed traders at baseline (pre-treatment, in 2015), midline (after one year of treatment), in 2016, and endline (after three years of treatment, in 2018). Third, we surveyed our farmer household sample, one at baseline in 2015 and again at endline in 2018.

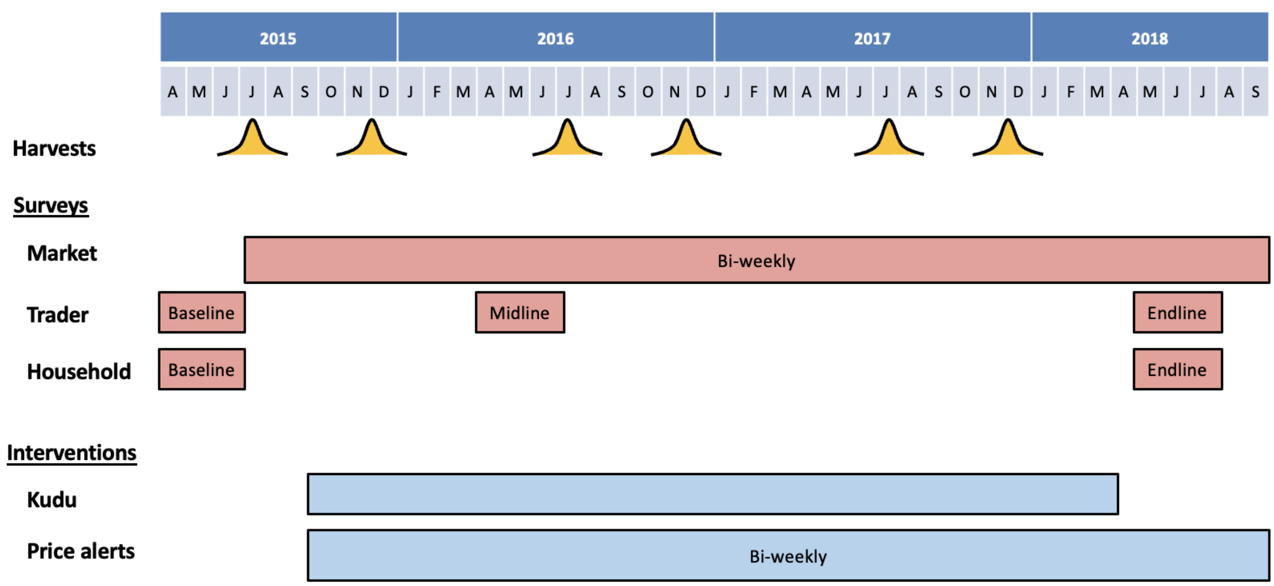


Figure A.3: **Sample and randomization design**

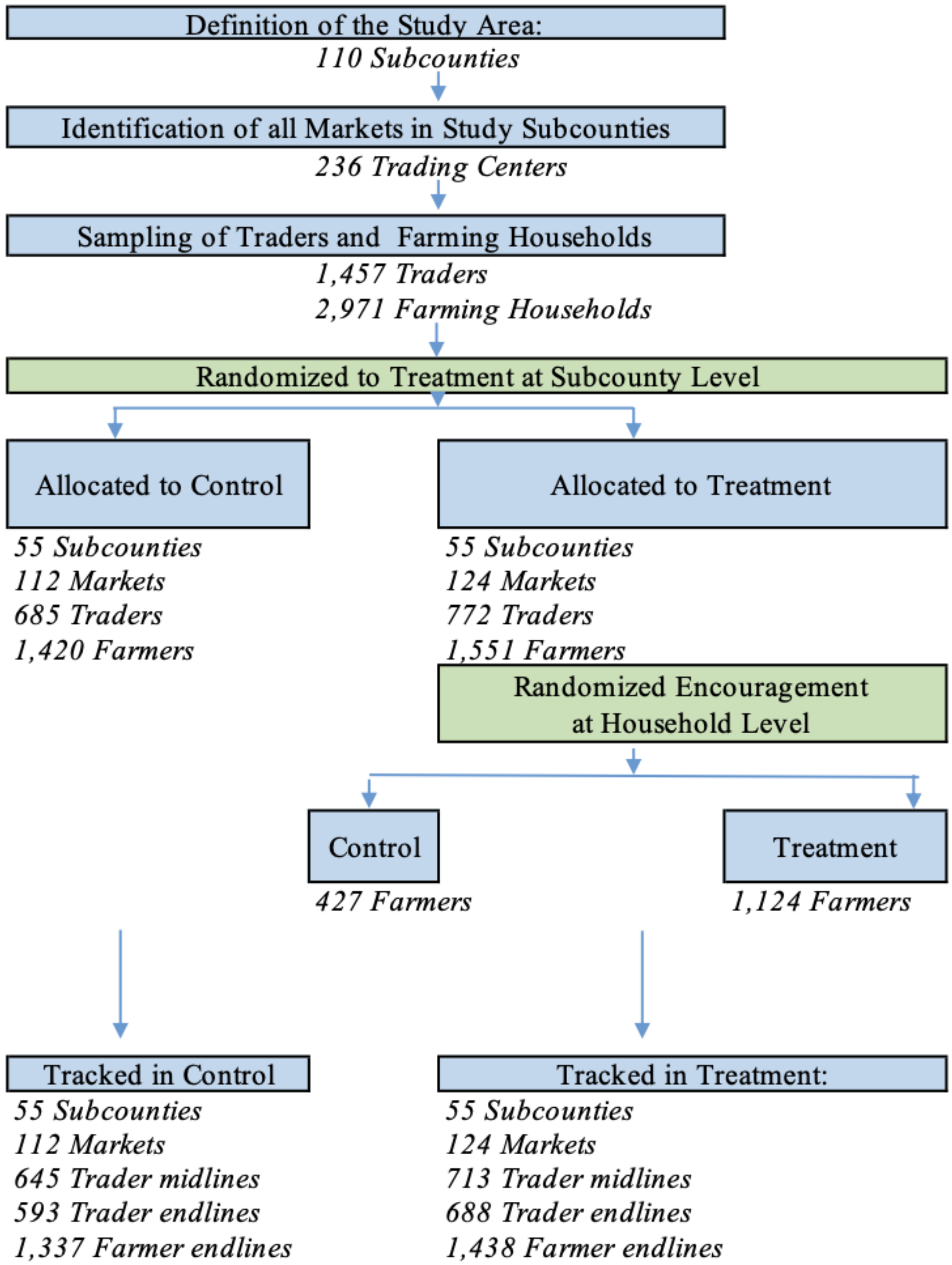


Figure A.4: **Attrition from the Market Survey.** The figure shows the Cumulative Density Functions (CDFs) of the fraction of intended surveys completed for each market, separately for treatment and control. A Kolmogorov-Smirnoff test fails to reject equality of the two distributions.

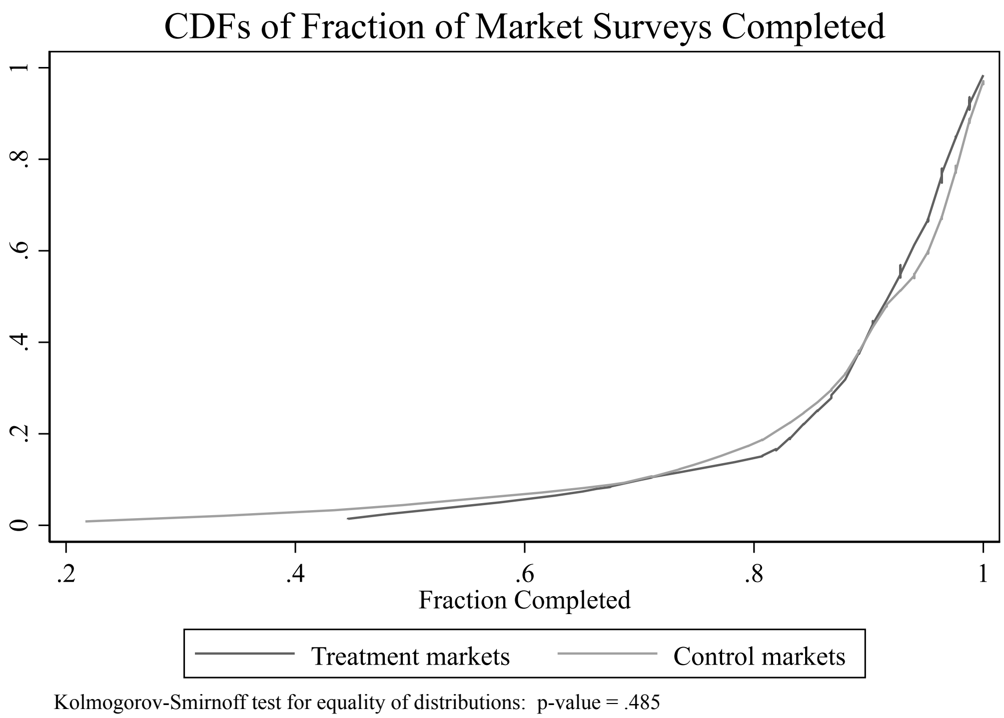


Figure A.5: **Volume of Bids and Asks.** Figure represents the spatial density of bids and asks over the study period.

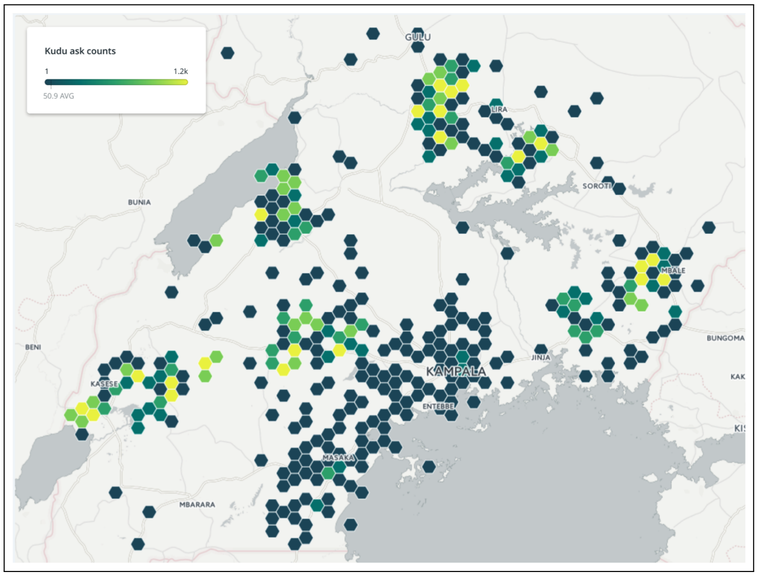


Figure A.6: **Maize Prices in Kudu vs. Market Survey.** Figure presents a comparison of average ask and bid prices on Kudu versus market prices.

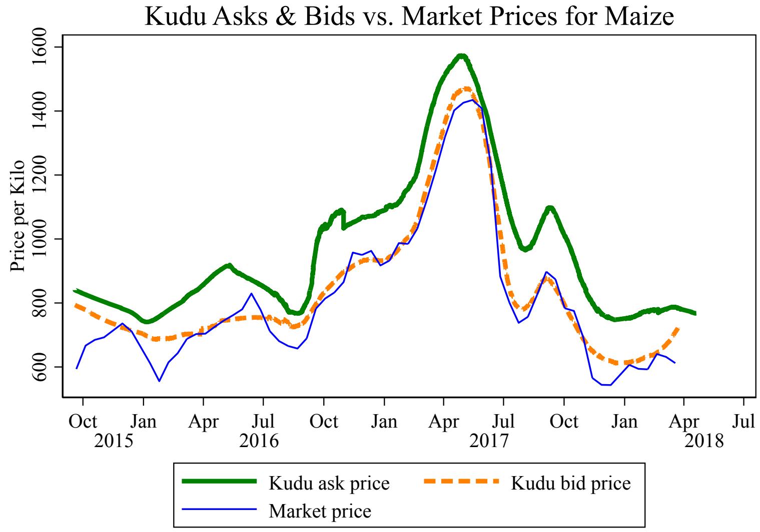


Figure A.7: **Distribution of Ask and Bid Prices, by Season.** Figure presents a box-and-whisker plot of the distribution of bid and ask prices posted on Kudu over the six seasons of the study.

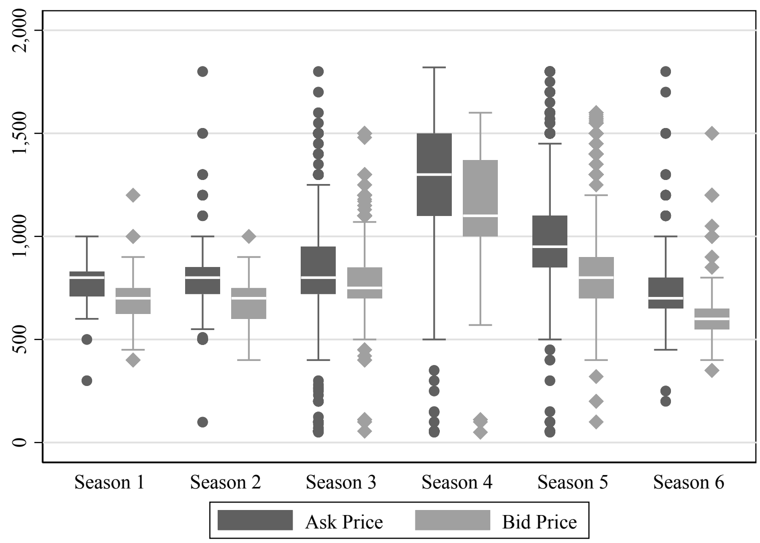


Figure A.8: **Cumulative transaction volumes on Kudu.** Cumulative transaction volumes of maize (in tons) on the Kudu platform over the study period.

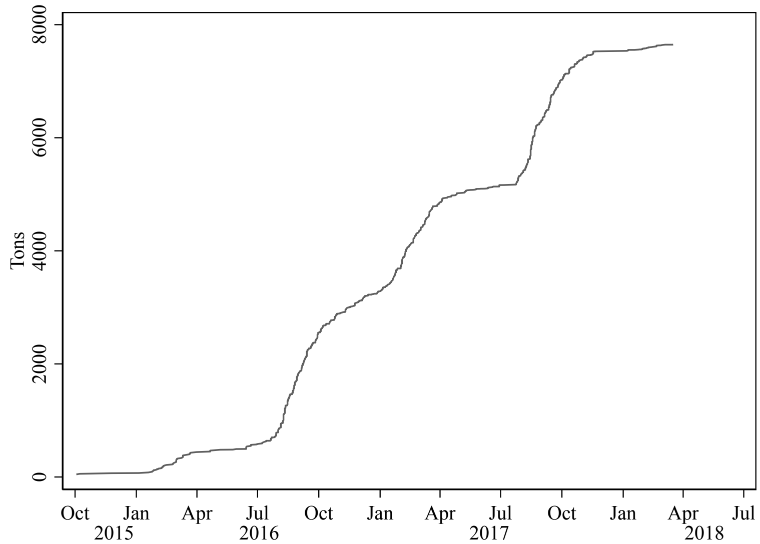


Figure A.9: **P-values from randomization inference for the reduced form effects on trade flows.** The figure shows the p-values resulting from a randomization inference procedure, in which 1,000 placebo randomized treatment assignments are draw. P-values from a two-sided test are presented in red and p-values from a one-sided test are presented in black. For the one-sided test, the null hypothesis being rejected is that the outcome is greater in control than in treatment. The left panel shows results for the probability of trade along any treated route, the middle panel for the number of traders trading along a route, and the right panel for trade volumes along the route.

Any Trade **Number Traders** **Trade Volumes**

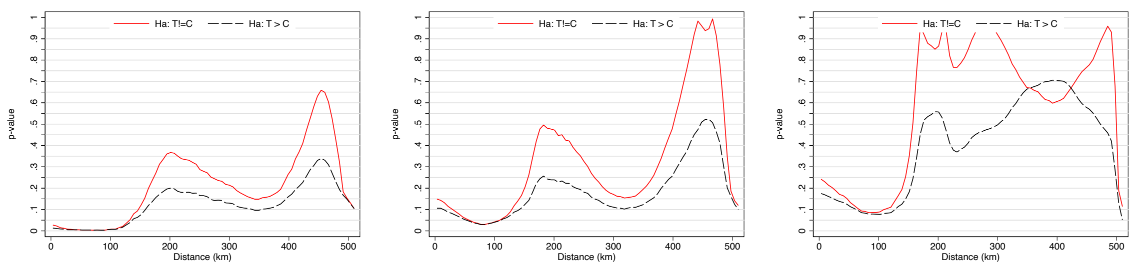


Figure A.10: **P-values from randomization inference for the reduced form effects on price gaps.** The figure shows the p-values resulting from a randomization inference procedure, in which 1,000 placebo randomized treatment assignments are draw. P-values from a two-sided test are presented in red and p-values from a one-sided test are presented in black. For the one-sided test, the null hypothesis being rejected is that the outcome is greater in treatment than in control.

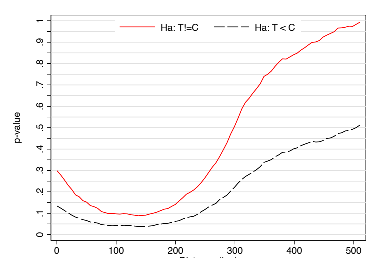
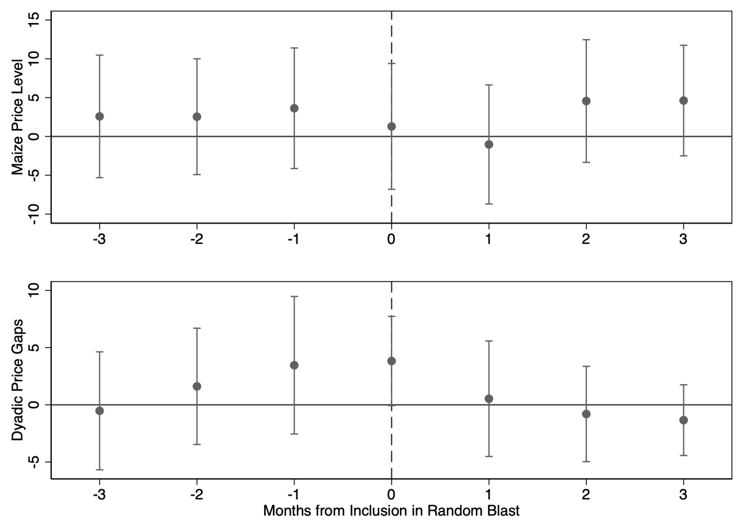


Figure A.11: **Impact of the Random Blast**. This figure presents an event study of the impact of the "Random Blast," a second sub-experiment to study the role of price information alone. In the "Random Blast," we randomly selected five treated markets in each round and broadcasted information about prices in those markets to the entire treated network. Though more short-run, this intervention was quite powerful, as it sent the market's price information to thousands of traders and farmers simultaneously. However, this price-only intervention also had no effect on price levels or gaps in the featured markets. The top panel presents monthly leads and lags of the treatment on price levels. The bottom panel examines impacts on price gaps between the Random Blast market and all other treatment markets that received information on that market's price.



Appendix B Tables

Table B.1: **Analysis of Variance in Market Prices.** Each coefficient in this table reports the R-squared from a different dummy variable fixed effects regression of prices in the market survey. The first column uses market fixed effects (and so measures cross-sectional variation in prices), the second column month-of-year fixed effects (and so measures the degree of typical seasonality), the third column includes fixed effects for each two-week round of the survey (and so measures the extent of overall time-series variation), and the fourth column includes both market and round fixed effects.

	Trading Center	Month of Year	Survey Round	TC and Round
Maize	0.04	0.18	0.84	0.87
Beans	0.26	0.15	0.34	0.55
Matooke	0.55	0.01	0.06	0.60
Tomato	0.30	0.04	0.09	0.39

Table B.2: **Trader Survey Attrition.** Analysis uses the full baseline sample of traders with a measure of subsequent attrition as the dependent variable. The first row examines midline attrition, the second row endline attrition prior to intensive tracking, the third row attrition during intensive tracking, and the final row overall endline attrition. Standard errors are clustered by subcounty.

	Treat	Control	Obs	T-C	
				diff	*p-val*
Baseline trader completes midline	0.92	0.94	1,457	-0.02	0.35
Tracked in original endline exercise	0.87	0.85	1,457	0.02	0.38
Found in Intensive Tracking	0.89	1.00	41	-0.11	0.08
Baseline trader completes endline	0.89	0.87	1,457	0.03	0.18

Table B.3: **Household Survey Attrition.** Analysis uses the full baseline sample of households with a measure of subsequent attrition as the dependent variable. The first row examines attrition prior to intensive tracking, the second row attrition during intensive tracking, and the final row overall endline attrition. Standard errors are clustered by subcounty.

	Treat	Control	Obs	T-C	
				diff	*p-val*
Tracked in standard exercise	0.92	0.93	2,971	-0.01	0.44
Tracked in intensive tracking	0.74	0.85	39	-0.11	0.35
Successfully tracked	0.93	0.94	2,971	-0.01	0.31

Table B.4: **Market Survey Balance.** Analysis uses the market-level average of outcomes from the two pre-treatment market survey waves to examine balance of the market survey. Standard errors are clustered by subcounty.

	Treat	Control	Obs	T-C	
				diff	*p-val*
Maize price	625.92	628.50	232	-2.58	0.86
Number of maize traders	8.82	8.84	233	-0.03	0.98
Maize quality	1.58	1.61	232	-0.04	0.57
Beans price	1,884.94	1,994.98	214	-110.04	0.16
Number of beans traders	5.14	5.01	233	0.13	0.88
Beans quality	1.45	1.50	214	-0.06	0.51
Bananas price	12,201.52	12,632.30	207	-430.79	0.67
Number of bananas traders	5.70	6.11	233	-0.41	0.74
Bananas quality	1.71	1.69	207	0.02	0.78
Tomato price	164.38	163.97	231	0.41	0.96
Number of tomato traders	9.87	10.53	233	-0.66	0.63
Tomato quality	1.64	1.63	231	0.01	0.82

Table B.5: **Dyadic market survey balance in absolute price gaps.** Analysis uses the dyad-level averages across the two pre-treatment rounds of the market survey to examine balance in absolute price gaps within dyad pairs. Standard errors are two-way clustered by each of the subcounties in the dyad, the unit of assignment.

	Maize	Beans	Bananas	Tomatoes
Both treated	10.56	22.11	116.2	1.916
	(7.715)	(31.59)	(354.4)	(2.753)
Mean DV	131.5	559.4	6361.7	72.37
N	26218	21129	20196	26149

Table B.6: **Trader Survey Balance.** Analysis conducted in baseline trader data but using only the attrited endline sample, with weights reflecting survey sampling and intensive tracking as used in the main analysis. The first two columns give the means in the control and treatment group respectively. The third column gives the total number of observations across the two groups. The last two columns give differences in means and the corresponding p-value. Standard errors for the p-values are clustered by subcounty, the unit of assignment.

	Treat	Control	Obs	T-C	
				diff	*p-val*
Female	0.07	0.06	1,281	0.01	0.64
Age	37.16	37.39	1,281	-0.23	0.76
Education	7.68	7.32	1,281	0.36	0.24
Age of business	10.86	10.92	1,178	-0.07	0.92
# of subcounties in which bought	1.15	1.12	1,281	0.03	0.46
# of subcounties in which sold	1.27	1.31	1,281	-0.03	0.65
Net revenue, mz & bn	21,946,001.68	28,474,012.42	1,275	-6,528,010.74	0.54
Business costs per month	6,290,868.45	6,050,540.21	1,281	240,328.24	0.80
Annual Revenue	47,550,250.45	45,657,411.81	1,278	1,892,838.64	0.82
Annual Costs	43,068,736.38	40,790,579.76	1,281	2,278,156.62	0.72
Volume buy (kgs), mz	112,323.01	100,580.90	1,281	11,742.10	0.63
Volume buy (kgs), bn	6,174.67	4,936.33	1,281	1,238.34	0.49
Volume sold (kgs), mz	157,676.55	161,821.69	1,281	-4,145.14	0.94
Volume sold (kgs), bn	6,667.08	5,906.31	1,281	760.77	0.71
Trade maize	0.92	0.94	1,281	-0.02	0.42
Trade beans	0.28	0.25	1,281	0.03	0.54
Annual profits	5,617,367.84	5,717,231.92	1,274	-99,864.08	0.92

Table B.7: **Household Survey Balance.** Analysis conducted in baseline household data but using the attrited endline sample, with weights reflecting survey sampling and intensive tracking as in the main analysis. The first two columns give the means in the control and treatment group respectively. The third column gives the total number of observations across the two groups. The last two columns give differences in means and the corresponding p-value. Standard errors for the p-values are clustered by subcounty, the unit of assignment.

	Treat	Control	Obs	T-C	
				diff	*p-val*
Number HH members	6.17	6.19	2,775	-0.03	0.86
Female	0.37	0.39	2,775	-0.01	0.66
Age	41.81	41.81	2,774	0.00	1.00
Highest grade completed	7.60	7.08	2,775	0.51	0.07
Food expenditure (month)	93,295.98	79,337.38	2,743	13,958.61	0.04
Land size (acre)	5.65	5.88	2,529	-0.23	0.60
Qtny sold, total (annual, kg)	1,133.87	1,056.52	2,775	77.35	0.70
Qtny harvest, total (annual, kg)	1,862.26	1,751.21	2,775	111.05	0.68
Number times sell	3.12	2.75	2,775	0.37	0.10
Percent of time sold at market	0.29	0.28	2,775	0.01	0.84
Sell in market	0.36	0.33	2,775	0.03	0.59
Distance to market	2.02	2.21	2,420	-0.20	0.63
Distance to Kampala	175.00	172.15	2,437	2.85	0.80
Assets (UGX)	2,508,859.59	2,297,790.91	2,775	211,068.68	0.62
Total exp (monthly, UGX)	219,099.68	191,827.79	2,775	27,271.89	0.09
Input exp (annual, UGX)	275,318.30	304,037.63	2,775	-28,719.33	0.46
Revenue, total (annual UGX)	637,169.99	555,801.61	2,775	81,368.38	0.43

Table B.8: **Predictors of take-up.** Columns 1-2 present the predictors of take-up among treated farmers, while 3-4 present those among treated traders, estimated via probit (with standard errors clustered at subcounty level). The outcome in Columns 1 and 3 is ever posting an ask or bid to Kudu, while the outcomes in Columns 2 and 4 are successfully completing a transaction on Kudu. Regressors are all measured at baseline and are: (log) quantity of maize sold or traded, age of respondent, years of education completed, and a dummy for being female.

	Farmers		Traders	
	Ever used Kudu	Completed deal on Kudu	Ever used Kudu	Completed deal on Kudu
Tons sold (log)	0.111***	0.019	0.128***	0.067**
	(0.032)	(0.066)	(0.032)	(0.034)
Age	-0.003	-0.001	-0.004	0.001
	(0.003)	(0.007)	(0.006)	(0.006)
Education	0.005	-0.009	0.003	-0.016
	(0.008)	(0.022)	(0.016)	(0.012)
Female	-0.314***	-0.272*	-0.009	-0.080
	(0.078)	(0.151)	(0.214)	(0.265)
Observations	1322	1322	716	716
Mean DV	0.26	0.02	0.80	0.22

Table B.9: **Impacts on farmer's production.** Analysis uses endline farmer surveys to measure impacts on production. Column one presents treatment effects on maize harvests (in kg), while column 2 presents an interaction of treatment and baseline surplus status of the area. Standard error are clustered by subcounty, the level of randomization.

	Harvests Maize KG	Harvests Maize KG
Treat	-15.54	-45.93
	(131.7)	(97.16)
Surplus		577.6**
		(250.7)
Treat*Surplus		108.9
		(329.6)
Observations	2768	2768
Mean of DV	1380.2	1380.2
R squared	0.343	0.357

Table B.10: **Impact of price information alone on price gaps.** This analysis uses dyadic market data in panel format with fixed effects for market survey wave, and analyzes the roll-in of the SMS price information-only intervention to control markets during the second half of the study. The sample for this analysis consists only of dyads in which neither market was treated in the original experiment, and 'Both Treated' means that both markets in the dyad were included in the roll-in treatment sample. Standard errors are two-way clustered at the subcounty level.

	All	Below 200km	Above 200km
Both treated by price info-only	4.17	6.89	2.88
	(8.31)	(10.14)	(7.32)
Dist (10km)	0.96^{***}	1.02	0.83^{**}
	(0.17)	(1.02)	(0.40)
Observations	220669	67899	152770
Mean DV	127.54	110.44	135.14

Table B.11: **Heterogeneity by unexpected price deviation.** This table examines whether dyadic connections through the platform induce faster convergence of prices when unexpected shocks hit. To calculate an unexpected dyadic price deviation, we first compute the average price for each district x month-of-year. Then, for each round of the market survey, we calculate the deviation in that round and district from the typical price at that location and time of year. Moving into the dyadic data structure, we then calculate the absolute value of the difference in these deviations across the dyadic pair. This then represents the component of the price gap between two districts that would not have been anticipated given the typical seasonal differences and hence provides a measure of the informational innovation present in the price discovery from the treatment. We then interact these price gaps with market-level treatment status to examine whether being treated causes faster price convergence for those treated pairs where the price information is revealing larger than expected price gaps. We look at this interaction starting with the contemporaneous shocks, and then lag the shocks by one, two, and three periods to allow for the fact that arbitrage will take time to lower these price gaps.

	Contemporaneous	First Lag	Second Lag	Third Lag
Both Treated x Shock	0.0113	-0.0123	0.0141	0.00593
	(0.0416)	(0.0376)	(0.0239)	(0.0175)
Dyadic shock dispersion	0.447***	0.162***	0.0257	0.0187
	(0.0351)	(0.0278)	(0.0168)	(0.0167)
Both treated	-5.051	-3.627	-4.698	-4.178
	(3.317)	(3.260)	(3.160)	(3.227)
Dist (10km)	0.478***	0.709***	0.823***	0.835***
	(0.0796)	(0.0793)	(0.0788)	(0.0798)
Mean DV	104.4	104.8	105.0	105.2
N	1440955	1416167	1390239	1366089

Table B.12: **Impacts on marketing.** Columns 1, 3, and 5 present treatment effects on the number of traders to which the farmer sold in the last 12 months, the number of *new* traders to which the farmer sold in the last 12 months, and whether the farmer sold anything at market in the last 12 months. Columns 2, 4, and 6 present an interaction of treatment and baseline surplus status of the area. Standard error are clustered by subcounty, the level of randomization.

	Number Traders		Number New Traders		Any Sales at Mkt	
Treat	-0.02	-0.02	0.02	0.05	0.00	0.01
	(0.08)	(0.09)	(0.03)	(0.03)	(0.03)	(0.03)
Surplus		0.39***		0.16***		-0.05
		(0.10)		(0.05)		(0.03)
Treat*Surplus		-0.01		-0.09		-0.04
		(0.13)		(0.09)		(0.05)
Observations	2768	2768	2768	2768	2774	2774
Mean of DV	1.05	1.05	0.22	0.22	0.26	0.26
R squared	0.05	0.07	0.01	0.01	0.04	0.05

Table B.13: **100% Scaled Treatment Effects.** Treatment on All (ToA) at 100% treatment saturation.

	Any Trade	Number Traders	Volume (tons)
Kudu	0.027***	0.014***	0.87***
	(0.0026)	(0.0012)	(0.12)
Dist (10km)	-0.01***	-0.01***	-2.48***
	(0.00)	(0.00)	(0.38)
Observations	11342	11342	11342
Mean DV	0.26	0.10	31.82

Appendix C Additional pre-committed analyses

For transparency, we describe here any additional analyses referred to in our pre-analysis plan, written in 2015, that we did not present above. First, we had intended to conduct an experiment to test credit constraints among traders by offering loans to a randomly selected subset of Commission Agents. We conducted a pilot for this experiment in the first season, issuing 62 short-term working capital loans to a group randomly selected from 124 CAs who expressed a desire for credit. In the end, the repayment rate on these loans was poor (78%) and our partner decided not to move this experiment to the intended scale, so we do not analyze it. Our PAP also specifies a set of hypotheses about convergence between spokes and hubs, and the differential effect of treatment for spokes in which the hub is and is not treated. In the end we were only able to map 84% of our spokes to hubs, and the analysis conducted within this reduced sample is typically inconclusive, suggesting that the trading networks may be more complex than our simple hub-and-spoke mapping at baseline supposed.

Appendix D Derivation of estimating equations

We specify variable trade costs as $\tau_{ij}^{\sigma-1} \equiv d_{ij}^{\gamma} e^{-u_{ij}}$, where d_{ij} is the distance between markets i and j. Fixed trade costs are: $F_{ij} \equiv \exp\left(\phi + \theta \mathbf{ln} d_{ij} + \beta_1 \mathbf{1}_{ij}^{K} + \beta_2 \mathbf{ln} d_{ij} \mathbf{1}_{ij}^{K} - v_{ij}\right)$ where $u_{ij} \sim N\left(0, \sigma_u^2\right)$ and $v_{ij} \sim N\left(0, \sigma_v^2\right)$, and $\mathbf{1}_{ij}^{K}$ is an indicator variable equal to one if both i and j are treated.

Using these assumptions about the form of trade costs, we can derive three estimating equations, as follows.

Any trade

Taking logs of Equation 2 and using the parameterization of τ_{ij} and F_{ij}, we get:

$$\zeta_0 + \zeta_j + \frac{(1-\sigma)}{\sigma_\eta^2} \ln p_i - (\gamma + \theta) \ln d_{ij} - \beta_1 \mathbf{1}_{ij}^{K} - \beta_2 \mathbf{ln} d_{ij} \mathbf{1}_{ij}^{K} + \eta_{ij} > 0 \tag{8}$$

where ζ_j is a destination market fixed effect, and $\eta_{ij} = v_{ij} + u_{ij}$ so that $\eta_{ij} \sim N\left(0, \sigma_\eta^2\right)$ where $\sigma_\eta^2 = \sigma_v^2 + \sigma_u^2$

Dividing through by σ_η^2 and making use of the normal distribution of η_{ij}, we get an equation for the probability of there being any trade from i to j:

$$\rho_{ij} = \Phi\left(\hat{\zeta}_0 + \hat{\zeta}_j + \frac{(1-\sigma)}{\sigma_\eta^2}\text{ln}p_i - \frac{(\gamma+\theta)}{\sigma_\eta^2}\text{ln}d_{ij} - \frac{\beta_1}{\sigma_\eta^2}\mathbf{1}_{ij}^K - \frac{\beta_2}{\sigma_\eta^2}\mathbf{ln}d_{ij}\mathbf{1}_{ij}^K\right) \tag{9}$$

where hatted variables are equal to their counterparts in 8, divided by σ_η^2, e.g. $\hat{\zeta}_0 = \frac{\zeta_0}{\sigma_\eta^2}$.

This yields our first moment condition:

$$\mathbf{Pr}\left(T_{ij}=1\right) - \Phi\left(\hat{\zeta}_0 + \hat{\zeta}_j + \frac{(1-\sigma)}{\sigma_\eta^2}\text{ln}p_i - \frac{(\gamma+\theta)}{\sigma_\eta^2}\text{ln}d_{ij} - \frac{\beta_1}{\sigma_\eta^2}\mathbf{1}_{ij}^K - \frac{\beta_2}{\sigma_\eta^2}\mathbf{ln}d_{ij}\mathbf{1}_{ij}^K\right) = 0$$

Number of traders

Estimation of a log-linear equation describing the number of traders serving a route will be biased by selection on the any trade margin – the set of routes with non-zero traders have unobserved features that are systematically different from those with zero traders. To avoid this problem, we rearrange Equation 4 to consider the fraction of traders based in market i that do not serve market j:

$$\text{ln}\left(1 - \tfrac{N_{ij}}{N_i}\right) = k\left(\text{ln}\tfrac{\sigma}{\sigma-1} + \text{ln}a_L - \tfrac{1}{1-\sigma}\text{ln}\sigma\right) + k\text{ln}p_i - k\text{ln}P_j + \tfrac{k}{1-\sigma}\text{ln}E_j + k\text{ln}\tau_{ij} - \tfrac{k}{1-\sigma}\text{ln}F_{ij}$$

Now using the parameterization of trade costs from the previous section, and collecting terms we get:

$$\text{ln}\left(1 - \frac{N_{ij}}{N_i}\right) = \varphi_0 + \varphi_j + k\text{ln}p_i - \frac{k(\gamma+\theta)}{1-\sigma}\text{ln}d_{ij} - \frac{k\beta_1}{1-\sigma}\mathbf{1}_{ij}^K - \frac{k\beta_2}{1-\sigma}\mathbf{ln}d_{ij}\mathbf{1}_{ij}^K + \varphi_{ij}$$

where $\varphi_{ij} = \frac{k}{1-\sigma}\left(v_{ij} + u_{ij}\right)$ is a mean zero normally distributed error term.

Value of trade on route

Both the any trade margin and the composition of heterogenous traders serving a route are important for determining the value of trade on a route, and we need to account for both factors when taking the model to the data.

Starting from Equation 4, taking logs, and making use of the parameterization of τ_{ij}, we get:

$$\text{ln}M_{ij} = (1-\sigma)\text{ln}\tfrac{\sigma}{\sigma-1} + (1-\sigma)\text{ln}p_i + \text{ln}N_i + (\sigma-1)\text{ln}P_j + \text{ln}E_j + (1-\sigma)\text{ln}\tau_{ij} + \text{ln}V_{ij}$$

where $V_{ij} = AW_{ij}$ and $W_{ij} = \max\left\{\left(\frac{a_{ij}^*}{a_L}\right)^{1-k-\sigma} - 1, 0\right\}$ and $A = \frac{k a_L^{1-\sigma}}{1-k-\sigma}$ is a constant.

Now using the parameterization of τ_{ij} from the previous section and collecting terms we get:

$$\text{ln}M_{ij} = \psi_0 + \psi_j + (1-\sigma)p_i - \gamma\text{ln}d_{ij} + u_{ij} + \text{ln}V_{ij} \tag{10}$$

Note that $\text{ln}M_{ij}$ is only defined when $T_{ij} = 1$. Conditioning on $T_{ij} = 1$ means that both $\text{ln}V_{ij}$ and u_{ij} are correlated with $\text{ln}d_{ij}$, and so if they are in the unobserved error term, estimates of γ will be biased. Therefore, we need estimates of these variables to include on the right side of the estimating equation. We can do this through a Heckman type correction for selection on both the any trade and trader composition margins:

$\ln\left\{exp\left[\delta\left(\hat{z}^*_{ij}+\hat{\bar{\eta}}^*_{ij}\right)\right]-1\right\}$ is a consistent estimate for $\mathbb{E}\left[\ln W_{ij}|.,T_{ij}=1\right]$, where where $\delta\equiv\frac{\sigma_\eta(k-\sigma+1)}{\sigma-1}$, $\hat{z}^*_{ij}=\Phi^{-1}\left(\hat{\rho}_{ij}\right)$ and $\hat{\rho}_{ij}$ is the predicted value from Equation 9, and δ is a coefficient to be estimated.

$B\frac{\phi\left(\hat{z}^*_{ij}\right)}{\Phi\left(\hat{z}^*_{ij}\right)}$ is a consistent estimate for $\mathbb{E}\left[\ln u_{ij}|.,T_{ij}=1\right]$, where $\hat{z}^*_{ij}$ is defined as above and $B=\frac{\mathrm{corr}(u_{ij},\eta_{ij})}{\left(\frac{\sigma_u}{\sigma_\eta}\right)}$ is a coefficient to be estimated.