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

Shilpa Aggarwal† Brian Giera‡ Dahyeon Jeong§ Jonathan Robinson¶
Alan Spearot‖

May 14, 2018

**Abstract**

While it is widely believed that poor market access limits agricultural productivity in rural Africa, quantifying the magnitude of this effect is complicated by a lack of data. To fill this gap, we collect granular data on farmer input and sales decisions, input and output prices, and travel costs in all 570 villages of the Kilimanjaro region of Northern Tanzania. We document substantial spatial price dispersion: going from the 10th to 90th percentile of the travel cost-adjusted price distribution increases prices by 40-50% of the mean, for both fertilizer and maize. We find that an additional log point of remoteness (measured as travel time from the regional hub) is associated with 50% lower market participation and input adoption. We develop a spatial model of input adoption and estimate that farmers behave as if they face travel costs of 3-4% ad-valorem per kilometer of travel (equivalent to a total of approximately 30% for the average purchase), which is approximately 3 times measured pecuniary travel costs. Holding exogenous local factors fixed, we estimate that reducing travel costs by 50% (approximately the effect of paving rural roads) would reduce the adoption-remoteness gradient by 45%.

*JEL Codes:* F14, O12, O13, O18, Q12

*Keywords:* market access, inputs, technology adoption, transport costs, roads

*We are grateful to Sheah Deilami-Nugent and David Roughgarden for outstanding work coordinating field activities, and BRAC Tanzania and IPA Tanzania for administrative support. We thank Patrick Olobo for his collaboration. Anouk Foure, Petra Gram, Matthew Krupoff and Alexandra Murphy provided excellent research assistance. We thank Lauren Falcao Bergquist, Alain de Janvry, Betty Sadoulet and seminar participants at Ashoka University, UC Berkeley, ISB, LSE, the University of Maryland, NEUDC, the University of Virginia, UCSC, the University of West Virginia and the WEAI for helpful comments. We acknowledge funding from the IGC (grant 1-VRC-VTZA-VXXXX-89307), the National Science Foundation (SES 1658942) and Innovations for Poverty Action. The data collection approval was approved by the IRBs of UCSC and the Indian School of Business, and by the Tanzania Commission for Science and Technology (COSTECH).

†Indian School of Business, email: shilpa_aggarwal@isb.edu
‡Amazon Research, email: bgiera14@gmail.com
§University of California, Santa Cruz, email: dajeong@ucsc.edu
¶University of California, Santa Cruz and NBER, email: jmrtwo@ucsc.edu
‖University of California, Santa Cruz, email: acspearot@gmail.com

# 1 Introduction

It is widely believed that poor access to markets – due mainly to poor transportation infrastructure – limits agricultural productivity in rural areas of developing countries, by making it harder to access productivity-enhancing inputs like fertilizer and to obtain high prices for harvest output (World Bank, 2008; 2017).[1] However, while remoteness no doubt limits market access, there is little research to rigorously quantify its effect, especially in the case of input adoption.

In this paper, we rigorously document market access among farmers in the Kilimanjaro region of Northern Tanzania. Our data collection exercise spans the entire supply chain in all 570 villages in the Kilimanjaro region, including (1) surveys with a sample of nearly 2,000 farmers in more than 180 randomly selected villages; (2) a census of all 395 agro-retailers, or "agrovets," in the region, and follow-up surveys with 369 agro-retailers that sell fertilizer; (3) a retrospective panel of monthly buying and selling prices of maize from a randomly selected sample of maize-sellers in each of the 98 markets in the region; (4) collection of information on road quality, travel times, and travel costs to all villages from their respective local markets as well as from 3 major urban centers, and travel times and costs between each market and each major urban center; and (5) driving times and distances pulled from Google Maps API for the universe of bilateral village pairs, as well as for pairs of villages and major urban centers in the region. This region includes a great deal of heterogeneity in remoteness – varying from villages just a few kilometers from the major town where fertilizer is distributed to villages located in remote mountains 200 km away – and so provides enough variation to examine fairly substantial changes in travel time to the urban hub.

We use our data to first document substantial spatial heterogeneity in input and output prices, for the universe of villages in the region: the price difference between the 90th and the 10th percentile of delivered input prices is equivalent to about 50% of the mean. This heterogeneity has important implications for the profitability of inputs, and hence on input adoption. We then conduct a reduced-form investigation of the correlation between input usage, market access and remoteness, which we define as travel time from the regional hub, Moshi. We find that input adoption and maize sales are much lower in remote areas: a 1 log point increase in travel time from Moshi is associated with a 60% reduction in the quantity used of fertilizer and a 40% reduction in maize sales. These results attenuate to about half this size when controlling for soil and farmer characteristics, but continue to be economically and statistically meaningful. Conditional on purchasing fertilizer, we observe that farmers 1 log point further away from the hub travel 50% further to buy the inputs. We also find that fertilizer is more expensive in rural areas: the marginal cost of transporting fertilizer to the shop increases by 50% with 1 log point distance, and markups increase by 10%.

On the output side, the story is similar: to measure general selling conditions, we construct the full set of travel cost-adjusted selling price for maize sales, which is the price available at every

[1] Transportation infrastructure is particularly underdeveloped in Africa as the continent has only 137 kilometers of roads per 1000 square kilometers of land area, with only a quarter paved. In contrast, the average for developing countries outside the region is 211 kilometers of roads per 1000 square kilometers, with more than half paved (World Bank, 2010). For comparison, the US has 679 kilometers per 1000 square kilometers, with nearly 2/3 paved.

weekly market in the region minus the cost of getting there and back from each village. Then, for each village, we find the maximum of this adjusted price. The standard deviation in the best travel cost-adjusted selling price of maize is also 15% of the mean, and the 90-10 ratio is similar (about 40%). Farmers are also much less likely to be in contact with a maize-buying intermediary ("agent") in remote areas, with a log point increase in travel time to Moshi being associated with a nearly 60% reduction in the likelihood of an agent visiting the farmer.[2]

The reduced form results suggest that more remote farmers suffer from reduced access to input retailers as well as selling opportunities. Examined on their own, this suggests a potentially large elasticity of adoption to the costs of reaching input and output markets. However, as other factors are also correlated with distance, we rely on a model to quantify the impact of access to input and output markets on input decisions. We develop a spatial model of fertilizer adoption in which the decision to adopt fertilizer is based on local output prices (including costs to reach the market), innate farmer productivity, the distribution of delivered input prices and retailer quality, and idiosyncratic shocks. Transportation costs affect the distribution of prices by increasing the cost for the farmer to reach a particular agrovet, and by increasing the costs of farmers to reach the local market to sell their harvest.

On the input side, the structure of the model (which is based on Eaton and Kortum, 2002) facilitates a decomposition of choosing an agrovet into (1) the decision whether to adopt anywhere; (2), the decision to travel to a particular location to acquire fertilizer; and (3) the decision to choose a particular agrovet within a location. Our novel farmer surveys reliably record the first two of these probabilities, and respectively, they allow us to calibrate local factors unrelated to prices that affect fertilizer adoption, and the implied trade costs of choosing between agrovets. On the latter, we derive a novel multinomial logit specification that estimates the implied iceberg trade costs to each location as a function of distance, while using location-specific dummies to account for other amenities available at each location. The results suggest that transportation costs are large; a simple iceberg cost that is a function of kilometers to each agrovet is estimated at approximately 3-4% per KM of travel. This is approximately 3 times the measured pecuniary costs of travel, and therefore suggest large non-pecuniary transport costs. These may include other factors such as the opportunity cost of the time to travel, risk aversion related to the risk of stockouts, or information frictions.

After estimating trade costs, we use the model to build a market-clearing condition for fertilizer sold by each agrovet, which is a function of the spatial distribution of fertilizer expenditures by each farmer and the probability that a farmer at each location adopts at a given agrovet. We balance these market clearing conditions by finding a vector of agrovet "non-price attributes" that exactly rationalize the market-shares of each agrovet. The recovered vector of non-price attributes allows us to assess the external validity of the model in two ways. First, these non-price attributes themselves are positively correlated with agrovet experience at that location. Second, using similar

[2]Maize-buying intermediaries - "agents" - typically visit villages soon after harvest and buy maize in bulk to profit from either inter-temporal or spatial arbitrage opportunities.

techniques to Berry (1994), we find that the mean implied mark-up in the data is 21%, which is similar in magnitude to those summarized in the reduced form.

To study the role of input and output market access on the adoption decision, and more generally, the remoteness gradient documented in the reduced form results, we run counterfactuals that adjust access to retailers and output markets. For input market access, our primary counterfactual is reducing trade costs incurred to reach retailers by 50%, which is roughly equivalent to a reduction in travel times due to road upgrading (Casaburi, Glennerster and Suri 2013). This policy increases mean adoption by 25 pecentage points (59% of the baseline mean), and also reduces the distance gradient by 45%. We also evaluate how the costs for retailers to source inputs from distributors affect the adoption decision. Fully subsidizing reported transport costs of retailers buying fertilizer from distributors, adoption rises by about 2pp, or 4.4%, but yields a 9.6% reduction in the remoteness-adoption gradient. On the output side, we run a similar counterfactual to the reduction in trade costs to obtain inputs, by reducing the trade costs to sell output at the local market by 50%. Here, the adoption rises by 6.7pp, or 15 percent, and there is effectively no change in the remoteness-adoption gradient. Finally, we offer farmers the market price in Moshi to simulate adoption effects when farmers have ready access to a major "urban" market. Here, we find large adoption effects (18pp, or 42%) and an adoption-remoteness gradient that is cut by 10.2%.

This paper sits at the intersection of trade and development economics, and we provide value to both literatures. Our primary question considers the impact of remoteness on the price, availability, and adoption of fertilizer by rural farmers. Sub-Saharan Africa has lagged far behind the rest of the developing world in agricultural technology adoption (World Bank 2007) despite evidence that improved technologies could generate large *yield* increases (i.e. Duflo, Kremer and Robinson 2008; Beaman et al. 2013; Stewart et al. 2005; Udry and Anagol 2006). The *profitability* of these technologies thus depends on the relative prices of fertilizer and crop output, and on the size of the yield increase. The literature is more divided on whether these technologies are profitable, with some papers finding large returns (i.e. Duflo, Kremer and Robinson 2008) and others lower or even negative returns (i.e. Beaman et al. 2013). While this previous literature has mostly focused on measuring yield increases, profitability is equally affected by access to technology and sales opportunities, the focus of this paper. Our results quantify the extent to which profitability, and thus adoption, will tend to be lower in more remote locations.

Our paper is also differentiated from much of the development literature by focusing on market access, rather than on demand side explanations like farmer knowledge and learning spillovers (Foster and Rosenzweig, 1995; Conley and Udry 2010; Bandiera and Rasul, 2006; Emerick, 2017), credit, liquidity or insurance constraints (Bardhan and Mookherjee, 2011; Maitra et al., 2017; Karlan et al., 2015), or behavioral explanations (Duflo, Kremer and Robinson, 2011; Hanna, Mullainathan, and Schwartzstein, 2014).[3] Our work is most closely related to Suri (2011), who shows that many Kenyan farmers with high gross returns to hybrid seeds choose not to adopt them because the fixed costs of obtaining seeds are too high, presumably due to travel costs. Our paper is differentiated

[3]See Foster and Rosenzweig (2010) and Jack (2013) for reviews of this literature.

by focusing on heterogeneity in market access, rather than on heterogeneity in returns.

Our paper is related to a rapidly growing literature about the effect of roads or other infrastructure improvements on development outcomes and on the spatial distribution of economic activity.[4] Many of these papers evaluate large-scale infrastructure programs as natural experiments, or by employing structural techniques, and thus provide causal evidence on the effect of *roads* on various outcomes. The key difference in our paper is that we focus narrowly on the specific effect of transportation costs on market access (i.e. the actual time and money costs of transportation and the presence of intermediaries and the prices they charge) in isolation, without changing other margins.[5] Building roads may change many outcomes other than just prices, including consumption diversity, human capital investment, migration, and occupation choice (Aggarwal, 2018; Adukia et al., 2016; Morten and Oliveira 2016; Asher and Novosad 2016), as well as many others such as electrification, proximity to health care, etc.[6] By contrast, our goal is to focus solely on the effect of remoteness on intermediary entry and pricing, with special emphasis on chemical fertilizer.[7]

Our work is related to a voluminous trade literature. Within this literature, price differentials across space can be attributed to three primary components – marginal trade costs (e.g. Donaldson, 2018; Eaton and Kortum, 2002; Keller and Shiue, 2007; Sotelo, 2016), spatially varying mark-ups (Atkin and Donaldson, 2015; Asturias et al., 2017), and the organization of intermediaries (Allen and Atkin, 2016; Dhingra and Tenreyro, 2017; Bergquist, 2017; Casaburi and Reed, 2017; Chatterjee, 2018). Simply quantifying these price differences is important for the literature, as there is a dearth of data studying rural markets, and in particular, access to inputs. We collected price and sales information by firm, input-type and brand - essentially "scanner" data - including wholesale prices for these items, which facilitates an exact measure of retail mark-ups. Further, our unique transportation surveys allow us to calculate the exact cost of acquiring inputs for all possible locations in our sample, providing a comprehensive mapping of input and output market access within the region.

Our paper is closely related to Atkin and Donaldson (2015), who estimate trade costs in a situation in which an oligopolist intermediary buys products at wholesale prices, transports them to distant markets and sells them directly to consumers. By contrast, we are interested in how trade costs affect the buying decisions of final consumers (in this case, farmers), as well as pricing decisions by retailers. Though not directly comparable since they are at different points in the supply

[4] A partial listing of papers includes Aggarwal (2018), Alder (2017), Adukia et al. (2016), Asher and Novosad (2016), Banerjee et al. (2012), Bird and Straub (2016), Bryan and Morten (2017), Gertler et al. (2014), Ghani et al. (2016), Khanna (2016), Shamdasani (2016), and Storeygard (2016). See Donaldson (2016) for a review.

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

[6] Indeed, several papers in this literature use overall economic development (as proxied by night lights) to capture the all-pervasive nature of the impacts generated by road construction. See, for example, Alder (2017), Khanna (2016), and Storeygard (2016).

[7] In the specific context of agricultural inputs, both Aggarwal (2018) and Shamdasani (2016) find evidence of increased input adoption in the wake of a pan-Indian rural road construction program. However, the impact documented by both of these papers is reduced form in nature and neither is able to establish either the impact of transport costs on decreasing adoption or the channels through which road construction encourages adoption.

chain, our average ad-valorem "travel costs" of farmers procuring fertilizer, as calculated through interviews with transport providers, turn out to be similar to those of the intermediaries in Atkin and Donaldson (2015). Our costs, however, are calculated over a much shorter trip.[8] When using our structural multinomial choice model to estimate the implied trade costs from revealed retailer choices, the results suggest even higher numbers, which are suggestive of other non-pecuniary costs of input purchases (search, hassle costs, etc). Accordingly, using our quantitative model, we find that farmers are particularly sensitive to the total costs of reaching retailers. On the output side of the market, we collect novel descriptive measures of intermediary behavior, in particular, the entry of output buying "agents." Allen and Atkin (2016) models a similar channel, where a perfectly competitive, heterogeneous group of traders travel from market to market exploiting all available arbitrage opportunities.[9] Different from the data used in their work, we measure intermediation directly at the level of the farmer - whether crops are sold, and if so, in what quantity and at what price. Indeed we have found an active supply network for maize that is run by intermediaries, we find that many farmers are not served by them, and that distance to the nearest market and nearest town significantly reduces the probability of being served. We also measure the proximity of farmers to their output market, and measure the effects cutting trade costs to reach this market. These effects are are positive and meaningful, but mostly uncorrelated with farmer remoteness.

Finally, much of the trade literature, which has documented larger gains from integration when there are input-output relationships (e.g. Yi, 2001; Costinot and Rodriguez-Clare, 2014; Sotelo, 2016) has only evaluated economies under the assumption of monopolistically competitive or purely competitive sectors at a fairly aggregate level (e.g. international trade by industry).[10] By contrast, our model yields a structural discrete choice problem in which farmers choose whether to adopt, and if so, choose the best agrovet from which to purchase fertilizer. [11]

The rest of this paper proceeds as follows. Section 2 provides background and context on our study region, and lays out the sampling strategy that was adopted for this project. Section 3 explains the data, and documents summary statistics about the various data-collection units. Section 4 presents our main results. We put our findings in the context of a spatial model, which

[8]Specifically, to find the best travel-adjusted price for fertilizer, our results suggest that for the typical village, the best option is 10km away. In Atkin and Donaldson, ad-valorem estimates are calculated based on the cost difference of a trip to the most remote location (500 miles away) relative to the least remote location (50 miles away), which is approximately a 720km difference.

[9]In Allen and Atkin (2016), when a particular market has excess supply, less efficient intermediaries enter that "route" to exploit the new arbitrage opportunity. In their work, they use this model to quantify the role of revenue volatility in crop choice, and use a highway project in India to evaluate how crop choice affects the gains from integration.

[10]Our work is closely related to Sotelo (2016) develops a model of regional trade in agriculture and road quality in Peru to study the impact of road and output shocks on regional welfare and crop choice. Our work differs in its focus on local intermediaries and how their presence affects the landscape of market access.

[11]Yi (2001) provides an influential take on the role of vertical relationships in the growth of vertical trade that is germane to our work. Intuitively, if inputs are traded from one country to another, and then final goods are traded back to the origin country, the role of distance is amplified by the multiple stages of production. That is, since borders must be crossed more than once, the costs of distance are amplified by the number of times the good crosses the border prior to consumption. Our field work has identified that economy in rural Tanzania is similar to this setting, where inputs are sourced from larger cities, and output, if sold at all, is trade back to these same cities.

is presented and calibrated in Section 5, and run policy counterfactuals in Section 6. Section 7 discusses the validity of our results outside of the study context of Northern Tanzania. Section 8 concludes with a discussion.

# 2 Background and Sampling Strategy

## 2.1 Background on fertilizer market and Kilimanjaro region

This primary study took place in the Kilimanjaro region[12] of Northern Tanzania. There are 570 villages in the region, and according to the 2012 census of Tanzania, the total population of the area is 1.6 million, about three-quarters of which is rural (National Bureau of Statistics, 2013). Our data collection covers the entire area of the region, a substantial area of 13,250 square-kilometers, roughly equivalent to the state of Connecticut in the USA or the country of Montenegro. Within Tanzania, Kilimanjaro is a relatively prosperous region, and agricultural productivity is relatively high. Using data provided by the 2012-13 wave of the National Panel Survey, we find that farmers in Kilimanjaro reported maize yields that are about 30 percent higher than the national average. Roads within Kilimanjaro are also marginally better than in Tanzania on the whole - according to numbers reported by the government of Tanzania, the paved trunk road density in Kilimanjaro is 2.2 percent of the total land area in the region (i.e. there are 2.2 kilometers of roads per 100 square kilometers of area), as opposed to only 0.7 percent for Tanzania as a whole. The density of the total network of trunk and regional roads is 7.4 percent in Kilimanjaro, but only half as much (3.7 percent) for the entire country of Tanzania (TanRoads and PMO-RALG, 2014).[13] The relative density of other minor roads is likely similar, although these numbers are harder to obtain. From an objective standpoint however, the road network in Kilimanjaro is quite poor. For instance, the density of the road network in the United States is 68 percent; the OECD average is 134 percent.[14]

Kilimanjaro has two growing seasons: a longer, more productive "long rains" season, which runs from March to June, and a less productive "short rains" season from October to January. Input usage tends to be much higher in the long rains, and some farmers decide not to plant in the short rains at all. Our main outcomes are based on behavior in the long rains.

We worked off of the list of villages included in the documents pertaining to the 2012 census of Tanzania, and did data-collection in the universe of villages (570 villages) listed as being in the Kilimanjaro region. As discussed in more detail below, we conducted farmer surveys in a subset of villages, and did a comprehensive census of agro-input retailers in the entire region.

Virtually all fertilizer is imported in Tanzania. While some developing countries (such as India) produce chemical fertilizer domestically, production capacity is virtually non-existent in sub-

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

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

[14] Information compiled from various online resources.

Saharan Africa, and therefore, many sub-Saharan African nations import the entirety of their fertilizer requirements (FAOSTAT Online database, 2016; Hernandez and Torero, 2011).[15] As a result, for these countries, transport costs from port to farm will directly affect prices. At present, we do not document the costs from port to distributor, but collect costs from that point on. In particular, we focus on the rural costs of intermediation and the costs at which farmers acquire inputs from retailers. To our knowledge, documenting the latter costs is entirely novel within the literature.

### 2.2 Sampling Strategy

The goal of this project is to construct a dataset representative of the entire region of Kilimanjaro. The main categories of data we set out to measure were: (1) surveys of farmers, fertilizer retailers, and maize buying agents; (2) transportation costs; and (3) prices. We initially set out to measure prices of a variety of goods. However, many villagers do not purchase most of their goods in their local village, and instead travel to local markets which operate one or several days a week. We decided to use these markets as the unit at which we would measure prices.

Thus, to construct our sample, we first assigned every village in our sample to a market catchment area. This was done by visiting ward offices (the ward is the lowest administrative level in Tanzania) and asking the ward officer to list the market that people from that village frequented. We use this market information in two main ways. First, we randomly selected markets for inclusion in the price collection from this list. Second, it was not feasible to travel individually from every village to a particular point to measure transport costs. Instead, we measure transport costs, requiring routes to go through the market center – we measure distances from every village to its closest market, and from every market to the main road. A map of the villages in our sample is included as Figures 1.

The geography of Kilimanjaro region provides for a setting with potential wide variation in transportation costs to Moshi. Closest to Moshi are semi-urban and rural districts that surround the city. While many villages may be located off main roads, their location is proximate to the main supply points in the region. In the northeastern part of Kilimanjaro, near the border with Kenya, many villages are by straight-line distance not far from Moshi, but the presence of Mt. Kilimanjaro is complicating for travel. Further removed from Moshi are villages near the town of Same, located to the southeast of Moshi and itself connected to Moshi by the main trunk road within the region. However, even along this road, travel times are not trivial, and many villages are located within, or on the other side of, the Pare Mountains (to the northeast of Same). Overall, the region provides substantial geographic variation that we now document in terms of the costs of travel.

[15] Tanzania has some limited domestic production capacity in the form of an Arusha-based company called Minjingu Mines and Fertilizer Ltd. Only a handful of retailers in our sample sell this brand of fertilizer, however.

# 3 Data and summary statistics

We have four main sources of data we use in this draft: agrovet surveys, farmer surveys, transport surveys, and maize price surveys. All were collected from January 2016 to December 2017.[16]

## 3.1 Agrovet surveys

First, we conducted a census of all agrovets in the region, finding a total of 395 agrovets. Of these agrovets, 376 sell fertilizer, and constituted our primary sample to conduct the detailed agrovet survey. We then revisited these agrovets to conduct a longer survey which took about 2 hours to complete. In total, 369 agrovets completed the survey. The survey asked questions about varieties of fertilizer sold, prices, quantities, and the wholesale costs of acquiring stock from the distributor. The survey took care to differentiate fertilizer types by distributor, brand, and type – thus the level of granularity should be akin to the barcode-level. The survey also included a number of questions about costs of travel to the distributor, as well as some background characteristics.

## 3.2 Farmer surveys

We conducted farmer surveys in a randomly sampled subset of about 180 villages. The farmer survey was collected in two waves, the first one in early 2016, and the second in late 2017. To the extent possible, we tried to survey mutually exclusive sets of villages during the 2 waves, and so our full sample for the farmer survey is akin to a repeated cross-section. During the first wave, we worked in 115 villages. Within a selected village, enumerators were instructed to first find a landmark.[17] Once the village center was identified, the enumerators randomly picked a direction to begin their fieldwork, and selected every third homestead, or the next homestead after five minutes of walking, whichever came first. Overall, we enrolled an average of 4.8 farmers in 115 villages. The survey itself included questions on input usage and prices, maize sales, harvest output, and related outcomes. The survey also included some household and demographic questions.

During Wave 2 of the farmer survey, we randomly sampled 97 villages, and data-collection began with one of our field officers visiting the local village office to procure a listing of all households in the village. From this list, we randomly selected 18 households for surveying (with an additional backup list of about 20), and conducted surveys with them. The nature of questions asked was largely the same as in the first round of data-collection.

[16] We also collected data on maize intermediaries who buy directly from farmers ("agents") and on larger warehouses that buy from these agents ("stores"), as well as logbooks of transactions from stores. We do not utilize this data in this version of the paper.

[17] These landmarks included a primary/secondary school within the village (1st choice), local church within the village (2nd), Boda stand within the village (3rd).

### 3.3 Measuring transport costs

We measured transportation costs in several ways. First, we collected GPS location for every village in Kilimanjaro,[18] from which we calculated driving times and distances using the Google API (via the statistical program R). Second, we conducted surveys of transportation operators in every village in our sample, which were either motorbike taxis ("Boda Bodas"), or consumer van taxis ("Dala Dalas"). In each village, we asked up to 3 operators how much it cost to travel to the major towns in Kilimanjaro (Arusha and Moshi), the capital city (Dar-es-Salaam) and the market center.

Third, enumerators recorded information on road quality and travel times as part of their field work. There are several major paved roads in Kilimanjaro. While not up to developed country standards, these roads are better maintained and most are paved. They are typically 2 lane roads. To get to a village, it is typically necessary to turn off one of these main roads and then travel for some time on unpaved feeder roads and village roads. To measure travel times, field officers used the following protocol. On a GPS unit, they recorded the point at which they had to turn off the main road, and then recorded the travel time, distance, and road quality on the road to the market center associated with the village. Once reaching the market, enumerators took a second form of transportation to the village, recording again distance, travel time, and road quality. We use this data to correlate costs of travel with road quality, and to estimate the percentage of roads which are paved (to inform later counterfactuals).

### 3.4 Maize prices

To measure maize prices, we visited markets post-harvest in September and October of 2017. During the visits, enumerators sampled up to 3 maize sellers per market to document pre- and post-harvest prices for maize during recent seasons. These data allow us to compare prices across markets at the same point in time, though they are not intended to be used in panel analysis.

### 3.5 Summary statistics

Summary statistics are provided in Table 1 for villages (Panel A) and roads (Panel B). The average village has 2,842 people, and is located 5.7 kilometers from the nearest market center. A round-trip to the market center takes about 40 minutes according to surveys (20 minutes according to Google maps), and costs about US $1.60. The average distance to the nearest major town of Moshi is about 66 km, and a round-trip there would take just under 3 hours and cost about $4.70. These travel costs are substantial for poor farmers making a few dollars a day.

There is also substantial variation in travel costs to these cities in the region, from towns just outside Moshi to remote villages in the mountains in Same District in the South of Kilimanjaro. The standard deviation of travel costs to Moshi is about 80% of the mean, while the minimum

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

travel cost is about \$0.30 and the maximum is \$22. In this context, it is reasonable to consider counterfactuals of even very large increases in travel costs.

Panel B shows information on the quality of the rural roads connecting markets and villages. Roads are about 1/3 paved, 1/3 dirt, and 1/3 gravel, and travel times according to google are fairly slow: 30.6 km/hour on rural roads compared to 49.5 km/hr on the main roads.[19]

Table 3 present summary statistics on our sample of farmers, and shows how these characteristics vary with distance from the regional hub. The average farmer in our sample is 51 years old and has about 7 years of education. Forty-six percent of our sample is female, and 73% is married. While poor, farmers in our sample are better off than in many parts of Africa: 90% of farmers have cell phones, 82% have mobile money accounts, and only 11% live in homes with thatch roofs. The average farmer has almost 3 acres of land, 28% have a market business, and the average farmer's income from activities other than farming is approximately \$430 per year. We find a correlation with remoteness for only a handful of these measures. Farmers in more remote areas are somewhat younger, are less likely to have cell phones, bank accounts, and mobile money accounts, and are less likely to have market businesses. On the other hand, they have more land and are less likely to have a thatch roof. On the whole however, the gradient in distance for these variables is not particularly steep.

Panel B shows buying and selling behavior. We find that 36% of farmers buy maize but sell none, 21% sell maize but buy none, and 36% do not buy or sell maize. Only 7% buy and sell maize. Forty percent of farmers buy more maize than they sell, compared to 22% who sell more than they buy. Panel C shows estimated yields from the FAO-GAEZ database. Yields are 0.8 metric tons per acre without inputs, compared to 3.4 metric tons per acre with inputs. We find that yields without inputs increase by 0.15 tons with every log point from Moshi, and yields with inputs decline by 0.225 tons, so that the difference is 0.37 tons lower. All else equal, this suggests we would expect lower input usage in rural areas. However, note that the size of the coefficient is modest, equivalent to about 14% of the constant, and so this difference would likely account for a relatively small part of the gap. We will return to this more seriously in the model. Finally, Panel D shows yields as measured in surveys. Reported harvest output is only 0.4 metric tons per acre, about half of the FAO-GAEZ predictions even for low input usage. Yields are substantially lower in remote areas (24% lower for each log point). This could be because of lower input usage due to higher prices (the focus of this paper) as well as other characteristics of rural farms such as soil.

# 4 Main results

## 4.1 Regression Specification

From the above data sources, we are able to construct transportation costs to every village in our sample, using either survey transport costs or Google maps. Our main empirical specification then

[19]However, note from panel A that travel times on google, at least on rural roads, are about half the travel times experienced by enumerators.

becomes:

$$m_{fvt} = \theta \log(hours_{vt}) + \beta X_{fvt} + \epsilon_{fvt} \tag{1}$$

where $m_{vt}$ is a measure of market access (or related outcome) at location $v$ in year $t$, and $hours_{vt}$ is the hours of travel to the regional hub city of Moshi (see Figures 1 for a map of the area, where Moshi is marked with a star).

The choice of controls varies based on the outcome of interest. We focus on two main sets of outcomes. First, we examine market access, measured at either the village and retailer level. These estimates include all villages in the region (N=570) and all retailers who completed surveys (N=351) . Second, we look at a set of farmer outcomes, including most importantly input adoption, maize sales and productivity, using our surveys from approximately 2,000 farmers in 187 randomly selected villages.

For the market access results, we are primarily interested in measuring market access for farmers and how it correlates with distance; we do not attempt to study *why* it is that remote villages might have limited access. For example, it is likely that intermediary retailer entry in rural areas is limited in part because farmers are poor and demand is limited; nevertheless, this does not change the reality that those farmers in those villages who do wish to purchase inputs must travel further. We therefore argue that the correct specification for market access is one with no controls. However, we also include regressions with controls in the Appendix.

On the other hand, for farmer outcomes such as input adoption, it is clear that usage will depend not only on market access but also other characteristics such as income and land suitability. Thus for the farmer results, we present with and without a vector of other controls, $\mathbf{X_{fvt}}$. These controls include a host of characteristics from the survey, such as land ownership, income, asset ownership, education and other demographic characteristics, as well as soil information from the FAO-GAEZ.

## 4.2 Market Access

### 4.2.1 Village-level market access

We now show some statistics on market access at the village level, for inputs as well as outputs. Since we visited *all* agro retailers and villages in the region, and have GPS coordinates for all of them, we have measures of input access for all villages (not just those in which farmers were surveyed). Therefore, we summarize access by calculating a travel cost-adjusted price of fertilizer for every village as follows:

$$r_v^{tc} = \min_j \{r_j + c_{jv}\} \tag{2}$$

where $r_j$ is the price at agrovet $j$ and $c_{jv}$ is the cost of transporting a bag of fertilizer from agrovet $j$ to village $v$.

For maize prices, we adopt a similar approach, but instead construct the maximum travel cost-adjusted selling price for maize:

$$p_v^{tc} = \max_m \{p_m - c_{mv}\} \tag{3}$$

Here, $p_m$ is the price of maize post-harvest for market $m$, and $c_{mv}$is the cost of traveling from village $v$ to market $m$.

We assume that farmers are free to travel to any market to buy or sell, but must incur a transportation cost, which we calibrate using information from transport surveys and google distances. The results of this calibration exercise are presented in Table 2.[20] We present estimates separately for rural roads, which are relevant for local travel and for which costs per km are higher, and for travel to the 3 major urban hubs for this region (Arusha, Moshi, and Dar-es-Salaam), where costs per km are substantially lower but distances are much longer. For many farmers, the rural road costs are more relevant since the distance to the hub is prohibitively far. We assume that a transaction is for a bag of fertilizer (50 kg). Since this is more than the average purchase, the estimated ad valorem travel cost will be a lower bound. We assume a farmer must make 3 one-way trips to purchase fertilizer (a round-trip for herself plus one additional trip for the bag of fertilizer or maize that she is carrying – this is based on qualitative field reports).

For trips that are entirely on rural roads, i.e., those between villages and their primary markets, the cost is \$0.23 per kilometer. To benchmark this, since a bag of fertilizer is worth about \$20 and the average distance a farmer must travel to minimize travel cost-adjusted prices is about 9 km, this will amount to roughly 10% ad valorem for the average purchase. For many farmers however, the cost may be much higher (and these farmers will likely not buy fertilizer at all). As expected, the calibrated cost of travel is much lower (\$0.06 per kilometer) for trips that use a mix of urban and rural roads (i.e. for trips where one of the terminal nodes is an urban center). We perform a similar calculation on the output side, and we are in the process of refining this measure further. For now, we assume that only the monetary cost of travel matters, and that travel costs are fungible with pecuniary costs. Later, we will show that farmers behave as if the costs they face are substantially higher than these pecuniary costs suggesting the presence of other costs such as search frictions.

We find substantial heterogeneity in these measures of travel cost-adjusted prices. Figure 2 plots CDFs of village-level "best" prices of inputs and output, adjusting for travel costs, and show tremendous heterogeneity in prices across villages. In Panel A, for maize prices, we observe that farmers the standard deviation in the best travel cost-adjusted selling price of maize is 15% of the mean, and the ratio of this price between the 90th and the 10th percentile ratio is similar about 40% of the mean. Panel B shows the distribution of fertilizer prices, where the standard deviation in travel cost-adjusted prices is 15% of the mean, and ratio of this price between the 90th and the 10th percentile is equivalent to about 50% of the mean.

In regression form (Table 4, Panel A), we find that delivered fertilizer prices are 10% with every $\log(hr)$ from Moshi, a substantial difference. These villagers must also travel further to obtain good

[20] In Appendix Table A1, we also present the correlation coefficients between the different measures of travel cost, travel time, and distance from the different surveys and Google maps to establish the quality of the data that underlies this calibration exercise.

prices (implying that prices in the rural areas are higher in general – which we will show in the next table). Somewhat counter-intuitively, we find that remote areas are *more* likely to have an agrovet: while only 26% of villages around Moshi have an agrovet, this increases by 50% with every $\log(hr)$ from Moshi. The explanation for this is that, since travel costs are large, agrovets cannot sell to farmers in remote areas without locating near them – and evidently it is profitable to do this in at least some cases. Further, many villages in remote areas include more land area. However, heterogeneity here is key: while a number of villages in remote areas may have an agrovet, the others are likely very far away from retailers and costs are likely very high for them. On the output side (Panel B), for every $\log(hr)$ from Moshi, villages are located 40% farther from their nearest market.

#### 4.2.2 Fertilizer retailers ("Agrovets")

Table 5 shows results for agrovets. While these numbers are conditional on the decision to enter a particular market, they are nevertheless illustrative. In this table, we use data on wholesale and retail prices. In addition, we collected information on the costs of accessing inputs for retailers. As it turns out, most agrovets pick up fertilizer from the wholesaler and transport it back themselves. We compute markups net of these travel costs. See Web Appendix Figure A1 for histograms of retail prices, wholesale prices, and markups, showing quite a bit of variation in retail prices and markups but less on the wholesale price (as expected, since retailers are traveling to buy inputs from a small set of wholesalers).

We find little effect of travel time on most types of fertilizer sold (other than the "other" category). We find that more remote agrovets sell less fertilizer though in total (a log point decreases sales by about a third). We also find that remote retailers charge higher prices: pooling all fertilizers together and including product fixed effects, for each $\log(hr)$ from Moshi, retail prices rise 0.94 percentage points, or 3.7%. We find that is due to a combination of higher markups and/or higher marginal costs of accessing inputs – most agrovets travel to town to buy fertilizer, so it is no surprise that wholesale prices vary much less with remoteness. Once we adjust mark-ups for travel costs, the mark-up increment for fertilizer rises approximately 10% for each $\log(hr)$ from Moshi.

### 4.3 Farmer outcomes

Tables 6 and 7 show results on farmer outcomes. Table 6 shows dramatically lower input usage in more remote areas: usage declines by 23 percentage points for each $\log(hr)$ away from Moshi, when not including farmer controls, and 13 percentage points when including a broad set of controls for farmer characteristics and soil quality. Benchmarking these effects relative to baseline adoption around Moshi, these translate to percentage declines of 44% and 25%, respectively, with every $\log(hr)$ away from Moshi. Impacts on quantities are even larger, equivalent to about a 64% decline without controls, and 33% with controls. Effects on improved seeds are are more modest, and insignificant in the case of adoption, in part because baseline usage rates are considerably higher. In Panel B, we quantify the distance disadvantage for those farmers who do buy inputs. Conditional on buying, the average farmer travels about 9 km to buy inputs, and farmers 1 log point away travel 4.53 km further. Web Appendix Figure A2 shows the CDF of distances traveled, showing that 40% of farmers buy inputs in their own village, and that only 30% travel more than 15 km. However, later we will document that to get the "best" price, which isn't necessarily at the local agrovet, farmers must travel farther in remote areas. See Web Appendix Figure A3 for a graphical representation of these results.

Table 7 focuses on the output side. Farmers in remote areas are less likely to sell maize to any type of buyer, and the total quantity sold is 20% lower even with controls. This is driven by reduced access to "agents" – maize-buying intermediaries who travel from major towns to buy maize from farmers, to be shipped to larger urban centers. Without controls, farmers who are an additional $\log(hr)$ from Moshi are 21 percentage points less likely to have been visited by an agent, equivalent to a 56% reduction in percentage terms relative to those close to Moshi. With controls, the effect is 14 percentage points, or 37% relative to villages nearby Moshi – still a large effect and highly significant. The reduction in quantity from sales to an agent is very similar to the effect on total sales, suggesting that most of the reduced sales are from this channel. On the other hand, farmers in remote areas are far more likely to buy maize: the probability of buying maize is roughly twice as high in remote areas. See Web Appendix Figure A4 for a graphical representation.

# 5 Model

The reduced form results summarized above suggest that more remote farmers suffer from reduced access to input retailers as well as selling opportunities. However, as other factors are also correlated with distance, we now quantify the impact of access to input and output markets by developing a spatial model of fertilizer adoption. In the model, we will be careful to develop a rigorous model of retailer choice, including for reasons unrelated to trade costs, as well as allowing for other factors that may affect adoption but not related to input access.

## 5.1 Model Preliminaries

### Production and Inputs

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

$$Y_i = \widetilde{\theta}_{i0} K_i^{\alpha_0} L_i^{1-\alpha_0} \tag{4}$$

Here, $\widetilde{\theta}_{i0}$ is baseline productivity without fertilizer for farmer $i$, $K_i$ is land held by farmer $i$, and $L_i$ is labor hired/used by farmer $i$. If the going wage rate in $i$ is $w_i$ and the selling price of maize is

$p_i$, holding land fixed, profits can be derived as:

$$\begin{aligned}\Pi_{i0} &= \alpha_0(1-\alpha_0)^{\frac{1-\alpha_0}{\alpha_0}}\widetilde{\theta}_{i0}^{\frac{1-\alpha_0}{\alpha_0}} p_i^{\frac{1}{\alpha_0}} w_i^{-\frac{1-\alpha_0}{\alpha_0}} K_i \\ &= \theta_{i0}\pi_{i0} \end{aligned} \tag{5}$$

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

The production function *with* fertilizer splits up variable inputs into labor and fertilizer, and also provides a productivity shock, $\widetilde{\theta}_{ijv}$:

$$Y_i = \widetilde{\theta}_{ijv} K^{\alpha} L^{(1-\alpha)\beta} M^{(1-\alpha)(1-\beta)} \tag{6}$$

Note that we are assuming that the exponents on capital and labor may be different for the technology with fertilizer, and that farmer $i$ will receive a productivity of fertilizer that potentially varies by agrovet $j$ in location $v$. We discuss the motivation for this assumption shortly. Writing the delivered price of fertilizer to $i$ from agrovet $j$ in $v$ as $r_{ijv}$, solving for the optimal labor and fertilizer inputs, profits from adoption can be written as:

$$\Pi_i = \theta_{ijv}\pi_i r_{ijv}^{-\sigma} \tag{7}$$

where $\sigma \equiv \frac{1-\alpha}{\alpha}(1-\beta)$, $\pi_i = p_i^{\frac{1}{\alpha_0}} w_i^{-\beta\frac{1-\alpha_0}{\alpha_0}} K_i$, and $\theta_{ijv} = \kappa_2\widetilde{\theta}_{ijv}^{\kappa_1}$.[21] Here, the profitability of fertilizer is a function of the productivity shock, $\theta_{ijv}$, which we will assume is random and with a central moment that is dependent on the source of the fertilizer, the (delivered) price of fertilizer itself, $r_{ijv}$, and profits based on local observables and technology $\pi_i$.

**Input and Agrovet Choice**

Farmers have a choice of whether to purchase fertilizer, and if so, where to purchase fertilizer, which itself is affected by prices for fertilizer at each agrovet location as well as the travel costs to get there and back. Suppose that the set of villages that contain an agrovet is defined as $\mathcal{V}$, where the price charged at location $v \in \mathcal{V}$ by agrovet $j$ is $r_{jv}$. The per-unit cost to the farmer $i$, inclusive of transport costs, will be written as $r_{ijv} = r_{jv}\tau_{iv}$, where $\tau_{iv}$ is an iceberg trade cost for farmer $i$ in traveling to $v$ and back. Further, we assume that $\theta_{ijv}$, is a random variable that could represent unobserved inputs purchased at agrovet $j$ or location $v$, or perhaps other networking and information that is acquired at location $v$ that may increase profitability. Whatever the interpretation, we assume that $\theta_{ijv}$ is distributed according to Fréchet distribution with location parameter $T_{jv}$ and dispersion

[21] $\kappa_1$ and $\kappa_2$ are constant functions of model parameters

parameter $\varepsilon$. Precisely:

$$\Pr\left(\theta_{ijv} < \theta\right) = \exp\left(-T_{jv}\theta^{-\varepsilon}\right) \tag{8}$$

Using this distributional assumption, the unconditional distribution of profits for farmer $i$ buying from agrovet $j$ in location $v$ is written as:

$$\Pr\left(\Pi_{ijv} < \pi\right) = \exp\left(-T_{jv}\pi_i^{\varepsilon} r_{ijv}^{-\varepsilon\sigma}\pi^{-\varepsilon}\right) \tag{9}$$

We also assume that the outside option of not buying fertilizer is random. Specifically, $\theta_{i0}$ is distributed Fréchet with location parameter $T_{i0}$. Thus, the distribution of profits is written as:

$$\Pr\left(\Pi_{i0} < \pi\right) = \exp\left(-T_{i0}\pi_{i0}^{\varepsilon}\pi^{-\varepsilon}\right) \tag{10}$$

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

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

$$\lambda_{ijv} = \frac{T_{jv}\pi_i^{\varepsilon} r_{ijv}^{-\varepsilon\sigma}}{T_{i0}\pi_{i0}^{\varepsilon} + \sum_{v'\in\mathcal{V}}\sum_{l\in\mathcal{J}_{v'}} T_{lv'}\pi_i^{\varepsilon} r_{ilv'}^{-\varepsilon\sigma}} \tag{11}$$

**Decomposing Choice Probabilities**

Note that (11) can be broken up into the probability of adoption for $i$, and the probability $i$ buys from $j$ at location $v$ conditional on buying:

$$\begin{aligned}
\lambda_{ijv} &= \underbrace{\frac{\sum_{v'\in\mathcal{V}}\sum_{l\in\mathcal{J}_{v'}} T_{lv'} r_{ilv'}^{-\varepsilon\sigma}}{T_{i0}\left(\frac{\pi_{i0}}{\pi_i}\right)^{\varepsilon} + \sum_{v'\in\mathcal{V}}\sum_{l\in\mathcal{J}_{v'}} T_{lv'} r_{ilv'}^{-\varepsilon\sigma}}}_{\mu_i} \cdot \underbrace{\frac{T_{jv} r_{ijv}^{-\varepsilon\sigma}}{\sum_{v'\in\mathcal{V}}\sum_{l\in\mathcal{J}_{v'}} T_{lv'} r_{ilv'}^{-\varepsilon\sigma}}}_{\lambda_{ijv|adopt}} && (12)\\
&= \mu_i \cdot \lambda_{ijv|adopt} && (13)
\end{aligned}$$

Conditional on buying, *all* $i$-specific variables drop out of the choice probability. Imposing the iceberg assumption for transport costs between farmer $i$ and location $v$, we can write:

$$\lambda_{ijv|adopt} = \frac{T_{jv} r_{jv}^{-\varepsilon_a}\tau_{iv}^{-\varepsilon_a}}{\sum_{v'\in\mathcal{V}}\sum_{l\in\mathcal{J}_{v'}} T_{lv'} r_{lv'}^{-\varepsilon_a}\tau_{iv'}^{-\varepsilon_a}} \tag{14}$$

where $\varepsilon_a = \varepsilon\sigma$. Since trade costs are specific to location, not agrovet, we can pull the iceberg cost out of the sum across agrovets at each location:

$$\lambda_{ijv|adopt} = \frac{T_{jv} r_{jv}^{-\varepsilon_a} \tau_{iv}^{-\varepsilon_a}}{\sum_{v' \in \mathcal{V}} \tau_{iv'}^{-\varepsilon_a} \sum_{l \in \mathcal{J}_{v'}} T_{lv'} r_{lv'}^{-\varepsilon_a}} \tag{15}$$

Finally, we can rearrange in the form of two probabilities: the probability of buying from *some* agrovet in location $v$, and then conditional on $v$, the probability of buying at agrovet $j$ within $v$. Precisely:

$$\lambda_{ijv|adopt} = \underbrace{\frac{\tau_{iv}^{-\varepsilon_a} \sum_{l \in \mathcal{J}_v} T_{lv} r_{lv}^{-\varepsilon_a}}{\sum_{v' \in \mathcal{V}} \tau_{iv'}^{-\varepsilon_a} \sum_{l \in \mathcal{J}_{v'}} T_{lv'} r_{lv'}^{-\varepsilon_a}}}_{\lambda_{iv|adopt}} \cdot \frac{T_{jv} r_{jv}^{-\varepsilon_a}}{\sum_{l \in \mathcal{J}_v} T_{lv} r_{lv}^{-\varepsilon_a}} \tag{16}$$

To characterize this probability, we will be estimating a functional form for trade costs, the elasticity ($\varepsilon_a$), and then also a vector of $T_{lv}$'s. We address each component in order.

## 5.2 Calibrating the Farmer's Problem

### Estimating Transport Costs through Location Choice

The adopt-location-agrovet decomposition detailed above is helpful in that for two of those stages, we have accurate measures of the probability in question. Adoption probabilities, $\mu_i$, which we will use later, are measured extensively within each village. More novel is our measurement of bilateral probabilities for choosing a particular location to purchase from some agrovet, $\lambda_{iv|adopt}$. While farmers had a somewhat difficult time recalling the exact agrovet they went to for inputs, in most cases they recalled the location (village) in which the agrovet was located. We now exploit this data to estimate transport costs, as revealed by location choices.

To begin, within a location $v$, we can construct the following index that measures the "return" of buying inputs at this location, not including the costs of transport:

$$\phi_v = \sum_{l \in \mathcal{J}_v} T_{lv} r_{lv}^{-\varepsilon_a} \tag{17}$$

Essentially, this is similar to a transformation of an inverse price index, but not accounting for trade costs. These $\phi_v$'s are integrated into the choice probability for location $v$ as follows:

$$\lambda_{iv|adopt} = \frac{\tau_{iv}^{-\varepsilon_a} \phi_v}{\sum_{v' \in \mathcal{V}} \tau_{iv'}^{-\varepsilon_a} \phi_{v'}} \tag{18}$$

To estimate equation (18), we need a dataset that identifies when each farmer $i$ chooses location $v$ to purchase fertilizer. Thus, defining $\mathcal{I}$ as the set of farmers who adopt adoption, and $\mathcal{V}$ as the set of locations with an agrovet, we construct a $\mathcal{I} X \mathcal{V}$ dataset of bilateral visit indicators. There will be many zeros in this dataset. For each bilateral combination, we will also measure distance between the farmer's village and the potential purchase location.

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

$$\lambda_{iv|adopt} = \frac{\exp\left(d_v - \varepsilon_a \log\left(\tau_{iv}\right)\right)}{\sum_{v'\in\mathcal{V}} \exp\left(d_{v'} - \varepsilon_a \log\left(\tau_{iv'}\right)\right)} \tag{19}$$

The main objective from this section is to assess the role of trade costs, so we need to specify a functional form for trade costs, $\tau_{iv}$. As a starting point, we will estimate a simple linear relationship between the elasticity adjusted log trade cost and distance, with dummy variables for whether location $v$ is outside the village's urban or rural market-catchment areas:

$$-\varepsilon_a \log\left(\tau_{iv}\right) = \beta_{dist} dist_{iv} + \beta_{out}^{urb} outsideurban_{iv} + +\beta_{out}^{rural} outsiderural,$$

Here, $dist_{iv}$ is the kilometer distance between farmer $i$ and location $v$, and $outsideurban_{iv}$ is equal to one if the village $i$ and agrovet location $v$ are in different market areas and both within Moshi Mjini (zero otherwise), and $outsiderural_{iv}$ is equal to one if the village $i$ and agrovet location $v$ are in different market areas and either are not in Moshi Mjini (zero otherwise). The idea for having the latter dummies is to account for potentially larger search and information costs outside the market area that the village is most familiar with, especially in rural areas where it may not be common to leave your market often. In contrast, within urban areas, it may be more common to travel across markets since they are more proximate or other differences in the composition of economic activity in towns (as opposed to rural areas).

To allow for a potentially non-linear cost of travel for farmers by distance (for example, if required technologies differ at longer distance), we will also use distance bins $D_{iv}^h$, which are equal to one if the distance between $i$ and $v$ is in bin $h$, and zero otherwise. These will be used as the primary specification as follows:

$$-\varepsilon_a \log\left(\tau_{iv}\right) = \sum_h \beta_h D_{iv}^h + \beta_{out}^{urb} outsideurban_{iv} + \beta_{out}^{rural} outsiderural \tag{20}$$

With the distance bins, the multinomial logit is written as:

$$\lambda_{iv|adopt} = \frac{\exp\left(d_v + \sum_h \beta_h D_{iv}^h + \beta_{out}^{urb} outurban_{iv} + \beta_{out}^{rural} outrural\right)}{\sum_{v'\in\mathcal{V}} \exp\left(d_{v'} + \sum_h \beta_h D_{iv'}^h + \beta_{out}^{urb} outurban_{iv'} + \beta_{out}^{rural} outrural_{iv'}\right)} \tag{21}$$

These cost estimates will include more than just the monetary costs of travel - there may be other hassle/search costs associated with fertilizer purchases, and the distance bins will absorb these aspects of the farmer decision problem as well.

Equation (21) can be estimated by McFadden's alternative-specific conditional logit. The results from doing so are presented in Table 8. In the first column, we present the linear specification of distance. Assuming $\varepsilon_a = -5$ (which will be supported by later results), the results suggest that $\log(\tau)$ increases by 0.035 per kilometer, or approximately 3.6% ad-valorem per kilometer. As technologies may change discretely depending on the distance to each agrovet (walking short distances, taking transit for long distances), our preferred specification using distance bins is presented in column 2 of Table 8. The estimates suggest costly travel for farmers acquiring fertilizer. To interpret the coefficients, we take two approaches. In the first, we can compare two locations with the same "return" from fertilizer, $d_v$, and that are assumed to be within the same market, and then focus on the reduction in probability if one is 1-5km way rather than 0-1 km away (essentially in the same village). In this case, the probability that one chooses the location 1-5km away compared to 0-1km away for idiosyncratic reasons that overcome trade costs is 0.478.[22]

We can also interpret the results as log changes in trade costs via:

$$\log(\tau_{iv}) = -\frac{1}{\varepsilon_a}\sum_h \beta_h D_{iv}^h \tag{22}$$

Dividing the coefficient estimates by $\varepsilon_a$ give us the log change in trade costs. Given the iceberg assumption, this is also interpreted as the log change in the delivered price. Thus, at the central estimate of $\varepsilon_a = -5$, the comparison is equivalent to a 3.2% ad-valorem trade cost over 5km. Using the upper bound of each bin as a conservative measure to calculate ad-valorem trade costs, the ad-valorem equivalent rises within each bin - for choosing an agrovet 15-20km away, the approximate ad-valorem equivalent trade costs is 7.9% per km.

Further, we note that, conditional on distance, there is negative effect on choosing an agrovet outside of the village's market catchment area when operating in rural areas. Precisely, the probability of choosing an agrovet outside the market catchment area is 0.12 relative to choosing one within. The ad-valorem equivalent cost of this friction is approximately 53%.

Finally, we repeat the exercise from the reduced form section of the paper and calculate best trade-cost-adjusted-prices for agrovets for all villages in the region, using the estimates of iceberg costs as described above. These results are presented in Figure 3. Here, there is significantly more heterogeneity in best trade cost adjusted prices for fertilizer.

### Calibrating Non-price Attributes of Each Agrovet

In the conditional multinomial logit used above, if enough farmers were sampled such that every location with an agrovet was chosen enough times, it would be possible to estimate precisely a value of $\phi_v$ for each location (up to the usual normalization), and use this in resulting counterfactuals. Unfortunately, funding was not sufficient to survey such a large sample, and thus, to recover all non-price attributes of all locations that contain an agrovet, we must employ a combination of agrovet revenue shares from our agrovet survey, and the spatial distribution of fertilizer expenditures from the farmer survey.

[22]This is calculated precisely by calculating the ratio of probabilities:

$$\frac{\lambda_{1-5km}}{\lambda_{0-1km}} = \frac{\exp(d_v - 0.738)}{\exp(d_v - 0)} = 0.478$$

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

$$\mathbb{E}\left[v_{jv}\right] = \sum_i L_i \mu_i \lambda_{ijv|adopt} \mathbb{E}\left[rm_{ijv}|adopt\ at\ jv\right] \tag{23}$$

where $\mathbb{E}\left[rm_{ijv}|adopt\ at\ j\right]$ is expected expenditures by $i$ on fertilizer from agrovet $j$ in location $v$, conditional on $j, v$ being the best option, and $L_i$ is the village population to use as weights in the demand equation. Noting that $\mu_i = \frac{\mathbb{E}[rm_i]}{\mathbb{E}[rm_i|adopt]}$, which is the ratio of the unconditional expected expenditures on fertilizer to the expected expenditures on fertilizer, given adoption, we can write:

$$\mathbb{E}\left[v_{jv}\right] = \sum_i L_i \lambda_{ijv|adopt} \frac{\mathbb{E}\left[rm_{ijv}|adopt\ at\ jv\right]}{\mathbb{E}\left[rm_i|adopt\right]} \mathbb{E}\left[rm_i\right] \tag{24}$$

Using the properties of the Fréchet distribution, it is straightforward to show that $\frac{\mathbb{E}[rm_{ijv}|adopt\ at\ jv]}{\mathbb{E}[rm_i|adopt]} = 1$; that is, the expected expenditures conditional on adoption anywhere is the same as the expected expenditures at some $j$, conditional on choosing $j$.[23] Thus, we have:

$$\mathbb{E}\left[v_{jv}\right] = \sum_i \lambda_{ijv|adopt} \mathbb{E}\left[rm_i\right] \tag{25}$$

Imposing the definition of $\lambda_{ijv|adopt}$:

$$\mathbb{E}\left[v_{jv}\right] = \sum_i L_i \left( \frac{T_{jv} \tau_{iv}^{-\varepsilon_a} r_{jv}^{-\varepsilon_a}}{\sum_{v' \in \mathcal{V}} \sum_{l \in \mathcal{J}_{v'}} T_{lv'} \tau_{iv'}^{-\varepsilon_a} r_{lv'}^{-\varepsilon_a}} \right) \mathbb{E}\left[rm_i\right] \tag{26}$$

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

$$\mathbb{E}\left[v_{jv}\right] = \sum_i L_i \left( \frac{\exp\left(-\sum_h \beta_h D_{iv}^h + \beta_{out}^{urb} outurban_{iv} + \beta_{out}^{rural} outrural_{iv}\right) \eta_{jv}}{\sum_{v' \in \mathcal{V}} \sum_{l \in \mathcal{J}_{v'}} \exp\left(-\sum_h \beta_h D_{iv'}^h + \beta_{out}^{urb} outurban_{iv'} + \beta_{out}^{rural} outrural_{iv'}\right) \eta_{lv'}} \right) \mathbb{E}\left[rm_i\right] \tag{27}$$

To implement this equation, we use observed agrovet fertilizer revenues for each agrovet to proxy for $\mathbb{E}\left[v_{jv}\right]$, and village-level fertilizer expenditures from the farmer's survey to proxy for $\mathbb{E}\left[rm_i\right]$. That is, for this equation, we take $i$ to represent villages and sum up expenditures within each village. However, we again run into an issue where our farmer survey was not large enough to survey farmers from every village in the region. Empirically, if we assume that our farmer survey captures the entire geography of demand, there will exist agrovets in other locations that appear more remote than they actually are since we did not survey farmers in that location. This will cause a bias in estimates of $\eta_{jv}$'s by assigning a large value for agrovet locations without any farmers surveyed to make-up for the incorrectly assigned remoteness. At present, the only solution to this problem is to

[23] See technical appendix for a proof

assume that all villages within a market-catchment area share the same characteristics as the (one) surveyed village in that area. Since village selection within a market catchment area was random, this should only add random measurement error to the village $i$ observables that are used in the calibration.

Two other empirical issues to consider are more straightforward. Since agrovet fertilizer revenues and farmer expenditures are from different surveys, and the latter aggregated from a farmer level sample, we normalize each to sum to one. After doing so, we can recover $\eta_{jv}$ by solving the non-linear system of equations formed using $\mathcal{J}$ agrovets and their revenue shares, as written in (27), under the normalizing assumption that $\sum_v \sum_j \eta_{jv} = 1$.[24] Then, after recovering the $\eta_{jv}$'s, we calibrate the value of $\varepsilon_a$ using the following regression (based on $\eta_{jv} = T_{jv} r_{jv}^{-\varepsilon_a}$):

$$\log(\eta_{jv}) = -\varepsilon_a \log(r_{jv}) + \underbrace{\beta_{exp} exper + district + u_{iv}}_{\log(T_{jv})} \tag{28}$$

where $\log(T_{jv})$ will be recovered as the components of the regression not equal to the log agrovet price. To account for other obvious factors that may be correlated with prices and location quality, we include agrovet-level experience at that location, $exper_{iv}$, and district fixed effects. Running this regression, instrumenting for current agrovet prices with one-year lagged prices, yields an estimate of $\varepsilon_a = 5.57$.

In the upper left panel of Web Appendix Figure A5, we summarize the relationship between $T_{jv}$'s and agrovet fertilizer revenues. Naturally, as displayed in the left-hand panel of Web Appendix Figure A5, there is a positive correlation between agrovet non-price attributes and market (revenue) shares. However, the lack of a perfect correlation implies that agrovet pricing and proximity to consuming farmers also matters for market shares. In the upper-right panel, we see that the distribution for $T_{jv}$'s is more disperse than market shares. Intuitively, if an agrovet has a better set of attributes, the agrovet will charge a higher price, which will mitigate the effect of this attribute advantage on market-shares. In the lower-left panel of Web Appendix Figure A5, we regress the values of $\log(T)$ on the number of years each agrovet has operated in their location. Though noisy, there is a clear, positive relationship between estimated non-price attributes and experience within each market.

**Calibrating Adoption**

For the final part of the farmers problem, we work to recover the (relative) productivity of not using fertilizer for each village, $T_{i0}$. To begin, recall that the probability of adoption is written as:

$$\mu_i = \frac{\sum_{v' \in \mathcal{V}} \sum_{l \in \mathcal{J}_{v'}} T_{lv'} \tau_{iv'}^{-\varepsilon_a} r_{lv'}^{-\varepsilon_a}}{T_{i0} \left(\frac{\pi_{i0}}{\pi_i}\right)^{\varepsilon} + \sum_{v' \in \mathcal{V}} \sum_{l \in \mathcal{J}_{v'}} T_{lv'} \tau_{iv'}^{-\varepsilon_a} r_{lv'}^{-\varepsilon_a}} \tag{29}$$

[24]This normalizing assumption is required since the probabilities within the sum in equation (27) are homogeneous degree zero in $\eta$'s.

Expanding the definition of $\frac{\pi_{i0}}{\pi_i}$, we have:

$$\mu_i = \frac{\sum_{v' \in \mathcal{V}} \sum_{l \in \mathcal{J}_{v'}} T_{lv'} \tau_{iv'}^{-\varepsilon_a} r_{lv'}^{-\varepsilon_a}}{T_{i0} p_i^{\varepsilon_p} w_i^{\varepsilon_w} + \sum_{v' \in \mathcal{V}} \sum_{l \in \mathcal{J}_{v'}} T_{lv'} \tau_{iv'}^{-\varepsilon_a} r_{lv'}^{-\varepsilon_a}} \tag{30}$$

where $\varepsilon_p = \varepsilon \left( \frac{\alpha - \alpha_0}{\alpha \alpha_0} \right)$ and $\varepsilon_w = \varepsilon \left( \beta \frac{1-\alpha}{\alpha} - \frac{1-\alpha_0}{\alpha_0} \right)$. Importantly, the elasticities of adoption with respect to price and wages will depend on the values of the parameters of the Cobb-Douglas production functions with and without fertilizer. Thus, before recovering the final productivity parameter for each village $i$, we must estimate the parameters of the production function for maize. For now, we will simply log-linearize the production functions for each technology, and estimate by non-linear least squares. Precisely, we will estimate:

$$\log(Y_i) = \alpha_0 \log(K_i) + (1 - \alpha_0) \log(L_i) + u_i \tag{31}$$

$$\log(Y_i) = \alpha \log(K_i) + (1 - \alpha) \beta \log(L_i) + (1 - \alpha)(1 - \beta) \log(M_i) + e_i \tag{32}$$

where as a reminder, $K_i$ is land, $L_i$ is labor used in the production of maize, and $M_i$ is fertilizer used in the production of maize. We estimate the two equations on the sample of farmers not adopting, and adopting, respectively. Using non-linear least squares, we estimate that $\alpha_0 = 0.87$, $\alpha = 0.61$, and $\beta = 0.1$. With $\alpha$ and $\beta$ we're also able to back-out an estimate for the dispersion parameter of the Fréchet distribution: $\varepsilon = 8.18$. With these estimates, we combine with the average harvest price minus trade costs (discussed below) at the market serving village $i$, $p_i$, and the going daily wage for agricultural work in village $i$, $w_i$, to recover the final parameter of the adoption decision, $T_{i0}$.

Before calibrating $T_{i0}$, we discuss $\mu_i$, the measure of input adoption in village $i$. As discussed in section three, we have extremely detailed information on input adoption in multiple years. For this version of the counterfactuals, we use the fraction of farmers in village $i$ who adopt fertilizer in 2017. To facilitate a feasible calibration, we winsorize the adoption data to fall between 0.05 and 0.95. With this data in hand, we recover $T_{i0}$'s by inverting equation (30) after imposing all estimates and the elasticity adjusted prices and wages.

### Agrovet Pricing and Markups

In the farmers problem, adoption was a function of a quality-adjusted delivered price for fertilizer at each agrovet option, as well as other terms that represent the relative incentives to abstain from using fertilizer. When we evaluate various trade shocks, we could do so while holding agrovet prices fixed. However, while this might be fine for local shocks, for a larger trade shock, such as a roads program, allowing for prices and mark-ups to change is crucial for a realistic counterfactual. We now derive the pricing problem for agrovets, and describe the calibration for mark-ups (which is similar to Berry 1994).

As is well-known, the first order condition for an oligopolist is a mark-up over marginal cost:

$$r_{jv} = \frac{\varepsilon_{jv}^{d}}{\varepsilon_{jv}^{d} + 1} c_{jv}$$

where $c_{jv}$ is the marginal cost for agrovet $j$ in location $v$, and $\varepsilon_{jv}^{d}$ is the elasticity of agrovet $j$ demand with respect to it's own price. Defining $\varepsilon_{jv}^{v}$ as the elasticity of revenue with respect to its own price, we have:

$$r_{jv} = \frac{\varepsilon_{jv}^{v} - 1}{\varepsilon_{jv}^{v}} c_{jv} \tag{33}$$

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

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

Using this elasticity formula calculated for each agrovet, we can then solve for the revealed marginal cost of selling fertilizer by using the mark-up equation. The predicted markups have a mean of 0.21, which while higher is not remarkably different from the measured mark-ups using wholesale prices (mean 0.14).

# 6 Counterfactuals

In this section, we use the calibrated and estimated parameters to evaluate a number of counterfactuals on input and output market access. To solve for the counterfactuals, we simply solve for a new vector of fertilizer prices that solves the first order conditions in (33), while taking into account equilibrium changes in the farmers problem in response to new agrovet prices and/or trade costs.

## 6.1 Experiments on Input Access

We begin by focusing on the effects of local access to fertilizer on adoption decisions. A general hypothesis that we have developed in the paper is that farmers are likely disadvantaged if agrovets are not in close proximity. While many villages have an agrovet at that location, many do not, and in some cases have to travel non-trivial distances to acquire fertilizer and other inputs. To study the role of access to inputs using a realistic counterfactual, we appeal to Casaburi, Glennerster and Suri (2013) and evaluate the effects of a 50% reduction in iceberg costs through a hypothetical roads improvement program. Such a cost reduction can also be motivated by local speeds on trunk roads in Kilimanjaro being approximately 50% lower than US speed (according to Google Maps). The results of this counterfactual are displayed in the top-left panel of Figure 4. Here, we find a large adoption effect of 25pp, or 59% higher than baseline. This counterfactual alone accounts for

45% of the adoption-remoteness relationship that is summarized in the reduced form results. Thus, we conclude that holding local factors fixed, access to input markets has a large effect on adoption levels, and contributes substantially to the reduced adoption levels in remote areas.

Also on the input side, we evaluate how the costs for retailers to source inputs from distributors affect the adoption decision. Through our detailed agrovet surveys, as summarized in the reduced form, we document that the costs of sourcing inputs from distributors rises significantly with distance from Moshi. So, in the second counterfactual, we fully subsidize this cost. The effects of this counterfactual are presented in the top-right panel of Figure 4. Here, adoption rises by about 2pp, or 4.4%, but yields a 9.6% reduction in the remoteness-adoption gradient. Thus, although effects on levels are modest, the effect of this subsidy on the remoteness gradient is non-trivial.

### 6.2 Experiments in Output Access

Also summarized in the reduced form, more remote villages tend to travel farther to reach their primary market. While this does not necessarily correlate with transport adjusted selling prices (since market clearing prices can also vary across space for a variety of reasons), some villages do travel far to reach their primary market, and these travel costs can reduce the margin available to selling their maize harvest. To model the trade costs associated with traveling to the village's primary market, we use the within-market costs of distance as estimated in the agrovet choice problem. That is, the farmer must ship $\tau_{im}$ units of maize to the market to effectively sell one unit. Thus the transport-adjusted selling price that we use for each village above when calibrating adoption decisions is equal to $p_i = p_m/\tau_{im}$, where $p_i$ is net selling price to farmer $i$ and $p_m$ is the price at the primary market for that village. To examine the impact of output market access on adoption, we now run an experiment cutting the iceberg costs to reach output markets by 50%. The results from this counterfactual are presented in the bottom-left panel of Figure 4. Here, adoption rises by 6.7pp, or 15 percent, but there is effectively no change in the remoteness-adoption gradient.

Finally, we offer farmers the market price in Moshi to simulate adoption effects when farmers have ready access to a major "urban" market. Precisely, we offer each farmer the maximum of the Moshi price and each village's transport cost adjusted price to the local market. Here, we find large adoption effects (18pp, or 42%) and an adoption-remoteness gradient that is cut by 10.2%. Thus, access to major markets may provide a significant boost to adoption levels, though this does not correlate substantially with the travel costs adjusted prices in local markets.

## 7 External Validity

As Kilimanjaro is a relatively prosperous region, this begs the question as to the role of remoteness in less developed regions. We address this using other datasets, looking at price dispersion and input adoption.

## 7.1 Price Dispersion

To address this, we assembled five secondary datasets[25] across 1,512 locations[26] in 56 African countries, and compare this to price data we collected between March and August 2016 with 251 retailers of various sorts (shops, agro-input dealers, and maize traders) in 82 markets.[27] To quantify price dispersion, we first decompose variation in spatial prices by running the following regression:

$$\log(p_{mcjt}) = \gamma_c + \gamma_j + \gamma_t + \epsilon_{mjt} \tag{34}$$

where $p_{mjt}$ (log) prices in market $m$ for product $j$ at time $t$ in country $c$, and the $\gamma$ terms are country, product, and time fixed effects. We calculate the standard deviation of the resulting residual. Results are reported in Web Appendix Table A2. In the secondary datasets, the standard deviation is 0.45 for all products, 0.34 for maize, and 0.12 for fertilizer; in our Tanzania data, the figures are 0.22, 0.14, and 0.09. The somewhat lower standard deviation in our data could be indicative of reduced measurement error, or that prices vary less within the geographic concentrated area of Kilimanjaro. Nevertheless, price dispersion is substantial within Kilimanjaro as well.

We also follow the literature[28], to run dyadic regressions to look at price gaps, as follows:

$$\log(\mid p_{mjt} - p_{m'jt} \mid) = \theta \log(c_{mm'}) + \gamma_m + \gamma_{m'} + \gamma_j + \epsilon_{mm'jt} \tag{35}$$

where $p_{mjt} - p_{m'jt}$ is the price gap between markets $m$ and $m'$ and $c_{mm'}$is the cost of transport between markets.[29] Results are presented in Web Appendix Table A3. For each dyad, we regress the absolute difference in log prices on two measures of distance: (1) kilometers between locations in Columns 1, 4, and 7, and (2) driving time between locations in Columns 2, 5, and 8 (both calculated via Google Maps API). We cluster standard errors by both the destination and origin market (Cameron, Gelbach and Miller 2012). In each of the secondary datasets, we find significant, positive coefficients, suggesting that price gaps are larger between more distant markets. The coefficients are economically meaningful: a doubling of travel costs would increase price gaps by about 1-3% in the secondary datasets. In Tanzania, we find that doubling distances would increase price gaps by a similar amount. Finally, we can use this data to provide some descriptive evidence

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

[26]These are not necessarily all unique locations. Though we have cleaned these datasets, there are some misspellings, different names for the same markets, and also differing levels of granularity in the datasets.

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

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

[29]These regressions are motivated by an assumption of free entry where an arbitrageur will enter if $\mid (p_m - p_{m'}) \mid \geq c_{mm'}$. While we know that free entry is not realistic in this context, we reproduce these results for comparability.

on road upgrading. We conjecture that price gaps should respond to the time it takes to travel from point to point, and not the geographic distance (since the time and other costs of traveling to sell items should be what is important). To examine this, we regress price gaps on both distance and duration in Columns 3, 6, and 9. Consistent with priors, we find that duration is significant, whereas distance is not – which suggests that improving road quality would reduce these gaps.

### 7.2 Fertilizer adoption

In Web Appendix Table A4, we assembled data from the World Bank LSMS-ISA household panel surveys for Ethiopia, Niger, Nigeria, Malawi, Tanzania, and Uganda, to study how remoteness affects fertilizer adoption. In the LSMS, measures of remoteness include distance to the main market, and distance to a population center. Using both measures of remoteness, we find a negative association between remoteness and technology adoption. However, since we cannot associate these adoption decisions with prices or precise measures of transport costs, we plan to continue our full field work in other regions of Tanzania. Further, we plan to do extensive work evaluating the sourcing decisions of output buying intermediaries, and how the presence of output buying intermediaries ultimately affects the decision by farmers to adopt fertilizer.

## 8 Conclusion

In this paper we collect extremely detailed and precise data on transportation costs, input and output prices, and the intensive and extensive margins of input purchases and output sales from market actors across the entirety of the supply chains for maize and fertilizer in all 570 villages in the Kilimanjaro region of Northern Tanzania. This data enables us to document large heterogeneity in market access, and study its implications for prices and for market participation. We find that there is large variation in prices, and the remotest villages are disadvantaged by as much as 50% of the mean relative to the least remote villages. The rates and magnitudes of fertilizer use and maize sales also display a large and significant distance gradient. Five different counterfactuals on the input and output sides suggest an important role for lowering access barriers on the input side for reducing this gradient, and a more modest one for those on the output side.

The results of these counterfactuals lead directly to the question of policy implications. Many African countries have experimented with input subsidies (some intermittently, and some more consistently), and these have had large adoption and usage effects by directly lowering the delivered price of fertilizer even though the transport cost may have stayed unaffected. However, most farmers fail to graduate out of the subsidy for a host of reasons, potentially including the fact that the market access issues remain unresolved. Therefore, policies that lower fertilizer prices through reducing transport costs can potentially have lasting effects, such as improving transportation linkages between markets and villages, and also between urban centers and villages. Initiatives to organize farmers into cooperative groups that enable them to defray the total costs of transportation over a large number of buyers may also be helpful.

## References


[1] Adukia, Anjali, Sam Asher, and Paul Novosad (2016). "Educational Investment Responses to Economic Opportunity: Evidence from Indian Road Construction." Unpublished

[2] Aggarwal, Shilpa (2018). "Do Rural Roads Create Pathways out of Poverty? Evidence from India." *Journal of Development Economics* 133: 375-395.

[3] Alder, Simon (2017). "Chinese Roads in India: The Effect of Transport Infrastructure on Economic Development." Unpublished.

[4] Allen, Treb and Costas Arkolakis (2016). "The Welfare Effects of Transportation Infrastructure Improvements." Unpublished.

[5] Asher, Sam and Paul Novosad (2016). "Market Access and Structural Transformation: Evidence from Rural Roads in India." Unpublished.

[6] Asturias, Jose, Manuel Garcia-Santana, and Roberto Ramos (2017). "Competition and the welfare gains from transportation infrastructure: Evidence from the Golden Quadrilateral of India." Unpublished.

[7] Atkin, David and David Donaldson (2015). "Who's Getting Globalized? The Size and Implications of Intra-National Trade Costs." NBER Working Paper No. 21439

[8] Bandiera, Oriana and Imran Rasul (2006). "Social networks and technology adoption in northern Mozambique." *Economic Journal* 116 (514): 869-902.

[9] Banerjee, Abhijit, Esther Duflo, and Nancy Qian (2012). "On the Road: Access to Transportation Infrastructure and Economic Growth in China," Unpublished.

[10] Bardhan, Pranab and Dilip Mookherjee (2011). "Subsidized Farm Input Programs and Agricultural Performance: A Farm-Level Analysis of West Bengal's Green Revolution, 1982-1995." *American Economic Journal: Applied Economics*, 3(4): 186-214.

[11] Baum-Snow, Nate (2007). "Did Highways Cause Suburbanization?" *Quarterly Journal of Economics*, 122, 775–805.

[12] Beaman, Lori, Dean Karlan, Bram Thuysbaert, and Christopher Udry (2013). "Profitability of Fertilizer: Experimental Evidence from Female Rice Farmers in Mali." *American Economic Review* 103 (3): 381-86.

[13] Bergquist, Lauren (2017). "Pass-Through, Competition, and Entry in Agricultural Markets: Experimental Evidence from Kenya." Unpublished.

[14] Berry, Steven (1994). "Estimating Discrete-Choice Models of Product Differentiation." *RAND Journal of Economics* 25 (2): 242-262.

[15] Bird, Julia and Stephane Straub (2014). "The Brasilia experiment: road access and the spatial pattern of long-term local development in Brazil." Unpublished.

[16] Cagley, Jessica, Mary Kay Gugerty, and Robert Plotnick (2009). "Political Economy of Fertilizer Policy in Tanzania." Prepared for the Farmer Productivity Team of the Bill and Melinda Gates Foundation.

[17] Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller (2011). "Robust inference with multiway clustering." *Journal of Business & Economic Statistics* 29 (2): 238-249.

[18] Casaburi, Lorenzo and Tristan Reed (2017). "Competition in Agricultural Markets: An Experimental Approach." Unpublished

[19] Chatterjee, Shoumitro (2018). "Market Power and Spatial Competition in Rural India." Unpublished.

[20] Conley, Timothy and Christopher Udry (2010). "Learning about a new technology: Pineapple in Ghana." *American Economic Review* 100 (1): 35-69.

[21] Costinot, Arnaud and Andres Rodriguez-Clare (2014). "Trade Theory with Numbers: Quantifying the Consequences of Globalization." In *Handbook of International Economic*s, Volume 4, Chapter 4. Editors: Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff. Elsevier.

[22] Dhingra, Swati and Silvana Tenreyro (2017). "Piggy-Back Exporting, Intermediation, and the Distributional Gains from Trade in Agricultural Markets." Unpublished.

[23] Donaldson, David (2015). "The Gains from Market Integration." *Annual Review of Economics* 7: 619-647.

[24] Donaldson, David (2018) "Railroads of the Raj: Estimating the impact of transportation infrastructure." *American Economic Review* 108 (4-5):899-934.

[25] Duflo, Esther, Michael Kremer, and Jonathan Robinson (2008). "How high are rates of return to fertilizer? Evidence from field experiments in Kenya." *American Economic Review* 98 (2): 482-488.

[26] Duflo, Esther, Michael Kremer, and Jonathan Robinson (2011). "Nudging farmers to use fertilizer: theory and experimental evidence from Kenya." *American Economic Review* 101 (6): 2350-2390.

[27] Eaton Jonathan and Samuel Kortum (2002). "Technology, Geography, and Trade." *Econometrica* 70: 1741–1779

[28] Emerick, Kyle (2017). "Trading Frictions in Indian Village Economies." Forthcoming, *Journal of Development Economics.*

[29] Foster, Andrew and Mark Rosenzweig (1995). "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture." *Journal of Political Economy* 103(6): 1176-1209

[30] Foster, Andrew and Mark Rosenzweig (2010). "Microeconomics of Technology Adoption." *Annual Review of Economics* 2: 395-424.

[31] Gertler, Paul, Tadeja Gracner, Marco Gonzalez-Navarro, and Alex Rothenberg (2014). "Road Quality and Local Economic Activity: Evidence from Indonesia's Highways." Unpublished.

[32] Ghani, Ejaz, Arti G. Goswami, and William R. Kerr (2016). "Highway to success: The impact of the Golden Quadrilateral project for the location and performance of Indian manufacturing." *Economic Journal* 126 (591): 317-357.

[33] Gollin, Doug, David Lagakos, and Michael Waugh (2014a). "The Agricultural Productivity Gap." *Quarterly Journal of Economics* 129 (2): 939-993.

[34] Gollin, Doug, David Lagakos, and Michael Waugh (2014b). "Agricultural Productivity Differences across Countries." *American Economic Review* 104 (5): 165-170.

[35] Gollin, Doug and Richard Rogerson (2014). "Productivity, Transport Costs, and Subsistence Agriculture." *Journal of Development Economics* 107: 38-48.

[36] Hanna, Rema, Sendhil Mullainathan, and Josh Schwartzstein (2014). "Learning through noticing: Theory and evidence from a field experiment." *Quarterly Journal of Economics* 129 (3): 1311-1353.

[37] Hernandez, Manuel and Maximo Torero (2011). "Fertilizer Market Situation. Market Structure, Consumption and Trade Patterns, and Pricing Behavior." International Food Policy Research Institute (IFPRI), Washington, DC, USA.

[38] Jack, B. Kelsey (2013). "Market Inefficiencies and the Adoption of Agricultural Technologies in Developing Countries." White Paper. Agricultural Technology Adoption Initiative, JPAL (MIT) and CEGA (UC Berkeley)

[39] Karlan, Dean, Robert Osei, Isaac Osei-Akoto, and Christopher Udry (2015). "Agricultural Decisions after Relaxing Credit and Risk Constraints." *Quarterly Journal of Economics* 129 (2): 597-652.

[40] Keller, Wolfgang and Carol Shiue (2007). "Markets in China and Europe on the Eve of the Industrial Revolution." *American Economic Review* 97 (4): 1189–1216.

[41] Khanna, Gaurav (2016). "The Road Oft Taken: The Route to Spatial Development." Unpublished.

[42] Maitra, Pushkar, Sandip Mitra, Dilip Mookherjee, Alberto Motta, and Sujata Visaria (2017). "Financing Smallholder Agriculture: An Experiment with Agent-Intermediated Microloans in India." *Journal of Development Economics* 127: 306 – 337.

[43] Morten, Melanie and Jaqueline Oliveira (2016). "The Effects of Roads on Trade and Migration: Evidence from a Planned Capital City." NBER Working Paper 22158.

[44] Redding, Stephen, and Daniel Sturm. (2008). "The Costs of Remoteness: Evidence from German Division and Reunification." *American Economic Review* 98 (5): 1766-97.

[45] Rothenberg, Alex (2013). "Transport Infrastructure and Firm Location Choice in Equilibrium: Evidence from Indonesia's Highways." Unpublished.

Sotelo, Sebastian (2016). "Domestic Trade Frictions and Agriculture."Unpublished.

[46] Stewart, W.M., D.W. Dibb, A.E. Johnston, and T.J. Smyth (2005). "The contribution of commercial fertilizer nutrients to food production." *Agronomy Journal* 97 (1): 1-6

[47] Storeygard, Adam (2016). "Farther on down the Road: Transport costs, trade and urban growth in Sub-Saharan Africa." *Review of Economic Studies* 83 (3): 1263-1295.

[48] Suri, Tavneet (2011). "Selection and Comparative Advantage in Technology Adoption." *Econometrica* 79 (1): 159-209

[49] Udry, Christopher and Santosh Anagol (2006). "The Return to Capital in Ghana." *American Economic Review* 96 (2): 388-393.

[50] World Bank (2010). *World Development Report 2008: Agriculture for Development.* Washington, DC: World Bank.

[51] World Bank (2010). "Africa's Infrastructure: A time for Transformation." Africa Development Forum. Washington, DC.

[52] World Bank (2017). "Enabling the Business of Agriculture 2017." Washington, DC: World Bank. doi:10.1596/978-1-4648-1021-3. License: Creative Commons Attribution CC BY 3.0 IGO

[53] Yi, Kei-Mu (2003). "Can Vertical Specialization Explain the Growth of World Trade?" *Journal of Political Economy* 111 (1): 52-102.

# A Deriving farmer profits, revenues, and input expenditures

To be written

# B Distributions of Fertilizer Expenditures

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

$$\mathbb{E}\left[rm_{ijv}|adopt\ at\ j\right] = \mathbb{E}\left[rm_i|adopt\right] \tag{36}$$

That is, that the expected fertilizer expenditures, conditional on adopting at location $j$, is the same as the expected fertilizer expenditure, conditional on adopting anywhere. This is a similar result to Eaton and Kortum (2003), where the price distribution conditional on being the lowest price supplier is the same as the unconditional price distribution at that destination. Here, we prove the similar result in the input adoption context.

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

$$\begin{aligned}
\Pr\left(\Pi_{ijv} > \pi\right) &= \Pr\left(\theta_{ijv}\pi_i r_{ijv}^{-\sigma} > \pi\right) && (37)\\
&= \Pr\left(\theta_{ijv} > \frac{\pi}{\pi_i} r_{ijv}^{\sigma}\right) && (38)\\
&= 1 - \exp\left(-T_{jv}\pi_i^{\varepsilon} r_{ijv}^{\varepsilon\sigma}\pi^{-\varepsilon}\right) && (39)
\end{aligned}$$

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

$$\Pr\left(\Pi_{ijv} > \pi\right) = 1 - \exp\left(-T_{jv}\gamma_{ijv}\pi^{-\varepsilon}\right) \tag{40}$$

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

$$\Pr\left(\Pi_{i0} > \pi\right) = 1 - \exp\left(-\Phi_{i0}\pi^{-\varepsilon}\right) \tag{41}$$

where $\Phi_{i0} = T_{i0}\gamma_{i0} \equiv \pi_i^{\varepsilon}$

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

$$\begin{aligned}
\Pr\left(\Pi_i^{max} > \pi\right) &= \Pr\left(\Pi_{ijv} > \pi \text{ for any } jv\right) && (42)\\
&= 1 - \Pr\left(\Pi_{ijv} < \pi \ \forall\ jv\right) && (43)
\end{aligned}$$

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

$$\begin{aligned}
\Pr\left(\Pi_i^{max} > \pi\right) &= 1 - \Pr\left(\Pi_{ijv} < \pi \ \forall\ jv\right) && (44)\\
&= 1 - \prod_{v'\in\mathcal{V}}\prod_{j\in\mathcal{J}_v} \Pr\left(\Pi_{ijv} < \pi\right) && (45)\\
&= 1 - \prod_{v'\in\mathcal{V}}\prod_{j\in\mathcal{J}_v} \exp\left(-\pi^{-\varepsilon}\right) && (46)
\end{aligned}$$

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

$$\Pr\left(\Pi_i^{max} > \pi\right) = 1 - \exp\left(-\Phi_i\pi^{-\varepsilon}\right) \tag{47}$$

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

$$G_i^{max}\left(\pi\right) = \Pr\left(\Pi_i^{max} < \pi\right) = \exp\left(-\Phi_i\pi^{-\varepsilon}\right) \tag{48}$$

with pdf:

$$g_i^{max}(\pi) \quad = \quad \varepsilon \Phi_i \pi^{-\varepsilon-1} \exp\left(-\Phi_i \pi^{-\varepsilon}\right) \tag{49}$$

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

$$\begin{aligned} \Pr(\Pi_i > \pi) \quad &= \quad \Pr(\Pi_{ijv} > \pi \text{ for any } jv \cup \Pi_{i0} > \pi) && (50) \\ &= \quad 1 - \Pr(\Pi_{ijv} < \pi \ \forall \ jv \ \cap \ \Pi_{i0} < \pi) && (51) \end{aligned}$$

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

$$\begin{aligned} \Pr(\Pi_i > \pi) \quad &= \quad 1 - \Pr(\Pi_{ijv} < \pi \ \forall \ jv \ \cap \ \Pi_{i0} < \pi) && (52) \\ &= \quad 1 - \Pr(\Pi_{i0} < \pi) \prod_{v' \in \mathcal{V}} \prod_{j \in \mathcal{J}_v} \Pr(\Pi_{ijv} < \pi) && (53) \\ &= \quad 1 - \exp\left(-T_{i0}\gamma_{i0}\pi^{-\varepsilon}\right) \prod_{v' \in \mathcal{V}} \prod_{j \in \mathcal{J}_v} \exp\left(-T_{jv}\gamma_{ijv}\pi^{-\varepsilon}\right) && (54) \end{aligned}$$

Using the definitions for $\Phi_{i0}$ and $\Phi_i$, this is simplified as:

$$\Pr(\Pi_i > \pi) \quad = \quad 1 - \exp\left(-\left(\Phi_{i0} + \Phi_i\right)\pi^{-\varepsilon}\right) \tag{55}$$

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

$$G_i(\pi) \quad = \quad \exp\left(-\left(\Phi_{i0} + \Phi_i\right)\pi^{-\varepsilon}\right) \tag{56}$$

with pdf:

$$g_i(\pi) \quad = \quad \varepsilon\left(\Phi_{i0} + \Phi_i\right)\pi^{-\varepsilon-1} \exp\left(-\left(\Phi_{i0} + \Phi_i\right)\pi^{-\varepsilon}\right) \tag{57}$$

**Profits conditional on adoption**

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

$$\Pr\left(\Pi_i^{max} < \pi \middle| adopt\right) \quad = \quad \frac{\Pr\left(\Pi_i^{max} < \pi \cap \Pi_i^{max} > \Pi_{i0}\right)}{\Pr\left(\Pi_i^{max} > \Pi_{i0}\right)} \tag{58}$$

This can be re-written as:

$$\begin{aligned}
\Pr\left(\Pi_i^{max} < \pi \middle| adopt\right) &= \frac{1}{\Pr\left(\Pi_i^{max} > \Pi_{i0}\right)} \int_0^{\pi} \Pr\left(s > \Pi_{i0}\right) g_i^{max}(s) ds \\
&= \frac{1}{\Pr\left(\Pi_i^{max} > \Pi_{i0}\right)} \int_0^{\pi} \exp\left(-\Phi_{i0} s^{-\varepsilon}\right) \varepsilon \Phi_i s^{-\varepsilon-1} \exp\left(-\Phi_i s^{-\varepsilon}\right) ds \\
&= \frac{1}{\Pr\left(\Pi_i^{max} > \Pi_{i0}\right)} \int_0^{\pi} \varepsilon \Phi_i s^{-\varepsilon-1} \exp\left(-\left(\Phi_{i0} + \Phi_i\right) s^{-\varepsilon}\right) ds \qquad (59)
\end{aligned}$$

Mulitplying by $\frac{\Phi_{i0}+\Phi_i}{\Phi_{i0}+\Phi_i}$, and then factoring out $\frac{\Phi_i}{\Phi_{i0}+\Phi_i}$, we have:

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

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

$$\begin{aligned}
\Pr\left(\Pi_i^{max} < \pi \middle| adopt\right) &= \int_0^{\pi} \varepsilon \left(\Phi_{i0} + \Phi_i\right) s^{-\varepsilon-1} \exp\left(-\left(\Phi_{i0} + \Phi_i\right) s^{-\varepsilon}\right) ds \qquad (60) \\
&= \Pr\left(\Pi_i < \pi\right) \qquad (61)
\end{aligned}$$

**Profits conditional on adoption from $j$**

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

$$\Pr\left(\Pi_{ijv} < \pi \middle| adopt\ from\ j\right) = \frac{\Pr\left(\Pi_{ijl} < \pi \cap \Pi_{ijv} > \Pi_{ij'l} \forall (j', l) \cap \Pi_{ijv} > \Pi_{i0}\right)}{\Pr\left(\Pi_{ijv} > \Pi_{ij'l} \forall (j', l) \cap \Pi_{ijv} > \Pi_{i0}\right)} \qquad (62)$$

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

$$\Pr\left(\Pi_{ijv} < \pi \middle| adopt\ from\ j\right) = \frac{1}{\lambda_{ijv}} \int_0^{\pi} Pr\left(s > \Pi_{ij'l} \forall (j', l) \cap s > \Pi_{i0}\right) g_{ijv}(s) ds \qquad (63)$$

Defining $\Phi_{i\not{j}v} = \left(\sum_{v' \in \mathcal{V}} \sum_{j \in \mathcal{J}_v} T_{jv} \gamma_{ijv}\right) - T_{jv} \gamma_{ijv}$, we can simplify $\Pr\left(s > \Pi_{ij'l} \forall (j', l) \cap s > \Pi_{i0}\right)$ as

$$\begin{aligned}
\Pr\left(s > \Pi_{ij'l} \forall (j', l) \cap s > \Pi_{i0}\right) &= \exp\left(-\Phi_{i0} s^{-\varepsilon}\right) \exp\left(-\Phi_{i\not{j}v} s^{-\varepsilon}\right) \qquad (64) \\
&= \exp\left(-\left(\Phi_{i0} + \Phi_{i\not{j}v}\right) s^{-\varepsilon}\right) \qquad (65)
\end{aligned}$$

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

$$\Pr\left(\Pi_{ijv} < \pi \middle| adopt\ from\ j\right) = \frac{1}{\lambda_{ijv}} \int_0^{\pi} \exp\left(-\left(\Phi_{i0} + \Phi_{i\not{j}v}\right) s^{-\varepsilon}\right) \varepsilon T_{jv} \gamma_{ijv} \pi^{-\varepsilon-1} \exp\left(-T_{jv} \gamma_{ijv} s^{-\varepsilon}\right) ds$$

Factoring out $\frac{T_{jv}\gamma_{ijv}}{\Phi_{i0}+\Phi_i}$, and then nothing that $\Phi_{i0} + \Phi_i = \Phi_{i0} + \Phi_{i\cancel{j}v} + T_{jv}\gamma_{ijv}$, we have:

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

Since $\lambda_{ijv} = \frac{T_{jv}\gamma_{ijv}}{\Phi_{i0}+\Phi_i}$, we land at the final result:

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

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

# C Mark-ups

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

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

Since fertilizer expenditures are proportional to profits, we have:

$$\mathbb{E}\left[v_{jv}\right] = \sum_i \lambda_{ijv|adopt} \mathbb{E}\left[rm_i\right] \tag{66}$$

Differentiating with respect to the fertilizer price, $r_{jv}$, we have:

$$\frac{d\mathbb{E}\left[v_{jv}\right]}{dr_{jv}} = \sum_i -\lambda_{ijv|adopt}\left(1 - \lambda_{ijv|adopt}\right)\frac{\varepsilon_a}{r_{jv}}\mathbb{E}\left[rm_i\right] + \sum_i \lambda_{ijv|adopt}\frac{d\mathbb{E}\left[rm_i\right]}{dr_{jv}}$$

It is straightforward to show that $\frac{d\mathbb{E}[rm_i]}{dr_{jv}} = -\mathbb{E}\left[rm_i\right]\left(1 - \mu_i + \frac{1}{\varepsilon}\right)\lambda_{ijv|adopt}\frac{\varepsilon_a}{r_{jv}}$. Hence:

$$\frac{d\mathbb{E}\left[v_{jv}\right]}{dr_{jv}} = -\sum_i \lambda_{ijv|adopt}\left(1 - \lambda_{ijv|adopt}\right)\frac{\varepsilon_a}{r_{jv}}\mathbb{E}\left[rm_i\right] - \sum_i \lambda_{ijv|adopt}\mathbb{E}\left[rm_i\right]\left(1 - \mu_i + \frac{1}{\varepsilon}\right)\lambda_{ijv}\frac{\varepsilon_a}{r_{jv}}$$

Defining $s_{ijv} = \frac{\lambda_{ijv|adopt}\mathbb{E}[rm_i]}{\sum_{i'}\lambda_{i'jv|adopt}\mathbb{E}[rm_{i'}]}$ as the expenditure share on $i$ within $j$, this can be simplified as:

$$\frac{d\mathbb{E}\left[v_{jv}\right]}{dr_{jv}}\frac{r_{jv}}{\mathbb{E}\left[v_{jv}\right]} = -\varepsilon_a \sum_i s_{ijv}\left(\left(1 - \lambda_{ijv|adopt}\right) + \left(1 - \mu_i + \frac{1}{\varepsilon}\right)\lambda_{ijv|adopt}\right) \equiv \varepsilon_v \tag{67}$$

which is the elasticity of agrovet revenues with respect to its price.

**Table 1. Summary statistics on villages in Kilimanjaro region**

| | (1)<br>Mean |
|---|---|
| **Panel A. Villages (N = 570)** | |
| Population | 2842.15 |
| | (1882.22) |
| Distance to nearest market center (km) - Google maps | 5.69 |
| | (9.10) |
| Time for round-trip journey to nearest market center (mins) - Google maps | 21.28 |
| | (32.48) |
| Time for round-trip journey to nearest market center - surveys | 40.43 |
| | (33.03) |
| Cost of round-trip journey to nearest market center (USD) - surveys | 1.59 |
| | (1.94) |
| Distance to Moshi (km) - Google maps | 65.76 |
| | (52.52) |
| Round-trip travel time to Moshi (mins) - Google maps | 177.23 |
| | (117.92) |
| Round-trip cost of travel to Moshi USD - surveys | 4.69 |
| | (3.76) |
| **Panel B. Road Quality (N = 570)** | |
| *Measurement of roads in field* | |
| Percent of road that is: | |
| Paved | 0.27 |
| Dirt | 0.35 |
| Gravel | 0.37 |
| | |
| Cost of trip from market center to village (paid by enumerator) | 0.91 |
| | (1.19) |
| *Google estimates* | |
| Travel speed on major roads - km/hr (Google) | 49.5 |
| | |
| Travel speed on feeder roads and rural roads - km/hr (Google) | 30.6 |

Notes: Standard deviations in parentheses.

**Table 2. Calibrating Travel Costs**

| | (1) | (2) |
|---|---|---|
| | Cost of transporting 50 kg bag of fertilizer from destination (USD) | |
| **Panel A. Rural Roads** | | |
| Google maps: kilometers to destination | 0.23*** | |
| | (0.022) | |
| Google maps: hours to destination | | 7.97*** |
| | | (0.808) |
| Number of villages | 533 | 533 |
| Number of observations | 533 | 533 |
| **Panel B. Travel to major cities** | | |
| **Panel B1. Market centers to major cities** | | |
| Google maps: kilometers to destination | 0.06*** | |
| | (0.001) | |
| Google maps: hours to destination | | 3.44*** |
| | | (0.056) |
| Number of markets | 93 | 93 |
| Number of observations | 279 | 279 |
| **Panel B2. Village centers to major cities** | | |
| Google maps: kilometers to destination | 0.06*** | |
| | (0.01) | |
| Google maps: hours to destination | | 3.60*** |
| | | (0.025) |
| Number of villages | 561 | 561 |
| Number of observations | 1660 | 1660 |

Notes: We assume purchasing fertilizer requires a round-trip for the farmers to the market center, plus the cost of transporting the fertilizer itself. Trips are calculated from the village center and so travel from home to the village center are not included. Standard errors clustered by location in parentheses. *, **, and *** indicate significance at 10%, 5%, and 1% respectively.
*, **, and *** indicate significance at 10%, 5%, and 1%.

**Table 3. Summary statistics on farmers, and correlation with remoteness**

| | (1) | (2) | (3) |
|---|---|---|---|
| | Mean | Regression on log (distance from regional hub) | |
| | | Constant | Coefficient |
| **Panel A. Demographic and background characteristics** | | | |
| Age | 51.07 | 51.40*** | -1.26** |
| | (15.10) | (0.52) | (0.60) |
| Female | 0.46 | 0.48*** | 0.00 |
| | (0.50) | (0.02) | (0.03) |
| Married | 0.73 | 0.73*** | 0.00 |
| | (0.44) | (0.01) | (0.01) |
| Household size | 4.60 | 4.55*** | 0.07 |
| | (2.17) | (0.08) | (0.09) |
| Years of education | 6.98 | 7.04*** | -0.13 |
| | (3.25) | (0.11) | (0.12) |
| Home has thatch roof | 0.12 | 0.11*** | -0.04** |
| | (0.32) | (0.02) | (0.02) |
| Has cell phone | 0.90 | 0.90*** | -0.03** |
| | (0.30) | (0.01) | (0.01) |
| Has bank account | 0.17 | 0.19*** | -0.08*** |
| | (0.38) | (0.01) | (0.02) |
| Has mobile money account | 0.83 | 0.84*** | -0.04** |
| | (0.38) | (0.01) | (0.02) |
| Acres of land | 2.92 | 2.78*** | 0.84*** |
| | (3.53) | (0.16) | (0.27) |
| Has market business | 0.28 | 0.31*** | -0.04* |
| | (0.45) | (0.02) | (0.02) |
| Annual total income from non-farming (USD) | 436.3 | 458.75*** | -48.62 |
| | (833.20) | (38.01) | (44.25) |
| **Panel B. Buying and Selling Maize** | | | |
| Farmer buys maize but sells none | 0.36 | 0.32*** | 0.13*** |
| | (0.48) | (0.02) | (0.03) |
| Farmer sells maize and buys none | 0.21 | 0.26*** | -0.11*** |
| | (0.41) | (0.02) | (0.02) |
| Farmer buys and sells maize | 0.07 | 0.07*** | -0.01 |
| | (0.25) | (0.01) | (0.01) |
| Farmer buys none and sells none | 0.36 | 0.34*** | -0.01 |
| | (0.48) | (0.02) | (0.02) |
| Net buyer (quantity bought > quantity sold) | 0.39 | 0.37*** | 0.11*** |
| | (0.49) | (0.02) | (0.03) |
| Net seller (quantity bought < quantity sold) | 0.23 | 0.27*** | -0.11*** |
| | (0.42) | (0.02) | (0.03) |
| **Panel C. Production Capacity (in kg/acre)[1]** | | | |
| FAO-GAEZ production capacity for low input level | 825.3 | 769.66*** | 145.94*** |
| | (267.50) | (23.84) | (25.70) |
| FAO-GAEZ production capacity for high input level | 3334 | 3,389.00*** | -225.98*** |
| | (678.70) | (58.65) | (86.01) |
| FAO-GAEZ production difference between high and low | 2509 | 2,619.34*** | -371.92*** |
| | (607.50) | (46.46) | (83.41) |
| **Panel D. Harvest Output** | | | |
| Total harvest output in 2016 long rains (kg) | 634.4 | 692.72*** | -122.85*** |
| | (632.00) | (34.79) | (29.82) |
| Harvest output per acre | 435.9 | 481.38*** | -115.65*** |
| | (389.10) | (17.80) | (25.61) |
| Value of harvest output at average regional post-harvest price | 158.4 | 172.39*** | -29.42*** |
| | (158.60) | (8.51) | (7.71) |

Notes: N = 1,977 farmers in 187 villages. In Column 1, standard deviations in. In Columns 2-3, standard errors in parentheses, clustered at the village level. *, **, and *** indicate significance at 10%, 5%, and 1%.

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

**Table 4. Remoteness and village-level market access**

| | (1) | (2) | (3) |
|---|---|---|---|
| | Mean | Regression on log (travel time from regional hub) | |
| | | Constant | Coefficient on log (travel time) |
| **Panel A. Input markets** | | | |
| Has an agrovet in village which sells fertilizer | 0.28 | 0.26*** | 0.13*** |
| | | (0.02) | (0.03) |
| Has at least 1 agrovet within 10 km of village which sells fertilizer | 0.90 | 0.90*** | -0.04** |
| | | (0.01) | (0.02) |
| Number of agrovets within 10 km of village which sells fertilizer | 8.08 | 8.40*** | -2.13*** |
| | (7.20) | (0.31) | (0.41) |
| Distance to nearest agrovet which sells fertilizer | 3.52 | 3.51*** | 0.09 |
| | (5.16) | (0.23) | (0.30) |
| Minimum travel-cost adjusted price for 50 kg of Urea[1] | 21.77 | 21.46*** | 2.01*** |
| | (3.07) | (0.12) | (0.15) |
| Distance to obtain minimum travel-cost adjusted price (km) | 9.30 | 8.94*** | 2.37*** |
| | (10.77) | (0.47) | (0.61) |
| Cost of travel to obtain minimum travel-cost adjusted price (USD) | 2.12 | 2.04*** | 0.54*** |
| | (2.46) | (0.11) | (0.14) |
| **Panel B. Output markets** | | | |
| Distance to the nearest output market (km) | 3.59 | 3.37*** | 1.46*** |
| | (5.24) | (0.23) | (0.30) |
| *Maize prices (120 kg bag of maize)* | | | |
| Market survey: maximum travel-cost adjusted price immediately after 2017 harvest (USD) | 40.52 | 40.45*** | 0.44 |
| | (6.03) | (0.27) | (0.35) |
| Farmer surveys: average "going price" in local village immediately after 2016 harvest[2] | 29.24 | 29.74*** | -1.16 |
| | (10.02) | (1.20) | (1.87) |
| Farmer surveys: average sales price among farmers after 2016 harvest[2] | 30.74 | 30.68*** | -0.11 |
| | (4.33) | (0.50) | (0.67) |

Notes: Data is from the universe of villages in Kilimanjaro region (N = 570). The unit of observation is the village. Travel costs imputed from transport surveys and Google maps. In Column 1, standard deviations are in parentheses. In Columns 2-4, standard errors in parentheses. *, **, and *** indicate significance at 10%, 5%, and 1%.

[1]We assume farmers buy a 50 kg bag in one trip (enough for 1 acre), the modal amount observed in our data, and must incur the cost of 3 trips to the retailer (a round-trip for herself, plus a trip for the bag of fertilizer).

[2]Data is only available from the second set of farmer surveys (1,400 farmers in 85 villages).

**Table 5. Remoteness and fertilizer retailer sales, prices, and markups**

| | (1) | (2) | (3) |
|---|---|---|---|
| | Mean | Regression on log (travel time from regional hub) | |
| | | Constant | Coefficient on log (travel time) |
| **Panel A. Agrovet-level (N = 356)** | | | |
| Sells Urea fertilizer | 0.98 | 0.97*** | 0.00 |
| | | (0.01) | (0.01) |
| Sells DAP fertilizer | 0.62 | 0.60*** | 0.07* |
| | | (0.03) | (0.04) |
| Sells CAN fertilizer | 0.14 | 0.14*** | 0.01 |
| | | (0.02) | (0.03) |
| Sells NPK fertilizer | 0.12 | 0.10*** | 0.04* |
| | | (0.02) | (0.02) |
| Sells other types of fertilizer[1] | 0.34 | 0.40*** | -0.18*** |
| | | (0.03) | (0.03) |
| Total bags of fertilizer sold | 183.9 | 204.17*** | -59.04* |
| | (467.20) | (27.52) | (35.31) |
| **Panel B. Prices and markups (Agrovet-brand level, N = 568)** | | | |
| Retail price for 50 kilograms | 25.42 | | 0.94*** |
| | (5.46) | | (0.17) |
| Wholesale price for 50 kilograms | 21.65 | | 0.27*** |
| | (4.35) | | (0.10) |
| Cost of transport from wholesaler (per 50 kg) | 0.58 | | 0.29*** |
| | (0.60) | | (0.06) |
| Markup (percentage points)[2] | 14.48 | | 1.47** |
| | (10.96) | | (0.61) |

Notes: In Column 1, standard deviations are in parentheses. In Columns 2-4, standard errors in parentheses. Regressions in Panel B including product fixed effects. *, **, and *** indicate significance at 10%, 5%, and 1%.

[1]Other types of fertilizer include local varieties SA, Yara, and Minjingu.

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

**Table 6. Remoteness and input market access and adoption**

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | | Regression on log (travel time from regional hub) | | |
| | Mean | Constant | Coefficient on log (travel time) | |
| | | | No controls | Controls for soil and farmer characteristics |
| ***Panel A: Input usage*** | | | | |
| Used chemical fertilizer in 2017 long rains | 0.45 | 0.52*** | -0.22*** | -0.13** |
| | | (0.03) | (0.04) | (0.05) |
| Quantity of chemical fertilizer used | 22.55 | 28.71*** | -18.61*** | -9.79*** |
| | (36.10) | (2.31) | (2.90) | (3.45) |
| Used improved seeds in 2017 long rains | 0.72 | 0.74*** | -0.05* | -0.03 |
| | | (0.02) | (0.03) | (0.04) |
| Quantity of improved seeds used | 5.85 | 6.36*** | -1.14** | -0.54 |
| | (6.58) | (0.34) | (0.47) | (0.56) |
| ***Panel B: Conditional on usage*** | | | | |
| Total input expenditures (USD) | 43.87 | 46.58*** | -5.35 | -6.34 |
| | (53.97) | (3.65) | (5.18) | (0.56) |
| Distance traveled to agrovet (km) | 10.80 | 9.27*** | 4.53*** | 1.33 |
| | (18.35) | (0.90) | (1.44) | (1.82) |
| Cost of travel to agrovet (USD) | 2.35 | 2.25*** | -0.30 | -0.44 |
| | (3.32) | (0.16) | (0.25) | (0.36) |
| Price of urea (per 50 kg bag, USD) | 23.84 | 23.88*** | 1.36*** | 0.30 |
| | (3.81) | (0.31) | (0.43) | (0.63) |
| Price of improved seeds (per kg, USD) | 2.75 | 2.75*** | 0.26** | 0.41*** |
| | (0.87) | (0.05) | (0.12) | (0.15) |
| Ad-valorem equivalent trade cost | 0.10 | 0.09*** | -0.01 | -0.01 |
| | (0.24) | (0.01) | (0.02) | (0.02) |

Notes: N = 2,155 farmers in 193 villages. See text for sampling details. Standard deviation in parentheses in Column 1. Standard errors, clustered at village level, in parentheses in Columns 2-4. *, **, and *** indicate significance at 10%, 5%, and 1%.

**Table 7. Remoteness and output market access and sales**

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | | Regression on log (travel time from regional hub) | | |
| | Mean | Constant | Coefficient on log (travel time) | |
| | | | No controls | Controls for soil and farmer characteristics |
| Sold maize after 2016 long rains | 0.28 | 0.33*** | -0.12*** | -0.08* |
| | | (0.02) | (0.03) | (0.04) |
| Total quantity sold (kg) | 181.6 | 225.79*** | -91.88*** | -49.58 |
| | (373.60) | (20.83) | (19.30) | (32.85) |
| ***Sales to agents at home*** | | | | |
| Agent visited homestead | 0.30 | 0.36*** | -0.18*** | -0.14** |
| | | (0.04) | (0.05) | (0.06) |
| Sold maize to an agent after 2016 long rains | 0.14 | 0.17*** | -0.08*** | -0.06* |
| | | (0.02) | (0.02) | (0.03) |
| Quantity sold to agents (kg) | 93.34 | 112.75*** | -56.40*** | -54.52** |
| | (266.40) | (12.40) | (12.71) | (22.25) |
| ***Sales at market*** | | | | |
| Sold maize at a market after 2016 long rains | 0.06 | 0.07*** | -0.03** | -0.04** |
| | (0.01) | (0.01) | (0.01) | (0.02) |
| Quantity sold at market (kg) | 30.31 | 37.56*** | -16.87** | -21.78 |
| | (155.20) | (5.78) | (6.79) | (13.51) |
| If sold at market, travel cost (USD) | 6.92 | 7.19*** | -1.51 | 13.34** |
| | (1.28) | (1.31) | (1.53) | (4.91) |
| Ad-valorem equivalent trade cost | 0.05 | 0.05*** | 0.01 | 0.08 |
| | (0.05) | (0.01) | (0.01) | (0.05) |
| ***Purchases*** | | | | |
| Farmer ever buys maize | 0.44 | 0.41*** | 0.11*** | 0.13*** |
| | | (0.02) | (0.02) | (0.04) |
| Quantity purchased in typical year | 91.91 | 81.74*** | 42.02*** | 46.26*** |
| | (163.50) | (5.12) | (8.22) | (13.25) |

Notes: N = 2,155 farmers in 193 villages. See text for sampling details. Standard deviation in parentheses in Column 1. Standard errors, clustered at village level, in parentheses in Columns 2-4. *, **, and *** indicate significance at 10%, 5%, and 1%.

**Table 8. Multinomial logit of agrovet choice**

| | (1) | (2) |
|---|---|---|
| | Agrovet Chosen | |
| Kilometers to agrovet | -0.177*** | |
| | (0.014) | |
| Dummies for agrovet distance bin: | | |
| between 1 and 5 km | | -0.738* |
| | | (0.433) |
| between 5 and 10 km | | -1.579*** |
| | | (0.479) |
| between 10 and 15 km | | -2.617*** |
| | | (0.501) |
| between 15 and 20 km | | -4.748*** |
| | | (0.618) |
| between 20 and 30 km | | -5.109*** |
| | | (0.608) |
| between 30 and 40 km | | -6.337*** |
| | | (0.672) |
| between 40 and 50 km | | -9.376*** |
| | | (1.206) |
| between 50 and 100 km | | -8.776*** |
| | | (0.852) |
| over 100 km | | -12.532*** |
| | | (1.505) |

Notes: Regressions also control for an indicator for whether the agrovet is located within the same administrative area of the village. Distances are calculated via Google maps. Standard errors in parentheses. *, **, and *** indicate significance at 10%, 5%, and 1%.

**Figure 1. Map of Survey Region and Villages**

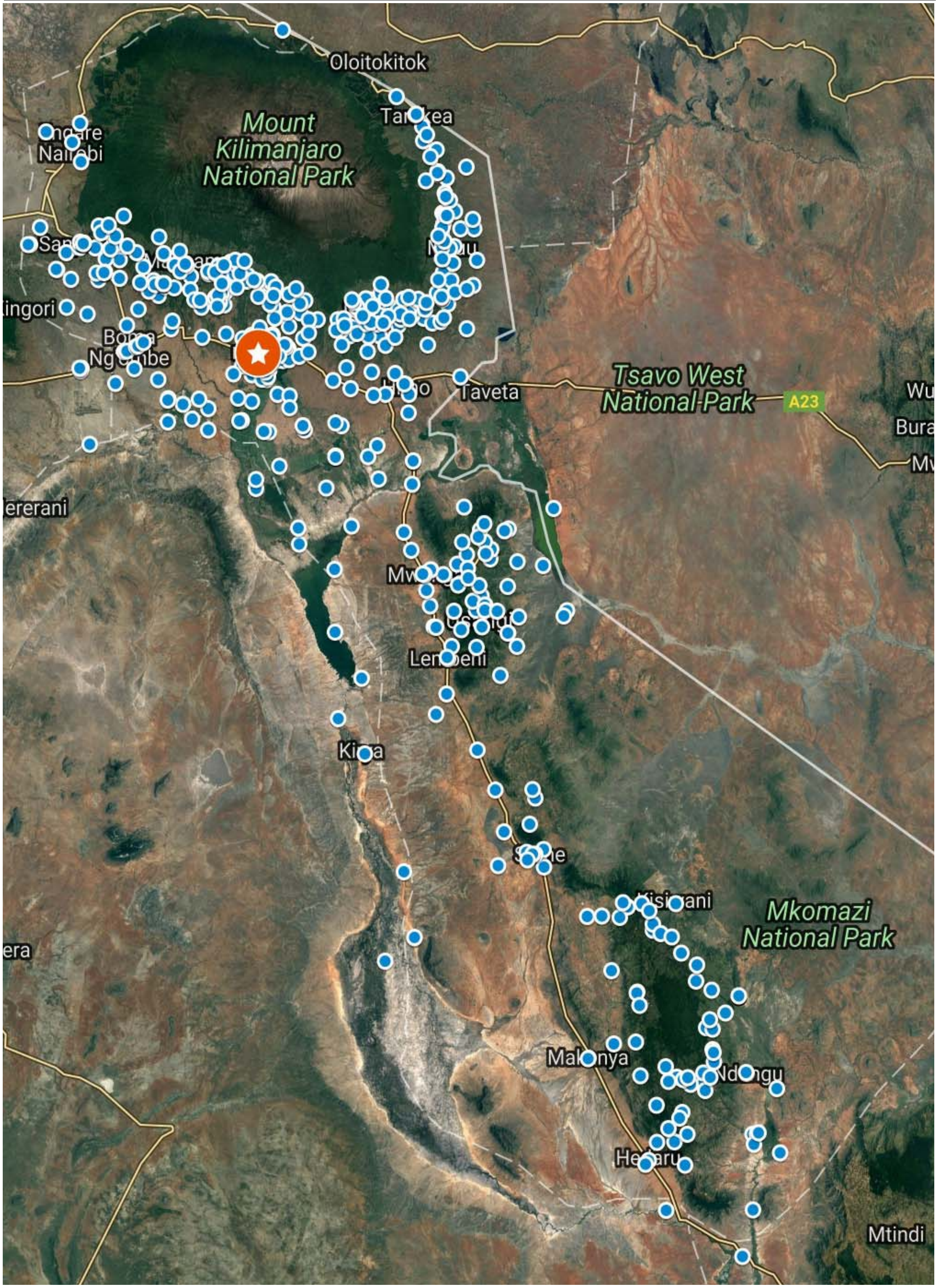
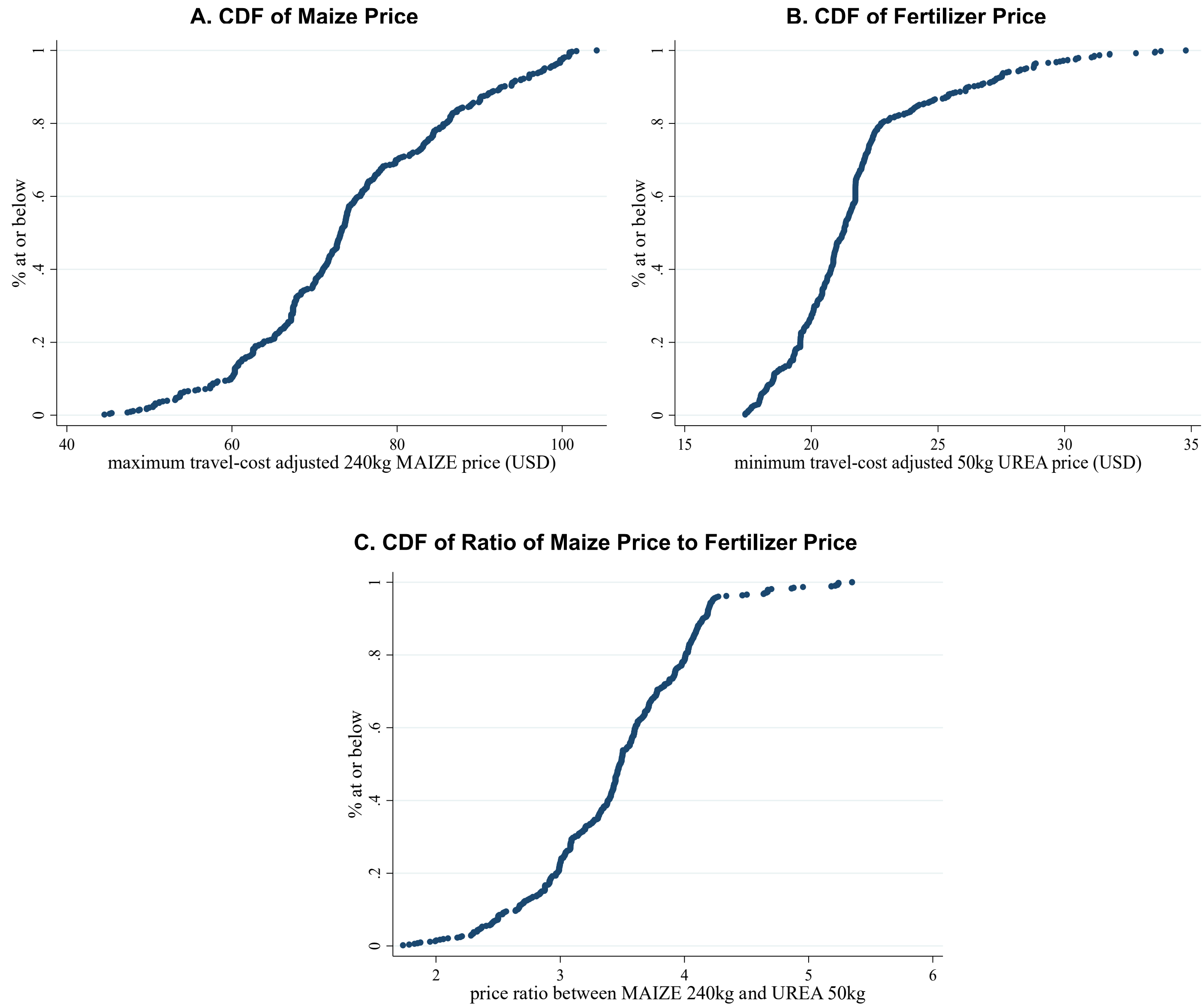
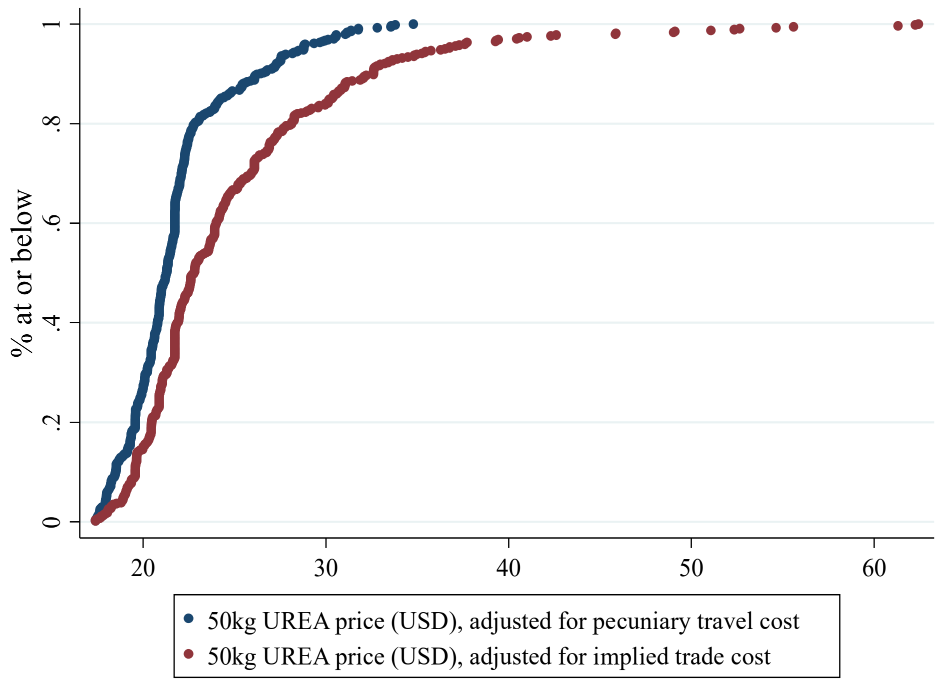


**Figure 2. CDF of travel-cost adjusted prices across villages**



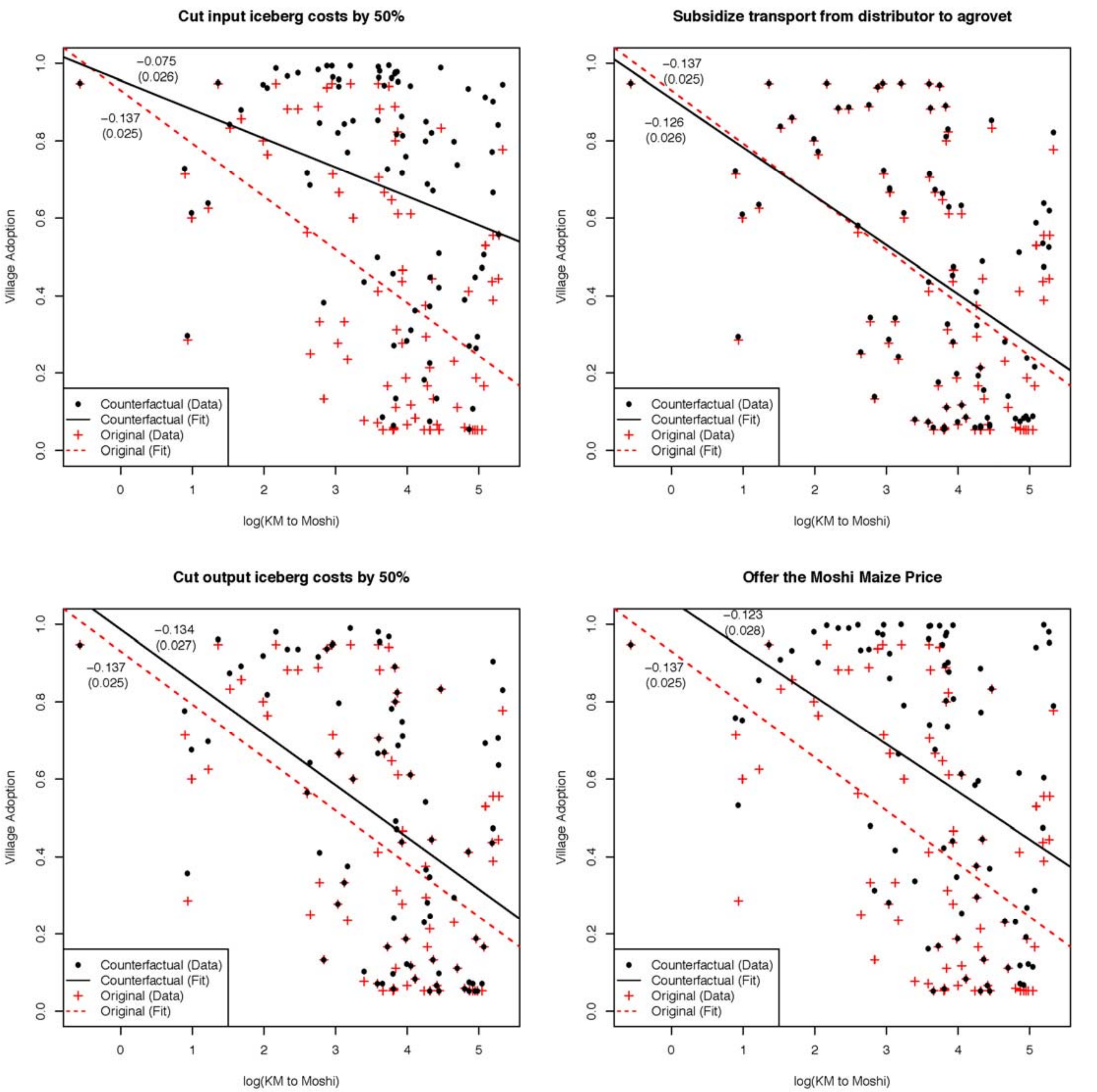

Notes: Each observation represents a village. Travel-cost adjusted prices are calculated through observed prices from an agrovet survey, a maize price survey at markets and transport cost information collected from interviews with transport operators.

**Figure 3. CDF of travel-cost adjusted prices (measured costs vs. estimated costs)**




Notes: Each observation represents a village. Travel-cost adjusted prices are calculated through observed prices from an agrovet survey, a maize price survey at markets and transport cost information collected from interviews with transport operators.

**Figure 4. Input Access Counterfactuals**




Notes: See text for discussion of counterfactuals.

**Web Appendix Table A1. Correlations in survey measures of travel time and Google maps**

| | (1) | (2) | (3) |
|---|---|---|---|
| | Surveys | Google maps | |
| | Cost of travel | Hours | Kilometers |
| **Panel A. Distance to Agrovets** | | | |
| Hours to agrovet reported by farmers who purchased inputs (from home to agrovet) | 0.267*** | 0.398*** | 0.399*** |
| Cost of travel reported by farmers who purchased inputs (from home to agrovet) | - | 0.267*** | 0.282*** |
| Google maps: hours from village center to agrovet | - | - | 0.969*** |
| **Panel B. Distances between markets and regional hub (Moshi)** | | | |
| Hours to Moshi reported by transportation operators | 0.748*** | 0.886*** | 0.850*** |
| Cost of travel to Moshi reported by transportation operators | - | 0.829*** | 0.833*** |
| Google maps: hours from market to Moshi | - | - | 0.976*** |
| **Panel C. Distances between villages and regional hub (Moshi)** | | | |
| Hours to Moshi reported by transportation operators | 0.672*** | 0.806*** | 0.788*** |
| Cost of travel to Moshi reported by transportation operators | - | 0.771*** | 0.795*** |
| Google maps: hours from market to Moshi | - | - | 0.976*** |

Notes: N = 318 farmers for Panel A. N = 562 villages for Panel B and C. In Panel A, google distances are calculated from the village commercial center and thus does not include distance from the farmer's home to the village center.
*, **, and *** indicate significance at 10%, 5%, and 1%.

**Web Appendix Table A2. Input and output market price dispersion across countries**

| | (1) Secondary Datasets[1] | (2) Tanzania Data[2] |
|---|---|---|
| Residual standard deviation in log prices for:[3] | | |
| All products | 0.45 | 0.22 |
| Maize only | 0.34 | 0.14 |
| Fertilizer only | 0.12 | 0.09 |

Notes:

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

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

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

**Web Appendix Table A3. Dyadic price dispersion**

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

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

**Web Appendix Table A4. Adoption in LSMS-ISA surveys**

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

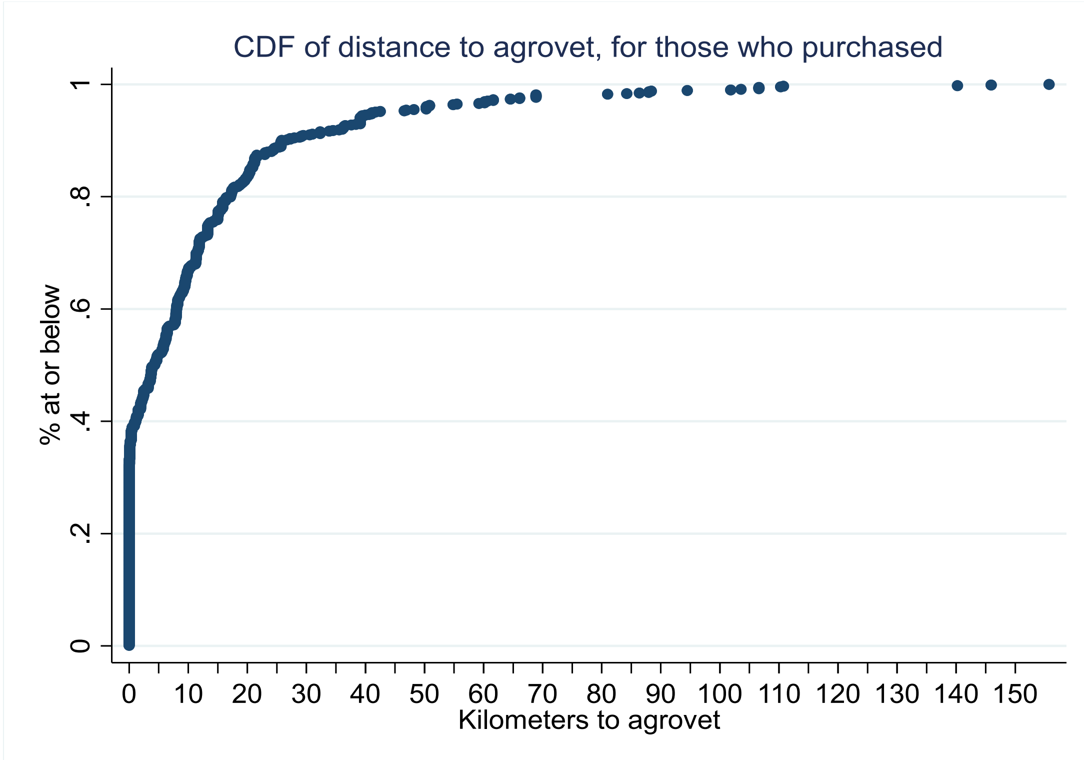
Notes: Regressions include World Bank LSMS-ISA household panel surveys in Ethiopia, Niger, Nigeria, Malawi, Tanzania, and Uganda. Standard errors clustered at the enumeration area level are in parentheses.

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

**Web Appendix Figure A1. Retail prices, wholesale prices, and markups for Urea**


Notes: Prices are for a 50kg bag of fertilizer. Urea is the most commonly used variety of fertilizer in Northern Tanzania. Markup is computed by "[retail price / (wholesale price + travel cost)] - 1."

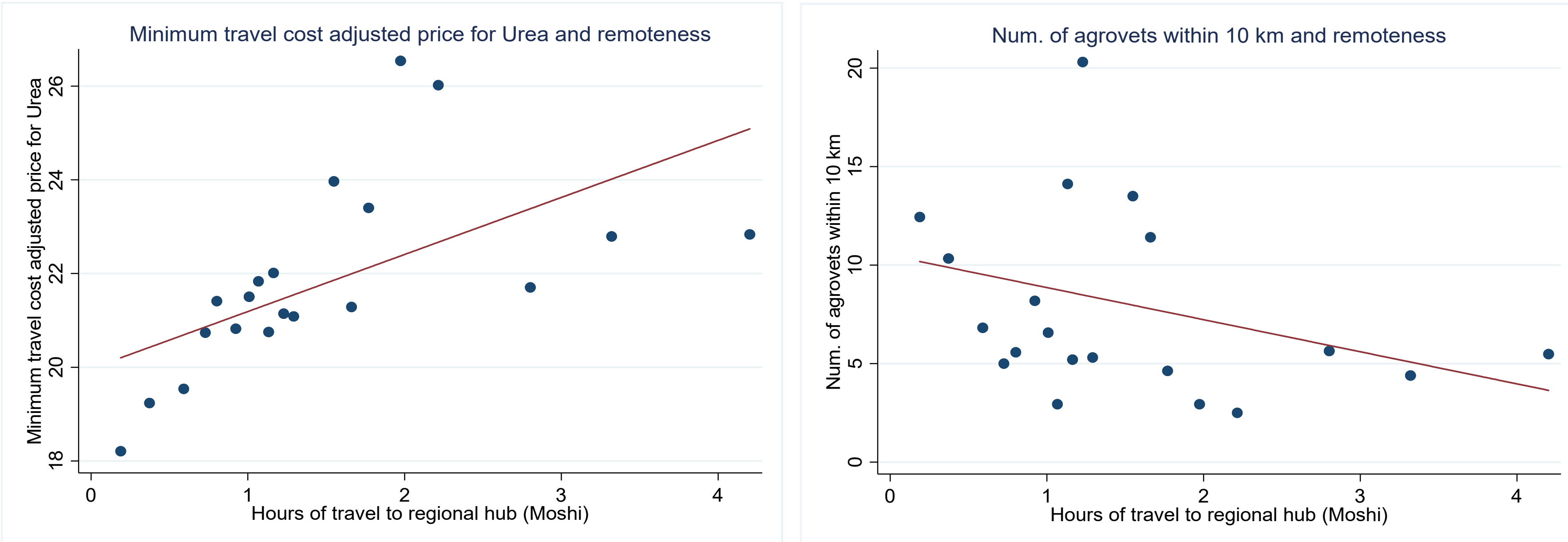
**Web Appendix Figure A2. CDF of distance farmers travel to purchase inputs**



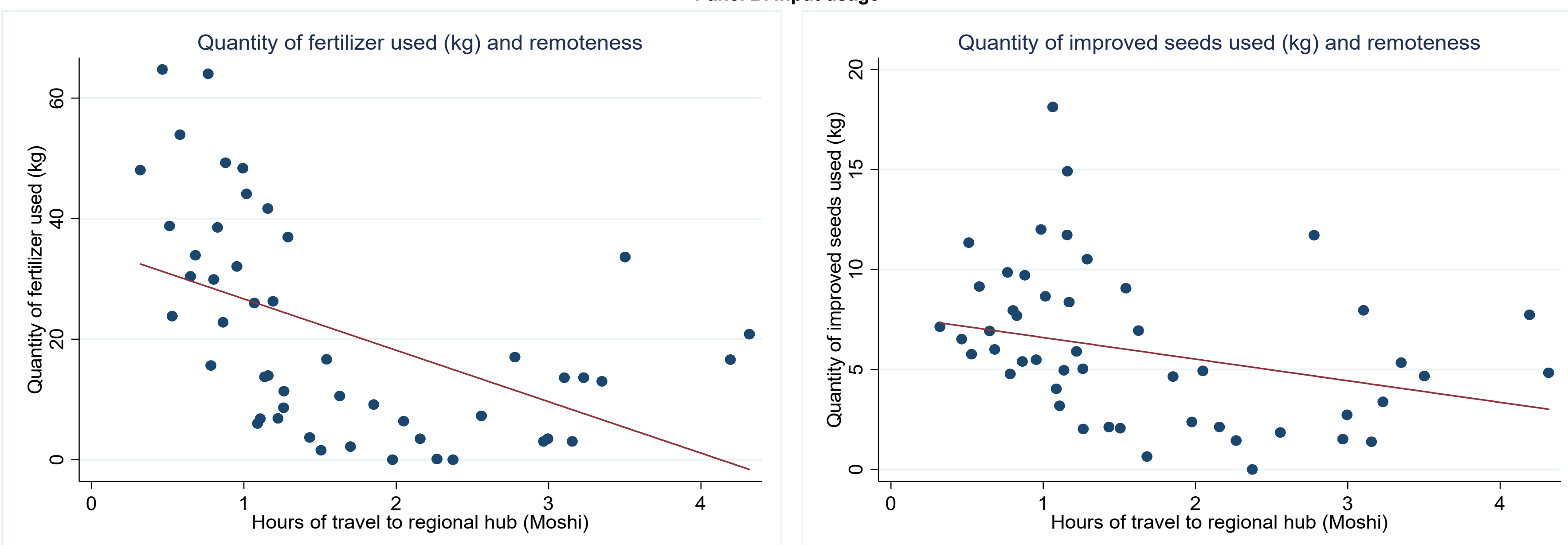

Notes: Each point represents a farmer.

**Web Appendix Figure A3. Graphical relationship of access to markets and input usage with respect to distance to regional hub**

**Panel A. Access to input markets**



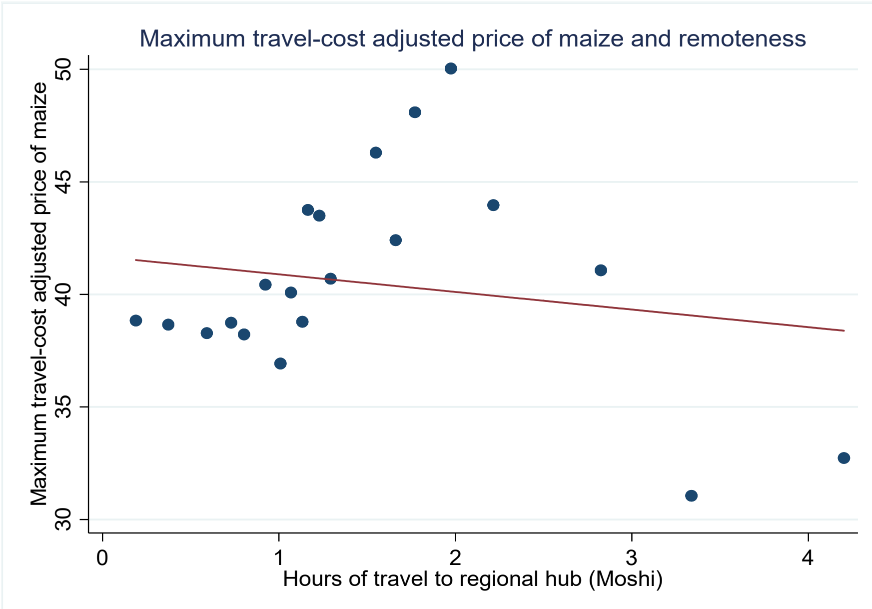

**Panel B. Input usage**

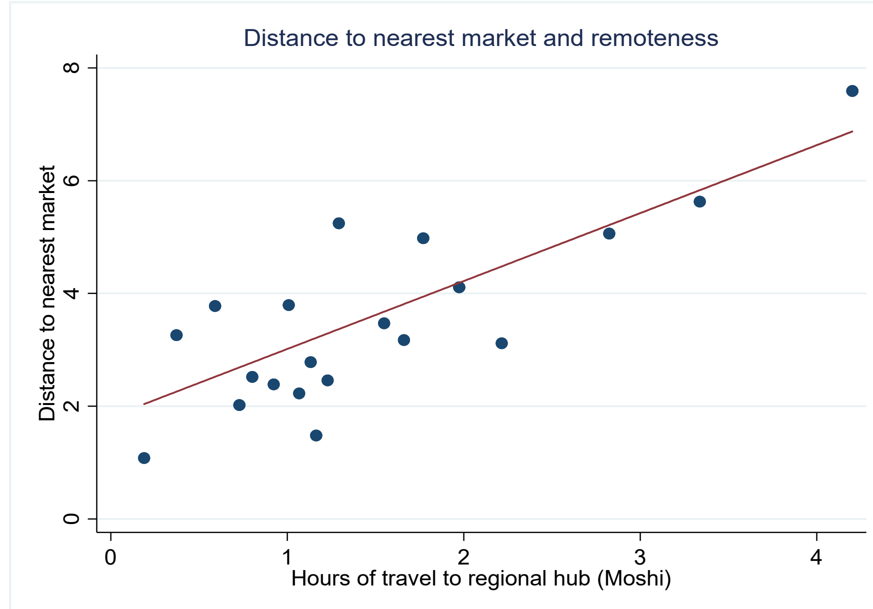



Notes: The unit of observation is the village in Panel A and farmers in Panel B.

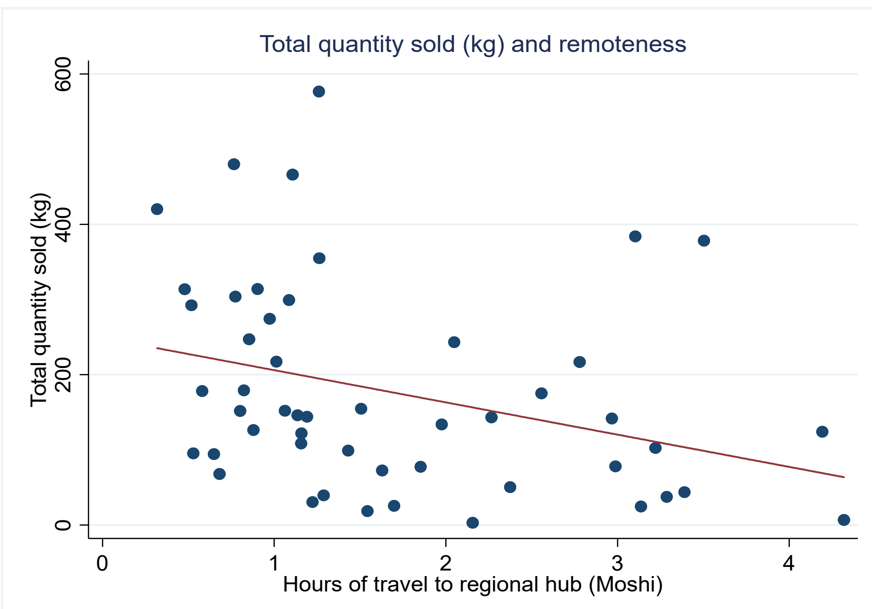
**Web Appendix Figure A4. Graphical relationship of access to markets and sales of output with respect to distance to regional hub**

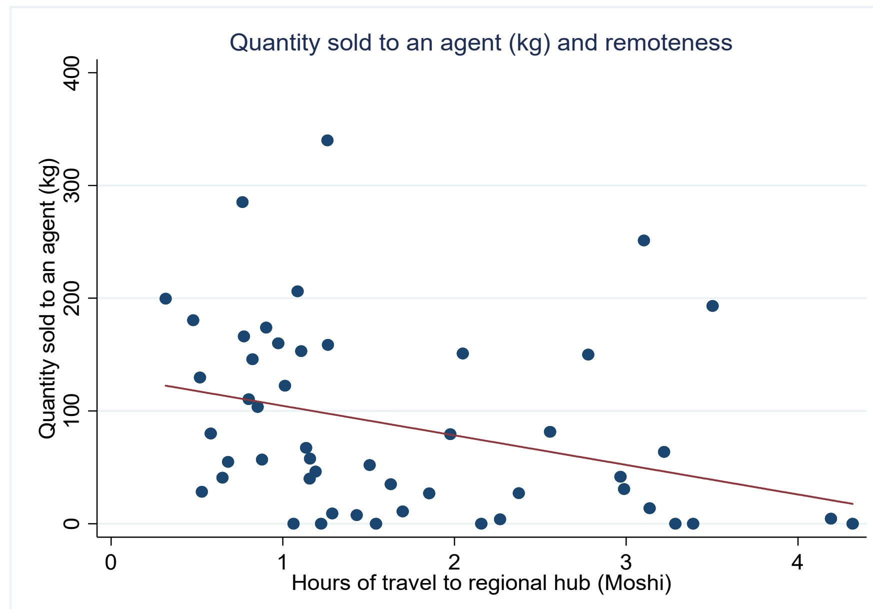
**Panel A. Access to output markets**



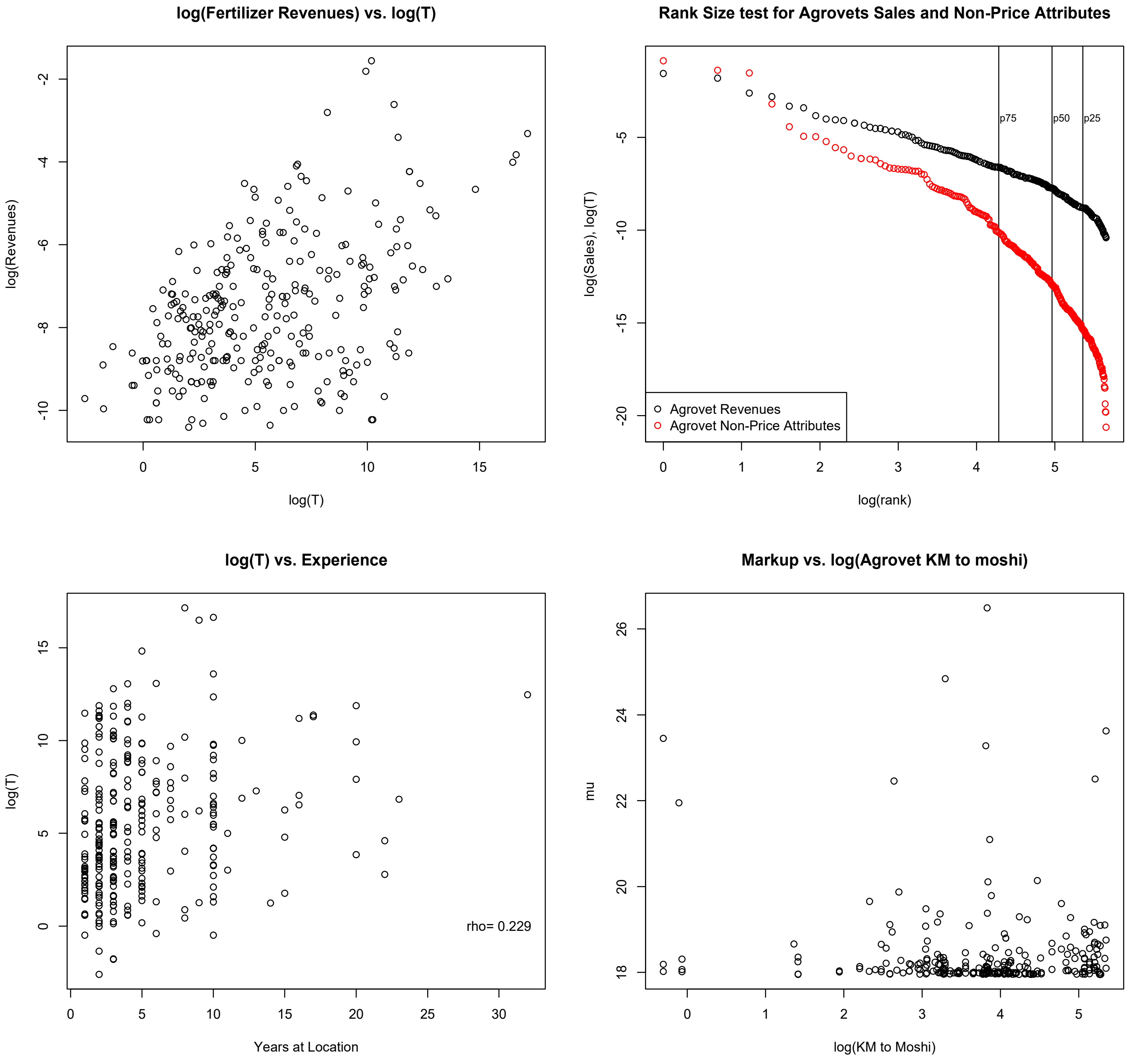

**Panel B. Sales**




Notes: The unit of observation is the village in Panel A and farmers in Panel B.

## Web Appendix Figure A5. Agrovet Quality Estimates




Notes: See text for estimation of T. Years at location measured from the Agrovet survey. Mark-ups estimated via Berry (1994) inverstion using estimated market-shares