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

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

February 18, 2019

Abstract

In this paper, we quantify market access in rural Tanzania, and the extent to which it constrains agricultural productivity. We collect granular data on farmer input and sales decisions, input and output prices, and travel costs in all 1,183 villages in two regions of Tanzania. We find that a village in the 90th percentile of the travel-cost adjusted price distribution faces input and output prices 40-55% less favorable than a village at the 10th percentile. In reduced form, an additional standard deviation of travel time is associated with 20-25% lower input adoption and output sales. We develop and quantify a spatial model of input adoption and conservatively estimate that farmers behave as if they face local travel costs of 5.7% ad-valorem per kilometer of travel, which is equivalent to 45% when traveling to the closest retailer. Holding exogenous local factors fixed, we estimate that reducing travel costs by 50% (approximately the effect of paving rural roads) would double adoption and reduces the adoption-remoteness gradient by 15%.

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 Faure, Petra Gram, Matthew Krupoff and Alexandra Murphy provided excellent research assistance. We thank Lauren Falcao Bergquist, Shoumitro Chaterjee, Alain de Janvry, Betty Sadoulet and seminar participants at Ashoka University, Cal Poly San Luis Obispo, UC Berkeley, ISB, LSE, the University of Maryland, NBER Trade and Agriculture Conference, NEUDC, the University of Virginia, UCSC, the University of West Virginia, the WEAI, the University of Pittsburgh, Stanford University, the University of Colorado, and Johns Hopkins SAIS for helpful comments. We acknowledge funding from the IGC (grant 1-VRC-VTZA-VXXXX-89307), the National Science Foundation (SES 1658942), Innovations for Poverty Action, ISB and UCSC. 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.

In this paper, we rigorously document market access among farmers in two regions – Kilimanjaro and Manyara – of Northern Tanzania. Together, these two regions represent 6 percent of the land area and population of the country. Our data collection exercise spans the entire supply chain of maize (output) as well as of fertilizer (input) in all 1,183 villages in these two regions, including (1) surveys with a random sample of 2,845 farmers in 246 randomly selected villages; (2) surveys with 532 agro-input retailers ("agrovets") that sell fertilizer (this sample represents the universe of retail locations available to farmers);[2] (3) a retrospective panel of buying and selling prices of maize from a randomly selected sample of maize-sellers in each of the 226 markets in the area; (4) collection of information on road quality, travel times, and travel costs to all villages from their respective local markets as well as from 5 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 across central and northern Tanzania.

We make three main contributions. First, we precisely document spatial price dispersion for input and output prices, inclusive of trade costs. To do this, we use our extensive travel cost data to estimate travel costs to every destination, and then take the most favorable prices for farmers. We find clear evidence of large and economically meaningful spatial heterogeneity in both input and output prices. For both, we find that the price difference between the 90th and the 10th percentile of delivered input prices is equivalent to about 50% of the mean. We find similar evidence of greatly reduced market access in remote rural areas using a variety of other measures.

Second, we conduct a reduced-form investigation of the correlation between usage and remoteness on the input side, and sales and remoteness on the output side, where remoteness of any location is proxied by two measures: the weighted average of distance from a set of 5 major urban centers, and an elasticity-weighted trade cost measure of remoteness from the same hubs (calculated analogously to Donaldson and Hornbeck, 2016). Using both measures, we find consistent and striking correlations between remoteness of villages and all measures of input usage and output sales. In particular, we find that a standard deviation increase in remoteness is associated with a 9-17 percentage point reduction in the probability of using fertilizer and a 6-9 percentage point

[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.

[2] As discussed later, we successfully surveyed 532 agrovet out of a universe of 585. For the remaining agrovets, we impute prices as described later.

reduction in the probability of selling maize. Put differently, we find that input usage in the most remote villages in our study sample is approximately only a third of that in the least remote villages, while maize sales are only 45 percent as high.

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

On the input side, the structure of the model (which is similar to Eaton and Kortum, 2002) facilitates a decomposition of choosing an agrovet into three components: (1) the decision whether to adopt anywhere; (2), the decision of which location to buy from; and (3) the decision of which retailer to pick within that location. Our novel farmer surveys reliably record the first two of these decisions, and thus allow us to calibrate (1) local factors that may affect adoption; and (2) the implied trade costs incurred while sourcing from each agrovet location. 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 (which may represent unmodeled factors like experience in selling inputs and likelihood of stockouts). The results suggest that transportation costs are large: our preferred specification yields estimates of local iceberg costs that are approximately 5.7% ad-valorem per kilometer of travel. This is approximately 45% to reach the average closest agrovet, and thus these costs are economically meaningful, and suggest significant non-pecuniary costs of travel. These may include other factors such as the opportunity cost of the time to travel, risk-aversion related to potential stockouts, or information frictions, among others. After estimating trade costs, we use the model to build a market-clearing condition for fertilizer for each agrovet, which is a function of the expected spatial distribution of fertilizer expenditures by each farmer and the probability that a farmer at each location adopts at a given agrovet. We balance these market clearing conditions by finding a vector of agrovet "amenities" that exactly rationalize the market-shares of each agrovet. After doing so, we are able to calculate a precise measure of market access for fertilizer (in the spirit of Redding and Venables 2004, Redding and Sturm 2007, Head and Mayer 2011, and Donaldson and Hornbeck 2016). The estimated measure of market access falls approximately 50% per standard deviation increase in remoteness.

Finally, we use the estimated parameters from the model to simulate market access counterfactuals. For input market access, our primary counterfactual is reducing trade costs incurred to reach retailers by 50%, which is roughly equivalent to the realized reduction in travel time due to road upgrading (Casaburi, Glennerster and Suri 2013). This policy roughly doubles adoption relative to baseline, and also reduces the distance gradient by 15%. When evaluating the effect on fertilizer expenditures, the gradient falls by about 50%. We also evaluate how the costs for retailers to source inputs from distributors affect the adoption decision. After cutting the wholesaler-retailer transportation costs by half, adoption rises by about 1 percentage point, or 4%, and yields a 4% reduction in the remoteness-adoption gradient. We also study hypothetical entry counterfactuals, where we find that while agrovet entry in remote areas has a larger effect on adoption, the profitability from doing so falls by 33 log points for each standard deviation increase in remoteness. Finally, subject to a number of caveats based on data availability for maize markets, we reduce transportation costs by 50% for reaching output markets, and find a slightly smaller change in adoption, though a slightly larger reduction in the adoption-remoteness gradient (when compared with input market access).

This paper sits at the intersection of trade and development economics. On the development side, our paper contributes to a literature examining why Sub-Saharan Africa has lagged behind the rest of the developing world in agricultural technology adoption (World Bank 2008). Many studies find evidence of large *yield* increases of using improved inputs (i.e. Duflo, Kremer and Robinson 2008; Beaman et al. 2013; Stewart et al. 2005; Udry and Anagol 2006), though the evidence is much more mixed on whether using these inputs is *profitable* (i.e. Duflo, Kremer and Robinson 2008; Beaman et al. 2013). Our results quantify the extent to which profitability, and thus adoption, will tend to be lower in more remote locations, due to less favorable input and output prices for farmers. In this sense, our work is closely related to Suri (2011), who shows that many Kenyan farmers with high gross returns to hybrid seeds choose not to adopt them because the fixed costs of obtaining seeds are too high, presumably due to travel costs. Our paper is differentiated by focusing on heterogeneity in market access, rather than on heterogeneity in returns. Related work in Minten, Koru, and Stifel (2013) also focuses on remoteness and profitability and documents significant farmer-to-retailer transaction costs to reach price-controlled input cooperatives in a rugged region in northern Ethiopia.[3]

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] which includes a host of outcomes other than just prices, including consumption diversity, farm investments, human capital investment, migration, and occupational choice (Aggarwal, 2018a; b; Adukia et al., 2016; Asher and Novosad 2016; Brooks and Donovan, 2017; Morten and Oliveira 2016). In our paper, we focus narrowly on the specific effect of transportation costs on market

[3]Specifically, the authors document the farmer-reported cost of renting cargo transport (predominantly a donkey in this region) and the time cost of travel for a trip to the market town along the only route of egress from their villages. Our study differs in its focus on access to all, privately owned intermediaries for inputs, the costs of transport along both rural and feeder roads, and also the quantification of the adoption decision through the lens of a spatial economic model.

[4]A partial listing of papers includes Aggarwal (2018a), 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.

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]

Our work is also related to a voluminous trade literature which attributes price differentials across space to three primary components – marginal trade costs (e.g. Donaldson, 2018; Eaton and Kortum, 2002; Keller and Shiue, 2007; Sotelo, 2018), 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). Our paper is most closely related to Atkin and Donaldson (2015), who estimate trade costs in a setting where an intermediary buys products at wholesale prices, transports them to distant markets and sells directly to consumers. By contrast, we are interested in how trade costs affect the buying decisions of final consumers (in this case, farmers) through the farmer access to retailers and output markets. Other general equilibrium trade models also assess the link between trade costs, price gaps and technology adoption. Porteous (2017) and Porteous (2019) assess the implications of intra- and international trade costs on output prices and technology adoption in Africa. The granularity of our analysis, and the focus on the farmer-retailer link, distinguishes our work. For example, Porteous (2017) focuses on trade between larger cities and markets, likely though lorry-freight transport, estimating a median freight cost of $0.287 per ton-km. Our local transport cost surveys, which are conducted with motorbike and van taxis, indicate trade costs which are an order of magnitude larger ($4.74 per ton-km). The associated ad-valorem equivalent trade costs in traveling to an agricultural retailer are 22-45%. These costs are larger than estimates in Atkin and Donaldson (2015), which are calculated over a much longer trip from source to consumer.[6]

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 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 on fertilizer market and study regions

This study took place in the Kilimanjaro and Manyara regions[7] of Northern Tanzania. The two regions are a combined 57,000 square-kilometers (6% of the land mass of Tanzania), contain 1,183 villages, and had a population of 3.1 million in 2012 (National Bureau of Statistics, 2013). Compared

[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, and which will be ready by 2020. Drones capable of transporting cargo of up to 20 kilos over a distance of 100 kms already exist.

[6] Specifically, ad-valorem estimates of 10-20% 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.

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

to developed countries, the quality of roads in Kilimanjaro and Manyara is very poor: for example, the paved trunk road density is 2.2 percent in Kilimanjaro (i.e. there are 2.2 kilometers of paved roads per 100 square kilometers of area), 0.15 percent in Manyara, and 0.7 percent in Tanzania overall (TanRoads and PMO-RALG, 2014), compared to 68 percent in the US and an OECD average of 134 percent[8]

The main crop grown in this area is maize. There are two growing seasons in this area: 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.

As in much of Sub-Saharan Africa, production capacity of fertilizer is virtually non-existent in Tanzania and almost all of what is used is imported via the port at Dar es Salaam (FAOSTAT Online database, 2016; Hernandez and Torero, 2011), and then transported throughout the country over surface roads, including to the study regions. In all of these respects, the study area is fairly similar to other countries throughout East Africa that predominantly grow maize and import fertilizer, such as Kenya, and perhaps a little bit better than landlocked countries such as Malawi and Uganda, that can receive fertilizer only after it has traversed the distance between a neighboring coastal nation's port and their shared border, and then must travel further inland to reach farmers.

3 Data and summary statistics

To construct our sample, we first assigned every village in the two regions 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 each village frequented. We use this market information in two main ways. First, we use this market designation as the unit over which to measure the price of maize, which is commonly transacted in such markets. Second, as it was not feasible to travel individually from every village to every possible destination to measure transport costs, we instead consider the market catchment area as the appropriate geography with which to measure transport costs. In particular, for travel *outside the market catchment area* (i.e. to retailers outside the market area and to regional hubs), we require routes to go through market centers – we measure distances from every village to its closest market, and from every market to the main road. This last step is often very low cost, since markets are typically located on or near the main road. For travel *inside the market catchment area*, we allow travel via routes that do not involve the market itself. A map of the villages in our sample is included as Figure 1.

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

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

2017 in Kilimanjaro and February to May 2018 in Manyara.[9]

3.1 Agrovet surveys

We conducted a census of all agricultural input retailers (known as "agrovets" locally) in the region, finding a total of 585 that sold either fertilizer or seeds. We then revisited these agrovets to conduct a longer survey which took about 2 hours to complete. Of these 585, we did surveys with 532 of them (see Web Appendix Table A1, which reports survey compliance and attrition). The survey asked questions about varieties of fertilizer sold, and their prices, quantities, and the wholesale costs of acquiring stock from the distributor. The survey took care to differentiate fertilizer varieties by distributor, brand, and type – thus the level of granularity should be akin to the barcode-level. The survey also included a number of questions about costs of travel to the distributor, as well as some background characteristics.

3.2 Farmer surveys

We conducted farmer surveys in 246 randomly selected villages in three waves. The first wave occurred in 115 villages in Kilimanjaro in early 2016, the second wave in 97 villages in Kilimanjaro in 2017, and the third wave in 50 villages in Manyara in 2018. The surveys included questions on input usage and prices, transport costs and agrovet choice, maize sales, harvest output, and a series of household and demographic questions. Though the exact questions varied from survey to survey, the general format was very similar across rounds. The main difference across rounds was the sampling procedure and the number of farmers enrolled per village: in round 1, households were selected from a random walk procedure[10], while in rounds 2-3 households were pre-identified from a village listing exercise conducted with local leaders. In Wave 1, we sampled only 5 households per village for budget reasons, while in Waves 2-3 we selected 18 households per village. While this differing sampling procedure could result in differential selection, we find no qualitative difference in results from the two methods, and thus we pool all surveys together in the analysis.[11]

3.3 Measuring transport costs

One of the primary contributions of this work is to carefully document transport costs incurred by farmers. We measured transportation costs in several ways. First, we collected the GPS location for

[9] 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.

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

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

every village in both the study regions,[12] from which we calculated driving times and distances using the Google API (via the statistical program R). Second, we conducted surveys of transportation operators in every village in our sample, which were either motorbike taxis ("Boda Bodas"), or consumer van taxis ("Dala Dalas"). In each village, we asked up to 3 operators how much it cost to travel to the major towns in Kilimanjaro (Arusha and Moshi), the capital city (Dar es Salaam), and importantly, the market center as defined for the sampling procedure. In Manyara, in addition to Arusha, Moshi and Dar es Salaam, we also asked about trip costs and times to Babati, Dodoma, and Tanga.

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

3.4 Maize prices

To measure maize prices, we visited markets post-harvest in September and October of 2017 in Kilimanjaro and February to May of 2018 in Manyara. During these visits, enumerators sampled up to 3 maize sellers per market 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 on villages

A map of Kilimanjaro and Manyara is shown in Figure 1. We surveyed all villages in these regions. Summary statistics on villages are provided in Table 1. The average village has 480 households (see table notes), and is located 6.5 kilometers from the nearest market center. It takes about 40 minutes by vehicle to reach the market and return, and a round-trip costs about $1.92 on average. The average village is over 70 km away from the nearest major hub, and travel there would be an almost 3 hours-long round-trip, and cost $6. For both measures, there are many villages that are much further away than this - the standard deviation of both is large.

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

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

4 Results

We now use our surveys to examine dispersion in input and output prices and other outcomes, and any relationship with the remoteness of each village in our study regions. We begin by documenting price dispersion in "best" input and output prices, after accounting for travel costs.

4.1 Travel cost-adjusted price dispersion

For each village, we assume that farmers are free to travel to any agrovet/market to buy inputs or sell output, but must incur a transportation cost, which we calibrate using information from transport surveys and google distances. Specifically, using Google API, we calculate the route from every village to every agrovet. This route will involve either (1) traveling only on local roads over a relatively short distance, or (2) using local roads to connect to trunk roads. We calibrate the costs of local and trunk roads using our transport operator surveys, and information collected by enumerators during their own travel. We present these results in Table 2. Columns 1-3 show the costs of traveling from market centers to hub towns, which involves primarily traveling on trunk roads. We find a cost of about \$0.021 per km, or \$1.26 per hour of travel. The remaining columns replicate this on rural roads, where in columns 7 and 8 we find higher costs for rural travel: \$0.088 per km, or \$2.61 per hour of travel.

Calculating the cost of travels from any village to any agrovet/market requires the cost data for two separate legs: (i) from village to its primary market on a main road and (ii) from the main road turnoff to the destination main road turn off. For the trip between village and its nearest main road, we measured the precise cost by having our enumerators to travel from every village to its main road turnoff. The travel cost between the origin main road turnoff and the destination main road turnoff is imputed by the travel cost on a paved road collected from our transport operator survey. On the other hand, when two villages are in the same market catchment area, farmers could go to the destination directly without going through the main road. For this case, we imputed the travel costs using the direct google distance between the two villages and the average travel cost per km on a rural road because we do not have the measured travel cost between the villages.

With these, we calculate a travel cost-adjusted price of fertilizer for every village as follows:

$$r_v^{tc} = \min_j \{r_j + c_{jv}\} \tag{1}$$

where r_j is the price at agrovet j and c_{jv} is the cost of traveling to an agrovet, and returning with a bag of fertilizer from agrovet j to village v. Farmers must therefore make a round-trip for themselves, and a one-way trip for the bag of fertilizer. To calibrate these costs, we use survey questions which asked those farmers who traveled to market about travel costs for themselves and the fertilizer (Web Appendix Table A2). We do this for a 50 kg bag of fertilizer, the modal amount purchased by farmers. We find that transporting a 50 kg bag of fertilizer costs about 69% as much as transporting a person for the same amount of time, implying therefore that a farmer must make 2.69 trips to buy a bag (2 for the farmerand 0.69 for the bag).

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

$$p_v^{tc} = \max_m \{p_m - c_{mv}\} \tag{2}$$

Here, p_m is the price of maize post-harvest for market m, and c_{mv} is the cost of traveling from village v to market m. We use a 120 kg bag for this calculation, and assume that the cost of transporting the bag is proportional to the weight. Thus, a trip to the market and back to sell 120kg of maize requires 3.7 trips (2 for the farmer and 1.7 for the bag).

We calculate these prices for every village-agrovet and village-market pair. We then take the minimum input price and the maximum output price faced by each village. Using thes "best" prices for inputs and output, our first main result is that 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 shows tremendous heterogeneity in prices across villages. In Panel A, for maize prices, we observe that one standard deviation in the best travel cost-adjusted selling price of maize is 23% of the mean, and the difference of this price between the 90th and the 10th percentile is similar to about 54% of the mean. Panel B shows the distribution of transportation cost-adjusted fertilizer prices, where the standard deviation in travel cost-adjusted prices is 19% of the mean, and difference of this price between the 90th and the 10th percentile is equivalent to about 43% of the mean.

4.2 Reduced form analysis

Section 4.1 documented substantial price dispersion in best input and output options, as calculated using our agrovet, transport operator, and maize market surveys. In this subsection, we explore the relationship of these best prices to the remoteness of the village, along with other important outcomes like input usage, selling behaviour, and access to retailers and markets. We begin by outlining a general specification to study remoteness, and then describing two measures of remoteness that we take to the data.

4.2.1 Specification

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

$$m_{vt} = \beta_r \cdot R_v + \epsilon_{vt} \tag{3}$$

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

For the measures of village-level market conditions estimated in (3), we include no controls. However, for farmer outcomes such as input adoption, it is clear that usage will depend not only on market access but also other characteristics such as income and land suitability. Therefore, these are estimated at the farmer level as:

$$m_{fvt} = \beta_r \cdot R_v + \beta_X X_{fvt} + \epsilon_{fvt} \tag{4}$$

where subscript f refers to farmer and X_{fvt} is a vector of other controls. 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. All farmer-level results are presented both with and without these controls.

4.2.2 Defining remoteness

To measure the remoteness of each village v, we will focus its proximity to a select number "hubs" that are within or nearby the study regions: Arusha, Babati, Dodoma, Moshi, and Tanga. The locations of interest are those where distributors for both maize and fertilizer are commonly located. While we do not have a complete listing of distibutors for the country, we know from our agrovet surveys and maize store surveys that distributors are only located in major cities and towns. Web Appendix Table A3 presents information on where input and output distributors are located, showing that nearly all of them are located in the major towns of Arusha, Moshi, and Babati. We extend this list to also include the regionally important cities of Tanga and Dodoma, and our complete set of hubs are marked with "stars" in Figure 1.[13]

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

$$remoteness_v = \sum_h d_{hv} p_h \tag{5}$$

where p_h is the (relative) population of hub h (i.e. the population of that hub divided by the population of all hubs) and d_{hv} is distance from village v to that hub. This measure is similar to Donaldson and Hornbeck (2016) given that relative population is used a proxy for the importance of each city in terms of availability of goods and average prices. However, the use of distance simplifies the construction of the measure substantially, where it is unnnecessary to calculate the ad-valorem costs of travel and estimate a distance elasticity. Despite this simplification, we show in the technical appendix that there exists a first-order approximation that links the two measures. Further, we standardize $remoteness_v$ to be mean 0 and standard deviation 1. The distribution of this variable is illustrated in Web Appendix Figure A1. The least remote areas (near major hubs) are approximately -2 standard deviations from the mean, while the most remote are +3 standard

[13] Even though our surveys collected information on the time and distance to Dar es Salaam, we do not include Dar es Salaam while calculating the remoteness measure as the high relative population of the capital leads it to overwhelm all the other hubs in our measure.

deviations away. The difference between these (5 standard deviations) is useful for benchmarking differences in outcomes between the most and least remote areas (similar to the approach taken in Atkin and Donaldson, 2015).

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

$$MA_v = \sum_h \tau_{hv}^{-\theta} p_h \tag{6}$$

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

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

where $cost_{hv}$ is the estimated cost to get from village v to hub h, 2.69 is the number of one-way trips required to travel to a destination and return back with a 50kg bag of fertilizer (see section 4.1), and *avgprice* is the average price of fertilizer in the sample (as measured by agrovets). We choose the average fertilizer price as a benchmark good to measure ad-valorem costs, since it is the focus of the paper, and also because agrovets commonly report traveling to hubs to get their fertilizer for inventories. To measure the elasticity term, $-\theta$, we appeal to estimation later in the paper where the substitution elastiticy across agro-retailers is estimated to be appoximately -5, though the results are robust to other estimates (eg. -8, as in Donaldson and Hornbeck).

Finally, to facilitate comparability with the standardized measure of population-weighted average distance, we standardize MA_v to be mean 0 and standard deviation 1, and also put a negative sign in front it, redefining the measure as standardized remoteness through market access. The distribution of this measure remoteness is also presented in Web Appendix Figure A1, where like the simple weighted distance measure, there are two modes of remoteness. The two measures are highly correlated – see Web Appendix Table A4.

4.2.3 Summary statistics and correlations with remoteness

Table 3 shows some statistics on demographic and background characteristics, buying and selling of maize, production capacity and harvest output, and displays how these variables vary with remoteness. From Panel A, we see a number of differences: farmers in more remote areas are less educated, own fewer assets, have less access to finance, and earn less income from sources outside of farming. These farmers also tend to have larger families and larger farms.

Panel B shows production capacity, based on GIS data from the FAO-GAEZ database, which provides information on counterfactual yields with and without inputs. The key concern here would be that more remote areas are less suitable for fertilizer, for whatever reason. At the mean, the FAO estimates that using inputs would more than quadruple yields. We do find that the yield increase is lower in remote areas – in villages one standard deviation away from the mean, yields using inputs are 260% of those without inputs, compared to over 300% at the mean. Comparing the most and least remote areas according to the weighted distance measure(roughly from -2 to 3 standard deviations), we estimate a yield increase of over 500% in the least remote areas compared to 200% in the most remote. Thus, while we should expect lower usage in remote areas since returns are lower, the yield increases are still extremely large everywhere. We control for these measures in our main regressions.

Finally, Panel C shows harvest output from the most recent long rains. While the relationship between total yields and remoteness depends on the measure used, yield per acre is lower in areas located farther away from hubs. In particular, one standard deviation increase in either measure of remoteness is associated with a reduction in harvest output per acre of about 20%. This is consistent with lower input usage in rural areas, or with differences in other factors such as soil quality.

In conclusion, Table 3 makes clear that it is difficult to pinpoint the role of input prices on outcomes, since access to roads is correlated with so many other characteristics. Ultimately this motivates the use of an economic model to conduct counterfactuals.

4.2.4 Access to input markets

Table 4 shows how the two remoteness measures correlate with access to input markets. We first tabulate access to retailers over a distance that is reasonably traveled by farmers. Web Appendix Figure A2 shows a CDF of the distance farmers travel to access inputs, conditional on purchase. We find that approximately 70% of purchases are made within 10 km of a farmer's village, and 85% within 20 km. We therefore show results within 10 km of a farmer's village. Panel A shows several measures of retailer activity in the area, including a dummy for whether there is at least one retailer within 10 km, the number of retailers within 10 km, and the minimum distance to a retailer. On each measure, we find clear evidence of reduced access to retailers in more remote villages, all significant at 1%.

Panel B of Table 4 shows our key measure of access, travel cost-adjusted prices. We find that one standard deviation of remoteness raises prices by \$2.3-2.45, equivalent to about 9.5-10% of the mean. This implies a difference in prices of approximately 50% between the most and least remote villages in our study sample. We then decompose this price difference into differences in the price itself, and in the travel cost. We find that the contributions of retail prices and transportation costs are approximately equal in their contribution toward the increase in delivered prices.

Here, it is useful to discuss the magnitude of transport costs incurred to buy fertilizer, whether to the best location, and per kilometer to the nearest location. Above, when evaluated at the local one-way travel cost of \$0.088 per km, the total cost of acquiring a 50kg bag of fertilizer is approximatly \$0.238 per kilometer. The average price of a bag of fertilizer is about \$25 (see Web Appendix Table A5), and thus, the ad-valorem travel cost relative to this price is about 0.95% per kilometer. Since the nearest agrovet is about 6.8km away on average, the ad-valorem cost of travel to the nearest agrovet is about 6.4% (evaluated at the sample average price). While these costs may seem modest, they do not account for two things. First, the nearest agrovet is not always the best agrovet, and in our above exercise finding the best transportation cost adjusted price (Table 4), transport costs are about 22% of the optimal purchase. Thus, transport costs are a substantial share of delivered costs in reaching the agrovest with the best prices. Second, there may be other, non-pecuniary costs of travel, like information on prices, opportunity costs of significant travel, and uncertainty regarding inventories at more distant locations, as well as differences in retailer quality that skew the decision on where to go and travels costs one is willing to incur. Accordingly, we later build a model to estimate transport costs based on revealed choices to agrovets that exhibit differences in both quality and price, and use this model to quantify the role transport costs in the adoption decision.

Continuing focus on the prices and remoteness, we look at agrovet pricing in Web Appendix Table A5. Note that these results should be taken with some caution, since they are conditional on entry (i.e. the agrovet location is not exogenous), and so we take these as more descriptive than definitive. Panel A shows the gradient of distance on sales and product variety, showing little heterogeneity except for the costs of reaching a wholesaler. In surveys, we find that nearly all agrovets travel themselves to distributors to access inputs, and we find that these costs are higher in remote areas (naturally), since they must travel further. Panel B shows the relationship on prices and markups. Note that these regressions are the fertilizer variety level (i.e. product barcode). We find some weak evidence that rural shops face slightly higher wholesale prices (significant at only 10%) and some stronger evidence that they charge higher retail prices. The effect is modest though statistically significant: a 1 standard deviation increase in distances is associated with 2.5% higher prices. We then calculate markups, inclusive of transport costs. We find no strong evidence of differential markups in remote areas – while retail prices are higher, this is largely due to transport costs directly passing through to farmers. From this, we conclude that much of the higher retail prices are a result of higher marginal costs of accessing inputs.[14]

4.2.5 Access to output markets

Table 5 performs a similar analysis, but on the output side. As before, Panel A shows that more remote villages are less likely to have a market within 10 km and have to travel farther to reach a market where maize is sold. Panel B1 shows travel cost-adjusted prices for maize. Since there are

[14] In Web Appendix Table A6, we conduct two robustness checks. First, we only surveyed retailers within the regional boundaries and thus have no information on retailers in neighboring regions. While most of these boundary areas are remote, it is nevertheless possible that there exist lower-priced retailers just over regional boundaries. Since we miss these, we may overstate travel cost-adjusted prices. To address this, we drop all villages within 10 km of regional boundaries – results are actually stronger. Second, while we had high survey completion rates among agrovets (91% – see Web Appendix Table A1), we nevertheless do not have the universe of retail options. This suggests that retail price heterogeneity is understated. However, it may affect the regression results. To address this, we conduct a bounding exercise in Web Appendix Table A4, Panel B. In this exercise, we estimate the distribution of prices within regions. We then assign prices in the tails of this distribution (the 10th or 90th percentile) to missing agrovets in a way that attenuate our regression results – for example, in remote areas, we assign agrovets low prices. This exercise lowers the coefficient marginally, but the qualitative results are unchanged.

large seasonal price fluctuations in rural Tanzania as in much of rural Africa,[15] we use a price for the single point in time which is most relevant for farmers: immediately post-harvest. Our surveys show that most farmers who sell do so shortly after harvest. We find that travel cost-adjusted prices of output are lower in remote areas, and consistent across both remoteness measures. As before, we decompose this into the retail price and the travel costs, finding that while retail maize prices rise modestly, transport costs to the closest maize market rise by \$3.9 with each standard deviation in remoteness, overwhelming the increase in the price of maize.

Finally, we show two other measures of price, both measured at the village level: (1) we asked farmers what the "going price" of maize was after the last harvest; (2) we collected information on sales prices for all farmers who sold maize. While this latter measure is affected by selection, we minimize this problem by taking the average sales prices across the village. Using both remoteness measures, the going price in the village, whether measured through surveys or actual sales of maize, falls when a village is more remote. This is intuitive if maize agents are traveling from the larger population centers (which are used to construct our remoteness measures), and offering lower selling prices to compensate for the higher costs of travel to more remote villages. We examine agent activity further in the Table 6from the perspective of the farmer.

4.2.6 Farmer decisions

The results so far show clear evidence of reduced market access on both input and output markets, and of higher prices for inputs, lower (travel cost-adjusted) prices for output, and lower "going" prices for output within the village. These results lead us to expect decreasing input usage and maize sales with increasing remoteness. We investigate this in Table 6, where we present results with and without a full set of farmer controls from the surveys. In Panel A, we present indicators for using inputs as well as the quantity used, for both seeds and fertilizer. In all specifications, these relationships are very strong (significant at 1%) and large. We find that use of fertilizer is 9-17 percentage points lower in villages 1 standard deviation away, and that of hybrid seeds is 5-7 percentage points lower. Since the distance between the least and most remote regions is about 5 standard deviations, the regressions predict approximately a minimum of 45 percentage point lower usage of fertilizerin the most remote villages, which translates to about 80% of the mean in the least remote areas. The effect for seeds is smaller but still evident.

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

Consistent with this, Panel C shows buying behavior. Remote farmers are more likely to buy

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

maize and are more likely to be net buyers of maize. Interestingly, we find a lot of heterogeneity in net buying behavior. We find that 37% of farmers buy maize but sell none, 24% sell maize but buy none, and only 8% buy and sell maize (the other 30% do not transact on either side of the market).

5 Model

The reduced form results suggest that more remote farmers suffer from reduced access to input retailers as well as selling opportunities for output, and that this ultimately affects their decision to adopt fertilizer. However, as these results are purely descriptive, we now push the data further and 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 framework of retailer choice, including for reasons unrelated to trade costs, as well as allowing for other factors that affect adoption but are not related to input access. Then, using a model specified measure of input access that we estimate, we run counterfactuals that study the role of transportation costs in the adoption decision.

5.1 Model Preliminaries

Production and Inputs

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

$$Y_{i0} = \widetilde{\theta}_{i0} K_i^{\alpha_0} L_{i0}^{1-\alpha_0} \tag{7}$$

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

$$\begin{aligned} \Pi_{i0} &= \alpha_0(1-\alpha_0)^{\frac{1-\alpha_0}{\alpha_0}} \widetilde{\theta}_{i0}^{\frac{1-\alpha_0}{\alpha_0}} p_i^{\frac{1}{\alpha_0}} w_i^{-\frac{1-\alpha_0}{\alpha_0}} K_i \\ &= \theta_{i0}\pi_{i0} \end{aligned} \tag{8}$$

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

The production function *with* fertilizer splits up variable inputs into labor and fertilizer, while maintaining the basic Cobb-Douglass assumption. When using fertilizer, farmers not only have a choice of how much fertilizer to buy, but also which retailer to choose. Supposing that farmer i buys fertilizer from agrovet j in location v,the production function is written as:

$$Y_{ijv} = \widetilde{\theta}_{ijv} \left(\theta_i K_i\right)^{\alpha} L_{ijv}^{(1-\alpha)\beta} M_{ijv}^{(1-\alpha)(1-\beta)} \tag{9}$$

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

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

$$\Pi_{ijv} = \theta_{ijv}\pi_i r_{ijv}^{-\sigma} \tag{10}$$

where $\sigma \equiv \frac{1-\alpha}{\alpha}(1-\beta)$, $\pi_i = \theta_i p_i^{\frac{1}{\alpha_0}} w_i^{-\beta\frac{1-\alpha_0}{\alpha_0}} K_i$, and $\theta_{ijv} = \kappa_2 \widetilde{\theta}_{ijv}^{\kappa_1}$.[16] Here, the profitability of fertilizer is a function of the productivity shock, θ_{ijv}, the (delivered) price of fertilizer itself, r_{ijv}, and deterministic profits based on local observables and technology, π_i.

Input and Agrovet Choice

Farmers have a choice of whether to purchase fertilizer, and if so, how much and where to purchase the fertilizer. These decisions areaffected by prices for fertilizer at each agrovet location, the productivity shock received in buying from a particular 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 τ_{iv} is an iceberg trade cost for farmer i in traveling to v and back. The assumption of iceberg trade costs will facilitate a decomposition of the model that aids in calibration.

Further, we assume that θ_{ijv} is a random variable that measures the benefit of purchasing at agrovet j in location v. These latter benefits could represent other inputs purchased in location v (hybrid seeds, for example), availability of extension services at location v, or perhaps other networking and information that is acquired at location v that may affect profitability. Further, it may represent the probability of getting bad or adulterated inputs at a given retail location, or given the functional form of (10), measurement error in the price at a retail location. Whatever the interpretation, for analytical convenience we assume that θ_{ijv} is distributed according to Fréchet distribution with location parameter T_{jv} and dispersion parameter ε. Precisely:

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

That is, while each farmer may get a random draw from this distribution, its central moments

[16] κ_1 and κ_2 are constant functions of model parameters

are specific to the retail location itself. Using this distributional assumption, the unconditional distribution of profits for farmer i buying from agrovet j in location v is written as:

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

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

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

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

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

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

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

$$\mu_i = \frac{\sum_{v'\in\mathcal{V}}\sum_{l\in\mathcal{J}_{v'}} T_{lv'} r_{ilv'}^{-\varepsilon_a}}{\frac{T_{i0}}{T_i}\left(\frac{\pi_{i0}}{\pi_i}\right)^{\varepsilon} + \sum_{v'\in\mathcal{V}}\sum_{l\in\mathcal{J}_{v'}} T_{lv'} r_{ilv'}^{-\varepsilon_a}} \equiv \frac{\Phi_i}{\Phi_{i0} + \Phi_i} \tag{12}$$

In (12), we define the two terms that fully characterize the adoption decision for each farmer i. First, we define farmer i's *market access* to inputs as $\Phi_i = \sum_{v'\in\mathcal{V}}\sum_{l\in\mathcal{J}_{v'}} T_{lv'} r_{ilv'}^{-\varepsilon_a}$, which after imposing the iceberg assumption and simplifying, can be written as

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

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

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

which is the relative profitability of using fertilizer (compared to to not using fertilizer), adjusted for local productivity factors that are unrelated to market access. To be clear, we will not be able to disentagle each component of Φ_{i0} lacking additional data and identifying variation in output markets or the micro-foundations of farm production. However, for purposes of calibrating the model-derived measure of market access and its relationship to the remoteness of villages from population centers, Φ_{i0} will be useful in absorbing all other variation in adoption as a residual.

5.2 Calibrating the Farmer's Problem

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

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

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

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

Estimating Transport Costs through Location Choice

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

$$\lambda_{iv|adopt} = \frac{\tau_{iv}^{-\varepsilon_a}\phi_v}{\sum_{v'\in\mathcal{V}} \tau_{iv'}^{-\varepsilon_a}\phi_{v'}} \tag{13}$$

To estimate equation (13), 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, and $\mathcal{V}$ as the set of locations with an agrovet, we construct a $\mathcal{I} \times \mathcal{V}$ dataset of bilateral visit indicators. There will be many zeros in this dataset. For each bilateral combination, we will also measure distance in kilometers between the farmer's village and the potential purchase location, $dist_{iv}$.

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

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

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

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

$$\lambda_{iv|adopt} = \frac{\exp\left(d_v + \sum_b \beta_b D^b_{iv}\right)}{\sum_{v' \in \mathcal{V}} \exp\left(d_{v'} + \sum_b \beta_b D^b_{iv'}\right)} \tag{14}$$

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 (14), and its simplified linear alternative, can be estimated by McFadden's alternative-specific conditional logit. The results from doing so are presented in Table 7. In the first column, we present the linear specification of distance. Assuming $\varepsilon_a \approx 5$ (which will be supported by later results), the results suggest that iceberg transport costs for fertilizer are 3.4% 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 7, where column 3 reports the ad-valorem equilivalent per kilometer when evaluated at the farthest distance that defines each bin, and with trade costs compounded each kilometer.[17]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 then focus on the reduction in probability if one is (0,5] km away rather than 0 km away (essentially in the same village). In this case, the probability that one chooses the location (0,5] km away compared to 0km away (in the home village) for idiosyncratic reasons that overcome trade costs is 0.25.[18]

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

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

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

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

$$\log\left(\tau_{iv}\right) = -\frac{1}{\varepsilon_a}\sum_b \beta_b D_{iv}^b \tag{15}$$

Dividing the coefficient estimates by ε_a gives us log trade costs. Given the iceberg assumption, this is also interpreted as the log change in the delivered price. Thus, at the central estimate of $\varepsilon_a \approx 5$, the comparison is equivalent to approximately 5.7% ad-valorem trade cost per km for the first three bins (up to 15km). When evaluated to the typical closest agrovet (6.7 km), the ad-valorem equivalent trade cost is 45%Beyond this, the ad-valorem equivalent per kilometer falls modestly, which is consistent with our transport surveys and also the likelihood that longer distances require a more efficient means of travel (though still at a high overall cost, on average).

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, which suggests sizable non-pecuniary costs of traveling to acquire fertilizer.

Model Calibration

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

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

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

where $\mathbb{E}\left[F_i | adopt\ at\ jv\right]$ is expected fertilizer expenditures by i, conditional on adopting at jv, and L_i is the village population to use as weights in the demand equation. As this conditional expectation is not observed in any practical way, we will appeal to the structure of the model to simplify to an unconditional expectation for fertilizer expenditures by i. Precisely, using the properties of the Fréchet distribution, it is straightforward to show that $\mathbb{E}\left[F_i | adopt\ at\ jv\right] = \mathbb{E}\left[F_i | adopt\right]$; that is, the expected expenditures conditional on adoption anywhere is the same as the expected expenditures at some j, conditional on choosing j.[19] Noting further that $\mu_i \mathbb{E}[F_i|adopt] = \mathbb{E}[F_i]$, we have:

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

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

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

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

$$\mathbb{E}[v_{jv}] = \sum_i L_i \left(\frac{\exp\left(-\sum_b \hat{\beta}_b D_{iv}^b\right) \eta_{jv}}{\sum_{v' \in \mathcal{V}} \sum_{l \in \mathcal{J}_{v'}} \exp\left(-\sum_b \hat{\beta}_b D_{iv'}^b\right) \eta_{lv'}} \right) \mathbb{E}[F_i] \tag{16}$$

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

However, we again run into a number of issues where our farmer survey was not large enough to survey farmers from every village in the region, and the population of farmers within each village. To this end, there are two issues to consider. First, within each village, while we surveyed approximately 18 maize farmers per village, in some villages zero or full adoption is reported. In reality, this may be accurate, or may be biased toward the bounds by a small sample. Further, village adoption at the bounds will also complicate the calibration of the overall adoption decision (which we describe in a moment). To facilitate a feasible calibration that is consistently applied across market clearing conditions and the adoption decision, we first winsorize the village adoption data to fall between 0.025 and 0.975.[20] Then, for those villages that report zero adoption in the sample, we assign a small value of $\mathbb{E}[F_i]$ that is calculated via the model using the reported land holdings of the village in the sample, the winsorized adoption share (0.025), and then the 1st percentile value of fertilizer expenditures per acre of land across the entire sample of farmers who adopt, $\left(\frac{F_i}{K_i}\right)_{1st}$.[21]

Focusing on the sampling of villages, if we assume that the farmer sample captures the entire geography of demand, there will exist agrovets in other locations that appear more remote than they actually are since no farmers were surveyed in that location. This will cause a bias in estimates of η_{jv}'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 assume that

[19] See technical appendix for a proof.

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

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

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

Two other empirical issues to consider are more straightforward. Since agrovet fertilizer revenues and farmer expenditures are from different surveys, and the latter aggregated from a farmer level sample, we normalize each to sum to one. After doing so, we can recover η_{jv} by solving the non-linear system of equations formed using $\mathcal{J}$ agrovets and their revenue shares, as written in (16), under the normalizing assumption that $\sum_v \sum_j \eta_{jv} = 1$.[22]

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

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

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

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

The estimated values of $\hat{\Phi}_i$ and $\hat{\Phi}_{i0}$ are presented in Figure 4. The solid line illustrates the fitted relationship between standardized remoteness and market access, $\hat{\Phi}_i$, and is significantly negative. Precisely, a one standard deviation increase in remoteness leads to a 0.77 reduction in log market access. In contrast, there is a mild positive but statistically insignificant relationship between standardized remoteness and the outside option, $\hat{\Phi}_{i0}$.

5.3 Estimating Elasticities for Counterfactuals

Above, we have fully calibrated the farmers problem in terms of the decision to adopt as a function quality-adjusted access to markets, and shown that there is a significant drop in market access for remote villages relative to those closer to urban hubs. The results also suggest a slightly higher outside option to using fertilizer in remote markets that may be due to local suitability for fertilizer or other market conditions. To evaluate the role of trade costs in market access and output prices in the outside option, we must push the model further to estimate the fundamental parameters of the productivity distribution and the production functions, with and without fertilizer.

We begin by estimating the composite elasticity of substitution between agrovet options, ε_a, which is a function of the native Fréchet dispersion parameter, and the land (α) and labor (β) shares in the production function with fertilizer. To estimate this elasticity, we log-linearize the

[22] This normalizing assumption is required since the probabilities within the sum in equation (16) are homogeneous degree zero in η's.

definition of η_{jv} to get:

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

As η_{jv} is calibrated using revenue-expenditure market clearing conditions, there is an obvious endogeneity problem in estimating ε_a. We address this endogeneity in two ways: First, we include as controls agrovet-level experience at that location, $exper_{iv}$, and district fixed effects, part to capture their correlatations with $\log(T_{jv})$. Thus the equation to estimate becomes:

$$\log(\hat{\eta}_{jv}) = -\varepsilon_a \log(r_{jv}) + \underbrace{\beta exper_{jv} + district + u_{jv}}_{\log(T_{jv})} \tag{17}$$

In addition to these simple controls, we instrument for current agrovet prices with one-year lagged prices. This yields an estimate $\varepsilon_a = 4.91$, which we use in later counterfactuals. A full set of estimates under OLS and IV, with regression diagnostics, is presents in Web Appendix Table A7.

Next, we estimate prodution parameters that are embeded in Φ_{i0}. Recall that $\Phi_{i0} = \frac{T_{i0}}{T_i}\left(\frac{\pi_{i0}}{\pi_i}\right)^{\varepsilon} = \frac{T_{i0}}{T_i} p_i^{\varepsilon_p} w_i^{\varepsilon_w}$, where $\varepsilon_p = \varepsilon\left(\frac{\alpha-\alpha_0}{\alpha\alpha_0}\right)$ and $\varepsilon_w = \varepsilon\left(\beta\frac{1-\alpha}{\alpha} - \frac{1-\alpha_0}{\alpha_0}\right)$; thus, the effects of any output price shock are a function of the relative importance of land in the production function. Further, β is important for the wage elasticy of adoption, as well as decomposing ε_a into the native dispersion parameter ε and production parameters. In Web Appendix Table A8, we detail a simple estimator for maize production with and without fertilizer, and using data from the Tanzania Living Standards Measurement Study and Integrated Surveys on Agriculture (LSMS-ISA) produce estimates for α (0.431) and α_0 (0.570). Also using the LSMS-ISA, we use reported wages and labor and fertilizer expenditures to calculate the share of labor in variable factors; β (0.75).[23] Using these estimates, it is straightforward to calculate that ε=14.96. While this may seem high, this essentially means that there is little idiosyncratic variation in quality-adjusted prices at each agrovet, around the T_{jv}'s. Practically, farmers are choosing the lowest quality adjusted price for each agrovet, with minimal other variation that distracts from prices, quality, and transport costs.

5.4 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 retail 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).

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

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

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

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

$$r_{jv} = \frac{\varepsilon^v_{jv} - 1}{\varepsilon^v_{jv}} c_{jv} \tag{18}$$

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

$$\varepsilon_v = -\varepsilon_a + \frac{\varepsilon - 1}{\varepsilon} \varepsilon_a \sum_i s_{ijv} \lambda_{ijv}$$

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

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 21.6% (median = 20.2%), which while higher is not remarkably different from the measured mark-ups in the reduced form (mean = 13%).

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 (18), 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 a number of 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. We study these issues in two ways: reduction in transport costs, and effects of agrovet entry.

Transportation Costs

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 in terms of adoption by each village are displayed in the top-left panel of Figure 5. The bottom left panel reports the effects on log fertilizer expenditures within each village. For clarity, we have grouped villages into 20 equally sized bins of standardized remoteness, and the points in the Figure represent average adoption within these groups. For interpretation, we have also plotted lines of best fit when regressing baseline or counterfactual adoption (or expenditures) on remoteness. For these regressions, we use the raw village data rather than binned.

In the top left panel, we find a large adoption effect of 27pp, or approximately twice baseline. This counterfactual alone accounts for 15% of the adoption-remoteness relationship at baseline. The results for expenditures are even more pronounced, where expenditures rise approximately 133 log points, though it is important to keep in mind that for many villages we are moving from an extremely low base. Nevertheless, the log-expenditure-remoteness gradient is cut in half by this counterfactual. 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 affects 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 a regional hub. So, in the second counterfactual, we subsidize this cost by 50%. The effects of this counterfactual are presented in the top-right panel of Figure 5. Here, adoption rises by about 1pp, or 4.4%, and yields a 4% reduction in the remoteness-adoption gradient. Thus, although effects on levels are modest, the effect of this subsidy on the remoteness gradient, despite any absorption by mark-ups, is non-trivial.

Entry

An overarching question throughout the paper has been why agrovet access is worse in remote areas, and in particular, why agrovets enter intensely in other areas. While we do not present an empirical model of entry (for example, as in Seim, 2006), we do run a simple counterfactual to examine profitability of entry and any corresponding effects on adoption. Specifically, we force a "median" agrovet (as defined by T_{jv} and marginal cost, within a district) to enter every village in the sample (one at a time, not simultaneously), and then measure the effects of that singular entry on adoption, and also measure the profitability of the entrant after entry. We do this for every village in the dataset, and then plot in the top panel of Figure 6 aggregate adoption (after entry) and entrant profits as a function of distance to the village in which the entry took place. Very clearly, profits are lower when entering more remote villages, though adoption effects are higher when entering the more remote villages. The former relationship is particularly strong, where a one standard deviation increase in remoteness reduces the profitability of hypothetical entry by 33 log points. Thus while access to agrovets in remote areas would improve adoption more so than entering less-remote areas, the profitability analysis supports the argument that this lack of entry in remote areas is logical.

6.2 Experiments in Output Access

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

First, as in the reduced form exercises, we assume that farmers optimally choose the best market to sell, subject to the measured/estimated costs of transportation from village to village. We measure the baseline best net-selling price, and then recalculate this best net-selling price after reducing transportation costs by 50%. With the baseline and counterfactual net selling price, the shock that is relevant to the model for each farmer is $\Phi_{i0}\left(\frac{p_{ic}}{p_{i0}}\right)^{\varepsilon_p}$, where the ratio of counterfactual prices p_{ic} to baseline output prices p_{i0} is interacted with the original calibrated parameter and raised to the price-elasticity ε_p. The results of this counterfactual shock are presented in the bottom left panel of Figure 6. Here, we see an adoption-remoteness gradient that is cut by a similar amount to the input market counterfactuals, and an adoption effect that is about 0.2, or 65%. Thus, this counterfactual has a lower effect on adoption when compared with a 50% cut in farmer-retailer transport costs, but a similar effect on the gradient. However, care must be taken in comparing the two. In the case of the cut in farmer-retailer transport costs, the transport cost cut is interacted with prices and calibrated agrovet quality terms, which adds noise to the shock. That is, while a farmer might be more likely to travel to any agrovet, the transport shock is not concentrated on the agrovet that is closest or that the farmer will necessarily choose. In contrast, the best net selling price is determined by a simple calculation of the maximum net price, with no probability of choosing different option. Thus, the noise in the farmer-retailer transport shock as it relates to distance will attenuate its effect on the gradient, though still provide a sizable effect on adoption.

Keeping in mind the same caveat for our final counterfactual, we now assume that farmers sell at their closest market, and experience a 50% reduction in iceberg costs to that market, as estimated by the agrovet choice problem. That is, the farmer must ship τ_{im} units of maize to the market to effectively sell one unit. Thus the transport-adjusted selling price that we use for each village above when calibrating adoption decisions is equal to $p_i = p_m/\tau_{im}$, where p_i is net selling price to farmer i and p_m is the price at the primary market for that village. To examine the impact of output market access on adoption, we now run an experiment cutting the iceberg costs to reach output markets by 50%. The results from this counterfactual are presented in the bottom-right panel of Figure 5. Here, adoption almost doubles, and the adoption-remoteness gradient falls by about 30%.

Overall, access to output markets appears to be an important component of the input adoption decision, and these effects deserve more detailed attention in follow-up work that more rigorously documents the movement of maize from the plot to the market and beyond.

7 External validity

While our data collection spans a large portion of Northern Tanzania, our data is still limited to only one region in one country. Do our results generalize? While we cannot provide a definitive answer, we provide some suggestive evidence in this section.

7.1 Price Dispersion

To address this, we assembled five secondary datasets[24] across 1,512 locations[25] in 56 African countries. We compare this to a small dataset we assembled between March and April 2016 with 251 retailers of various sorts (shops, agro-input dealers, and maize traders) in 82 markets in the Kilimanjaro region.[26] To quantify price dispersion, we first decompose variation in spatial prices by running the following regression:

$$\log(p_{mcjt}) = \gamma_c + \gamma_j + \gamma_t + \epsilon_{mjt} \tag{19}$$

where p_{mcjt} (log) prices in market m for product j at time t in country c, and the γ terms are country, product, and time fixed effects. We calculate the standard deviation of the resulting residual. Results are reported in Web Appendix Table A9. 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,[27] to run dyadic regressions to look at price gaps, as follows:

[24] 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.

[25] 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.

[26] 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.

[27] 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).

$$\log(| p_{mjt} - p_{m'jt} |) = \theta \log(c_{mm'}) + \gamma_m + \gamma_{m'} + \gamma_j + \epsilon_{mm'jt} \tag{20}$$

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.[28] Results are presented in Web Appendix Table A10. 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. 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 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 A11, 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-ISA, 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.

8 Conclusion

In this paper we collect detailed 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 1,183 villages in the Kilimanjaro and Manyara regions 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 that the most remote villages face prices substantially higher than the least remote (equivalent to about 40% of the mean). The rates and magnitudes of fertilizer use and maize sales also display a large and significant distance gradient. 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 qualified one for those on the output side.

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

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 (2018a). "Do Rural Roads Create Pathways out of Poverty? Evidence from India." *Journal of Development Economics* 133: 375-395.

[3] Aggarwal, Shilpa (2018b). "The Long Road to Health: Healthcare Utilization Impacts of a Road Pavement Policy in Rural India." Unpublished

[4] Aggarwal, Shilpa, Eilin Francis, and Jonathan Robinson (2018) "Grain Today, Gain Tomorrow: Evidence from a Storage Experiment with Savings Clubs in Kenya," *Journal of Development Economics* 134: 1-15.

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

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

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

[8] 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.

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

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

[11] 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.

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

[13] Bergquist, Lauren, Marshall Burke, and Edward Miguel (2017). "Selling Low and Buying High: an Arbitrage Puzzle in Kenyan Villages." 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] Casaburi, Lorenzo and Tristan Reed (2017). "Competition in Agricultural Markets: An Experimental Approach." Unpublished

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

[18] 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.

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

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

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

[22] Donaldson, Dave and Richard Hornbeck (2016). "Railroads and American economic growth: A "market access" approach." *Quarterly Journal of Economics* 131 (2): 799-858.

[23] 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.

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

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

[26] 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.

[27] 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.

[28] Head, Keith and Thierry Mayer (2011). "Gravity, market potential and economic development", *Journal of Economic Geography*, Vol 11, pp. 281–294

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

[30] Minten, Bart, Bethlehem Koru and David Stifel (2013). "The last mile(s) in modern input distribution: Pricing, profitability, and adoption", *Agricultural Economics*, Vol 44, pp. 629-646

[31] Porteous, Obie (2019). "Trade and Agricultural Technology Adoption: Evidence from Africa." Unpublished.

[32] Porteous, Obie (2017). "High Trade Costs and Their Consequences: An Estimated Dynamic Model of African Agricultural Storage and Trade", *Forthcoming, American Economic Journal: Applied Economics.*

[33] Redding, Stephen and Anthony Venables (2004), "Economic geography and international inequality", *Journal of International Economics,* 62: 53 – 82.

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

[35] Seim, Katja (2006), "An Empirical Model of Firm Entry with Endogenous Product-Type Choices", *The RAND Journal of Economics*, 37-3, pp 619-640

[36] Sotelo, Sebastian (2018). "Domestic Trade Frictions and Agriculture." Unpublished.

[37] 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.

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

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

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

[41] 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

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

Appendix A: Remoteness and Market Access

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

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

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

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

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

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

Collecting terms:

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

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

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

Standardizing this equation gives us:

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

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

Appendix B: Deriving farmer profits, revenues, and input expenditures

The production function under basic technology is:

$$Y_i = \widetilde{\theta}_{i0} K_i^{\alpha} L_i^{1-\alpha} \tag{21}$$

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

$$\Pi_{i0} \quad = \quad \max_{L_i} \left\{ p_i \widetilde{\theta}_{i0} K_i^{\alpha} L_i^{1-\alpha} - w_i L_i \right\} \tag{22}$$

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

$$(1-\alpha) p_i \widetilde{\theta}_{i0} K_i^{\alpha} L_i^{-\alpha} = w_i \tag{23}$$

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

$$w_i L_i = (1-\alpha) p_i \widetilde{\theta}_{i0} K_i^{\alpha} L_i^{1-\alpha} = (1-\alpha) R_{i0} \tag{24}$$

and substituting into the profit function, we have:

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

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

$$\begin{aligned} \Pi_{i0} &= \alpha_0 (1-\alpha_0)^{\frac{1-\alpha_0}{\alpha_0}} \widetilde{\theta}_{i0}^{\frac{1}{\alpha_0}} p_i^{\frac{1}{\alpha_0}} w_i^{-\frac{1-\alpha_0}{\alpha_0}} K_i \\ &= \theta_{i0} \pi_{i0} \end{aligned} \tag{25}$$

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

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

$$Y_i = \widetilde{\theta}_{ijv} \left(\theta_i K_i\right)^{\alpha} L_{ijv}^{(1-\alpha)\beta} X_{ijv}^{(1-\alpha)(1-\beta)} \tag{26}$$

The profit maximization problem when using fertilizer is written as:

$$\Pi_{i0} \quad = \quad \max_{L_i, X_{ijv}} p_i \widetilde{\theta}_{ijv} \left(\theta_i K_i\right)^{\alpha} L_{ijv}^{(1-\alpha)\beta} F_{ijv}^{(1-\alpha)(1-\beta)} - w_i L_{ijv} - r_{ijv} F_{ijv} \tag{27}$$

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

$$c_{ijv} M_{ijv} = (1-\alpha) p_i \widetilde{\theta}_{ijv} \left(\theta_i K_i\right)^{\alpha} L_{ijv}^{(1-\alpha)\beta} F_{ijv}^{(1-\alpha)(1-\beta)} = (1-\alpha) R_{ijv} \tag{28}$$

and

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

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

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

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

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

$$\Pi_i = \theta_{ijv} \pi_i r_{ijv}^{-\sigma} \tag{29}$$

where $\sigma \equiv \frac{1-\alpha}{\alpha}(1-\beta)$, $\pi_i = p_i^{\frac{1}{\alpha}} w_i^{-\beta \frac{1-\alpha}{\alpha}} K_i$, and $\theta_{ijv} = \kappa_2 \widetilde{\theta}_{ijv}^{\kappa_1}$.[29] Here, the profitability of fertilizer at this location is a function of the productivity shock, θ_{ijv}, the (delivered) price of fertilizer itself, r_{ijv}, and profits based on local observables and technology π_i.

[29] κ_1 and κ_2 are constant functions of model parameters

Appendix C: Distributions of Fertilizer Expenditures

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

$$\mathbb{E}\left[rF_i \mid adopt\ at\ j\ in\ v\right] = \mathbb{E}\left[rF_i | adopt\right] \tag{30}$$

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 when choosing that agrovet. Thus, we focus all proofs on the distribution of profits, and then the analogue to revenues and input expenditures follows directly. To begin, we first derive the distribution of profits for farmer i who buys from agrovet j in location v.

$$\begin{aligned}
\Pr\left(\Pi_{ijv} > \pi\right) &= \Pr\left(\theta_{ijv}\pi_i r_{ijv}^{-\sigma} > \pi\right) && (31)\\
&= \Pr\left(\theta_{ijv} > \frac{\pi}{\pi_i} r_{ijv}^{\sigma}\right) && (32)\\
&= 1 - \exp\left(-T_{jv}\pi_i^{\varepsilon} r_{ijv}^{\varepsilon\sigma}\pi^{-\varepsilon}\right) && (33)
\end{aligned}$$

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

$$\Pr\left(\Pi_{ijv} > \pi\right) = 1 - \exp\left(-T_{jv}\gamma_{ijv}\pi^{-\varepsilon}\right) \tag{34}$$

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

$$\Pr\left(\Pi_{i0} > \pi\right) = 1 - \exp\left(-\Phi_{i0}\pi^{-\varepsilon}\right) \tag{35}$$

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

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

$$\begin{aligned}
\Pr\left(\Pi_i^{max} > \pi\right) &= \Pr\left(\Pi_{ijv} > \pi \text{ for any } jv\right) && (36)\\
&= 1 - \Pr\left(\Pi_{ijv} < \pi\ \forall\ jv\right) && (37)
\end{aligned}$$

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

$$\begin{aligned}
\Pr\left(\Pi_i^{max} > \pi\right) &= 1 - \Pr\left(\Pi_{ijv} < \pi \ \forall \ jv\right) && (38)\\
&= 1 - \prod_{v' \in \mathcal{V}} \prod_{j \in \mathcal{J}_v} \Pr\left(\Pi_{ijv} < \pi\right) && (39)\\
&= 1 - \prod_{v' \in \mathcal{V}} \prod_{j \in \mathcal{J}_v} \exp\left(-\pi^{-\varepsilon}\right) && (40)
\end{aligned}$$

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

$$\Pr\left(\Pi_i^{max} > \pi\right) = 1 - \exp\left(-\tilde{\Phi}_i \pi^{-\varepsilon}\right) \tag{41}$$

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

$$G_i^{max}\left(\pi\right) = \Pr\left(\Pi_i^{max} < \pi\right) = \exp\left(-\tilde{\Phi}_i \pi^{-\varepsilon}\right) \tag{42}$$

with pdf:

$$g_i^{max}\left(\pi\right) = \varepsilon \tilde{\Phi}_i \pi^{-\varepsilon-1} \exp\left(-\tilde{\Phi}_i \pi^{-\varepsilon}\right) \tag{43}$$

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

$$\begin{aligned}
\Pr\left(\Pi_i > \pi\right) &= \Pr\left(\Pi_{ijv} > \pi \text{ for any } jv \cup \Pi_{i0} > \pi\right) && (44)\\
&= 1 - \Pr\left(\Pi_{ijv} < \pi \ \forall \ jv \ \cap \ \Pi_{i0} < \pi\right) && (45)
\end{aligned}$$

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

$$\begin{aligned}
\Pr\left(\Pi_i > \pi\right) &= 1 - \Pr\left(\Pi_{ijv} < \pi \ \forall \ jv \ \cap \ \Pi_{i0} < \pi\right) && (46)\\
&= 1 - \Pr\left(\Pi_{i0} < \pi\right) \prod_{v' \in \mathcal{V}} \prod_{j \in \mathcal{J}_v} \Pr\left(\Pi_{ijv} < \pi\right) && (47)\\
&= 1 - \exp\left(-T_{i0}\gamma_{i0}\pi^{-\varepsilon}\right) \prod_{v' \in \mathcal{V}} \prod_{j \in \mathcal{J}_v} \exp\left(-T_{jv}\gamma_{ijv}\pi^{-\varepsilon}\right) && (48)
\end{aligned}$$

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

$$\Pr\left(\Pi_i > \pi\right) = 1 - \exp\left(-\left(\tilde{\Phi}_{i0} + \tilde{\Phi}_i\right)\pi^{-\varepsilon}\right) \tag{49}$$

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

$$G_i\left(\pi\right) = \exp\left(-\left(\tilde{\Phi}_{i0} + \tilde{\Phi}_i\right)\pi^{-\varepsilon}\right) \tag{50}$$

with pdf:

$$g_i(\pi) = \varepsilon\left(\tilde{\Phi}_{i0}+\tilde{\Phi}_i\right)\pi^{-\varepsilon-1}\exp\left(-\left(\tilde{\Phi}_{i0}+\tilde{\Phi}_i\right)\pi^{-\varepsilon}\right) \tag{51}$$

Profits conditional on adoption

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

$$\Pr\left(\Pi_i^{max}<\pi\middle|adopt\right) = \frac{\Pr\left(\Pi_i^{max}<\pi\cap\Pi_i^{max}>\Pi_{i0}\right)}{\Pr\left(\Pi_i^{max}>\Pi_{i0}\right)} \tag{52}$$

This can be re-written as:

$$\begin{aligned}\Pr\left(\Pi_i^{max}<\pi\middle|adopt\right) &= \frac{1}{\Pr\left(\Pi_i^{max}>\Pi_{i0}\right)}\int_0^\pi \Pr\left(s>\Pi_{i0}\right)g_i^{max}(s)ds\\ &= \frac{1}{\Pr\left(\Pi_i^{max}>\Pi_{i0}\right)}\int_0^\pi \exp\left(-\tilde{\Phi}_{i0}s^{-\varepsilon}\right)\varepsilon\tilde{\Phi}_i s^{-\varepsilon-1}\exp\left(-\tilde{\Phi}_i s^{-\varepsilon}\right)ds\\ &= \frac{1}{\Pr\left(\Pi_i^{max}>\Pi_{i0}\right)}\int_0^\pi \varepsilon\tilde{\Phi}_i s^{-\varepsilon-1}\exp\left(-\left(\tilde{\Phi}_{i0}+\tilde{\Phi}_i\right)s^{-\varepsilon}\right)ds \qquad (53)\end{aligned}$$

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

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

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

$$\begin{aligned}\Pr\left(\Pi_i^{max}<\pi\middle|adopt\right) &= \int_0^\pi \varepsilon\left(\tilde{\Phi}_{i0}+\tilde{\Phi}_i\right)s^{-\varepsilon-1}\exp\left(-\left(\tilde{\Phi}_{i0}+\tilde{\Phi}_i\right)s^{-\varepsilon}\right)ds \qquad (54)\\ &= \Pr\left(\Pi_i<\pi\right) \qquad (55)\end{aligned}$$

Profits conditional on adoption from j

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

$$\Pr\left(\Pi_{ijv}<\pi\middle|adopt\ from\ j\ in\ v\right) = \frac{\Pr\left(\Pi_{ijl}<\pi\cap\Pi_{ijv}>\Pi_{ij'l}\forall(j',l)\cap\Pi_{ijv}>\Pi_{i0}\right)}{\Pr\left(\Pi_{ijv}>\Pi_{ij'l}\forall(j',l)\cap\Pi_{ijv}>\Pi_{i0}\right)} \tag{56}$$

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

$$\Pr\left(\Pi_{ijv}<\pi\middle|adopt\ from\ j\ in\ v\right) = \frac{1}{\lambda_{ijv}}\int_0^\pi Pr\left(s>\Pi_{ij'l}\forall(j',l)\cap s>\Pi_{i0}\right)g_{ijv}(s)ds \tag{57}$$

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

$$\begin{aligned} \Pr\left(s > \Pi_{ij'l} \forall (j', l) \cap s > \Pi_{i0}\right) &= \exp\left(-\tilde{\Phi}_{i0} s^{-\varepsilon}\right) \exp\left(-\tilde{\tilde{\Phi}}_{ijv} s^{-\varepsilon}\right) && (58) \\ &= \exp\left(-\left(\tilde{\Phi}_{i0} + \tilde{\tilde{\Phi}}_{ijv}\right) s^{-\varepsilon}\right) && (59) \end{aligned}$$

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

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

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

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

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

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

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

Appendix D: Production Function Estimation with and without Fertilizer

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

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

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

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

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

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

Here, DFT_{ipt} is a district-time variable, with and without fertilizer use, meant to absorb differences is local wages and prices, and other local and shocks, that may vary by time and whether fertilizer is used. While one could argue that local wages and prices should vary at a more granular level, this is about as far as we can push the data given the other sets of fixed effects that are utilized. $Plot_{ip}$is a fixed effect to absorb plot-specific sources of productivity differences. Within these fixed effects, we estimate α_0 and α using labor per unit of land and an interaction with a dummy variable identifying fertilizer use. Appendix Table A7 reports these estimates. In the preferred specification, we find that $\alpha_0 = 0.57$ and $\alpha = 0.421$.

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

Appendix E: Mark-ups

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

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

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

$$\mathbb{E}\left[v_{jv}\right] = (1-\beta)\frac{1-\alpha}{\alpha}\sum_i \lambda_{ijv}\mathbb{E}\left[\Pi_i\right] \tag{60}$$

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

$$\frac{d\mathbb{E}\left[v_{jv}\right]}{dr_{jv}}\frac{r_{jv}}{\mathbb{E}\left[v_{jv}\right]} = \sum s_{ijv}\left(\frac{d\lambda_{ijv}}{dr_{jv}}\frac{r_{jv}}{\lambda_{ijv}} + \frac{d\mathbb{E}\left[\Pi_i\right]}{dr_{jv}}\frac{r_{jv}}{\mathbb{E}\left[\Pi_i\right]}\right) \tag{61}$$

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

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

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

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

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

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

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

$$\frac{d\mathbb{E}\left[v_{jv}\right]}{dr_{jv}}\frac{r_{jv}}{\mathbb{E}\left[v_{jv}\right]} = -\varepsilon_a\sum_i s_{ijv}\left((1-\lambda_{ijv}) + \frac{1}{\varepsilon}\lambda_{ijv}\right) \tag{62}$$

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

$$\varepsilon_v \equiv \frac{d\mathbb{E}\left[v_{jv}\right]}{dr_{jv}}\frac{r_{jv}}{\mathbb{E}\left[v_{jv}\right]} = -\varepsilon_a + \frac{\varepsilon-1}{\varepsilon}\varepsilon_a\sum_i s_{ijv}\lambda_{ijv} \tag{63}$$

Table 1. Summary statistics on villages

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

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

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

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

Table 2. Calibrating Travel Costs

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	From market center to major hub town (Transport Operator Surveys)			From village center to market center					
				Enumerator's Trips			Transport Operator Surveys		
	Cost	Cost	Hours	Cost	Cost	Hours	Cost	Cost	Hours
Panel A. Costs from Markets									
Google maps: kilometers to destination	0.021***								
	(0.000)								
Google maps: hours to destination		1.261***	0.998***						
		(0.028)	(0.032)						
Number of markets	201	201	201						
Number of observations	900	900	893						
Panel B. Costs from villages									
Google maps: kilometers to destination				0.117***			0.088***		
				(0.011)			(0.009)		
Google maps: hours to destination					3.544***	0.724***		2.609***	0.839***
					(0.274)	(0.069)		(0.252)	(0.076)
Number of villages				1127	1033	1036	1133	1133	1027
Number of observations				1127	1033	1036	1133	1133	1027

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

Table 3. Remoteness and farmer characteristics

	(1)	(2)	(3)
	Mean	(Standardized) coefficient from regression of dependent variable on remoteness measure based on (population-weighted):	
		Distance to hubs	Elasticity-adjusted travel costs to hubs
Panel A. Demographic and background characteristics			
Age	49.76	-0.98*	-1.45***
	(15.23)	(0.52)	(0.50)
Female	0.45	-0.02	-0.02
		(0.02)	(0.02)
Married	0.76	0.00	0.01
		(0.01)	(0.01)
Household size	4.95	0.26**	0.36***
	(2.78)	(0.11)	(0.10)
Years of education	6.58	-0.31***	-0.47***
	(3.56)	(0.11)	(0.12)
Home has thatch roof	0.17	0.03	0.04**
		(0.02)	(0.02)
Has cell phone	0.89	-0.03***	-0.03***
		(0.01)	(0.01)
Has bank account	0.15	-0.05***	-0.05***
		(0.01)	(0.01)
Has mobile money account	0.77	-0.08***	-0.08***
		(0.02)	(0.01)
Acres of land	5.46	1.37**	2.65***
	(13.89)	(0.57)	(0.68)
Has market business	0.28	-0.05***	-0.06***
		(0.01)	(0.01)
Annual total income from non-farming (USD)	408.9	-74.72**	-87.91***
	(772.60)	(30.25)	(28.88)
Panel B. Production Capacity (in kg/acre)[1]			
FAO-GAEZ production capacity for low input level	788.3	70.07***	53.84***
	(290.70)	(21.21)	(19.10)
FAO-GAEZ production capacity for high input level	3325	-296.16***	-291.20***
	(876.00)	(57.38)	(58.96)
FAO-GAEZ production difference between high and low	2536	-366.23***	-345.04***
	(744.90)	(46.71)	(49.04)
Panel C. Harvest Output			
Total harvest output in 2016 long rains (kg)	928.7	-16.62	136.62**
	(1360.00)	(51.18)	(54.35)
Harvest output per acre	455.3	-85.11***	-82.61***
	(384.30)	(17.46)	(15.60)
Value of harvest output at average regional post-harvest price	201.9	-3.61	29.70**
	(295.60)	(11.13)	(11.82)

Notes: N = 2,845 farmers in 246 villages. In Column 1, standard deviations are in parentheses. Columns 2 and 3 show regression coefficients from separate regressions of the dependent variable on a measure of remoteness (equations 3 and 4 in the paper). See text for further discussion of these measures. In those columns, standard errors in parentheses, clustered at the village level.
*, **, and *** indicate significance at 10%, 5%, and 1%.
[1]Regressions for production capacity are at village level.

Table 4. Remoteness, access to input markets and retail price heterogeneity

	(1)	(2)	(3)
	Mean	(Standardized) coefficient from regression of dependent variable on remoteness measure based on (population-weighted):	
		Distance to hubs	Elasticity-adjusted travel costs to hubs
Panel A. Summary measures of access to input retailers			
Has at least 1 agrovet within 10 km of village	0.75	-0.14***	-0.13***
which sells fertilizer or seeds		(0.01)	(0.01)
Number of agrovets within 10 km of village	7.79	-2.93***	-4.23***
which sells fertilizer or seeds	(8.96)	(0.25)	(0.23)
Distance to nearest agrovet	6.79	3.17***	2.69***
which sells fertilizer or seeds	(15.15)	(0.47)	(0.46)
Panel B. Travel-cost adjusted prices faced by farmers			
Minimum travel-cost adjusted price	24.19	2.33***	2.45***
for 50 kg of Urea (USD)[1]	(4.66)	(0.12)	(0.12)
Decomposition of price between retail price and cost of transportation			
Retail price at the location with the lowest	19.82	1.09***	1.28***
travel-cost adjusted price (USD)	(2.63)	(0.07)	(0.07)
Cost of travel to obtain minimum travel-cost	4.372	1.24***	1.17***
adjusted price (USD)	(4.39)	(0.13)	(0.13)

Notes: The unit of observation is the village. Data is from the universe of villages in Kilimanjaro and Manyara regions (N = 1,183). Travel costs imputed from transport surveys and Google maps. In Column 1, standard deviations are in parentheses. Columns 2 and 3 show regression coefficients from separate regressions of the dependent variable on a measure of remoteness (equations 3 and 4 in the paper). See text for further discussion of these measures. In those columns, standard errors in parentheses, clustered at the village level.

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

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

Table 5. Remoteness, access to output markets and output price heterogeneity

	(1)	(2)	(3)
	Mean	(Standardized) coefficient from regression of dependent variable on remoteness measure based on (population-weighted):	
		Distance to hubs	Elasticity-adjusted travel costs to hubs
Panel A. Summary measures of access to output markets			
Has at least 1 maize seller within 10 km of village	0.67	-0.16***	-0.17***
		(0.01)	(0.01)
Number of maize sellers within 10 km of village	1.89	-1.06***	-1.44***
	(2.48)	(0.07)	(0.06)
Distance to nearest output market with maize sellers (km)	8.67	5.65***	4.75***
	(14.17)	(0.42)	(0.42)
Panel B1. Maximum imputed travel-cost adjusted price if farmers were to sell in a local market			
Market survey: maximum travel-cost adjusted price immediately after 2017 harvest (USD)[1]	30.30	-3.08***	-3.19***
	(7.24)	(0.20)	(0.20)
Decomposition of price between retail price and cost of transportation			
Retail price at the location with the highest travel-cost adjusted price (USD)	39.34	0.80***	0.22**
	(3.17)	(0.09)	(0.10)
Cost of travel to obtain the highest travel-cost adjusted price (USD)	9.05	3.88***	3.41***
	(7.06)	(0.18)	(0.19)
Panel B2. Price available within village by maize-buying intermediaries immediately after last season's harvest			
Farmer surveys: average "going price" in local village immediately after previous harvest[2]	25.86	-1.31**	-2.68***
	(6.24)	(0.52)	(0.48)
Farmer surveys: average village sales price after previous harvest[2]	30.38	-1.88**	-3.61***
	(8.60)	(0.75)	(0.70)

Notes: The unit of observation is the village. Data is from the universe of villages in Kilimanjaro and Manyara regions (N = 1,183). Travel costs imputed from transport surveys and Google maps. In Column 1, standard deviations are in parentheses. Columns 2 and 3 show regression coefficients from separate regressions of the dependent variable on a measure of remoteness (equations 3 and 4 in the paper). See text for further discussion of these measures. In those columns, standard errors in parentheses, clustered at the village level. *, **, and *** indicate significance at 10%, 5%, and 1%.

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

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

Table 6. Remoteness and input market access and adoption

	(1)	(2)	(3)	(4)	(5)
		(Standardized) coefficient from regression of dependent variable on remoteness measure based on (population-weighted):			
	Mean	Distance to hubs		Elasticity-adjusted travel costs to hubs	
		No controls	Controls for soil and farmer characteristics	No controls	Controls for soil and farmer characteristics
Panel A: Input usage					
Used chemical fertilizer in previous long rains	0.39	-0.17***	-0.09***	-0.20***	-0.13***
		(0.03)	(0.03)	(0.03)	(0.03)
Quantity of chemical fertilizer used (kg)	19.84	-13.06***	-6.46***	-14.21***	-8.83***
	(31.63)	(2.15)	(1.74)	(1.95)	(1.84)
Used improved seeds in previous long rains	0.66	-0.07***	-0.05**	-0.11***	-0.10***
		(0.02)	(0.02)	(0.02)	(0.03)
Quantity of improved seeds used (kg)	6.29	-1.30***	-1.21***	-1.06***	-0.97**
	(8.21)	(0.36)	(0.44)	(0.33)	(0.43)
Panel B. Maize sales					
Sold maize after previous long rains	0.32	-0.09***	-0.06**	-0.07***	-0.04
		(0.02)	(0.03)	(0.02)	(0.03)
Total quantity sold (kg)	388.1	-97.86***	-112.16**	-4.08	-16.78
	(1142.00)	(35.11)	(48.86)	(41.06)	(48.90)
Sales to agents at home					
Agent visited homestead	0.31	-0.14***	-0.09***	-0.12***	-0.07*
		(0.03)	(0.03)	(0.03)	(0.04)
Sold maize to an agent after previous long rains	0.17	-0.07***	-0.04**	-0.04***	-0.02
		(0.02)	(0.02)	(0.01)	(0.02)
Quantity sold to agents (kg)	142	-46.39***	-39.11**	-12.54	4.12
	(433.70)	(13.80)	(18.61)	(12.73)	(19.36)
Sales at market					
Sold maize at a market after previous long rains	0.06	-0.03***	-0.03***	-0.02***	-0.02**
		(0.01)	(0.01)	(0.01)	(0.01)
Quantity sold at market (kg)	34.42	-14.61***	-15.26**	-9.55	-10.93
	(197.10)	(5.32)	(7.72)	(6.16)	(7.52)
Panel C. Maize purchases					
Farmer ever buys maize	0.48	0.11***	0.08***	0.11***	0.09***
		(0.02)	(0.02)	(0.02)	(0.02)
Quantity purchased in typical year (kg)	152.3	75.67***	65.05***	93.94***	80.19***
	(315.50)	(15.79)	(17.92)	(16.29)	(14.77)
Net buying					
Farmer buys maize but sells none	0.37	0.11***	0.08***	0.11***	0.08***
		(0.02)	(0.03)	(0.02)	(0.02)
Farmer sells maize and buys none	0.24	-0.09***	-0.06***	-0.08***	-0.06***
		(0.02)	(0.02)	(0.01)	(0.02)
Farmer buys and sells maize	0.08	-0.00	0.00	0.01	0.02
		(0.01)	(0.01)	(0.01)	(0.01)
Net buyer (quantity bought > quantity sold)	0.44	0.10***	0.07**	0.11***	0.08***
		(0.02)	(0.03)	(0.02)	(0.03)
Net seller (quantity bought < quantity sold)	0.26	-0.08***	-0.06**	-0.06***	-0.04
		(0.02)	(0.03)	(0.02)	(0.03)

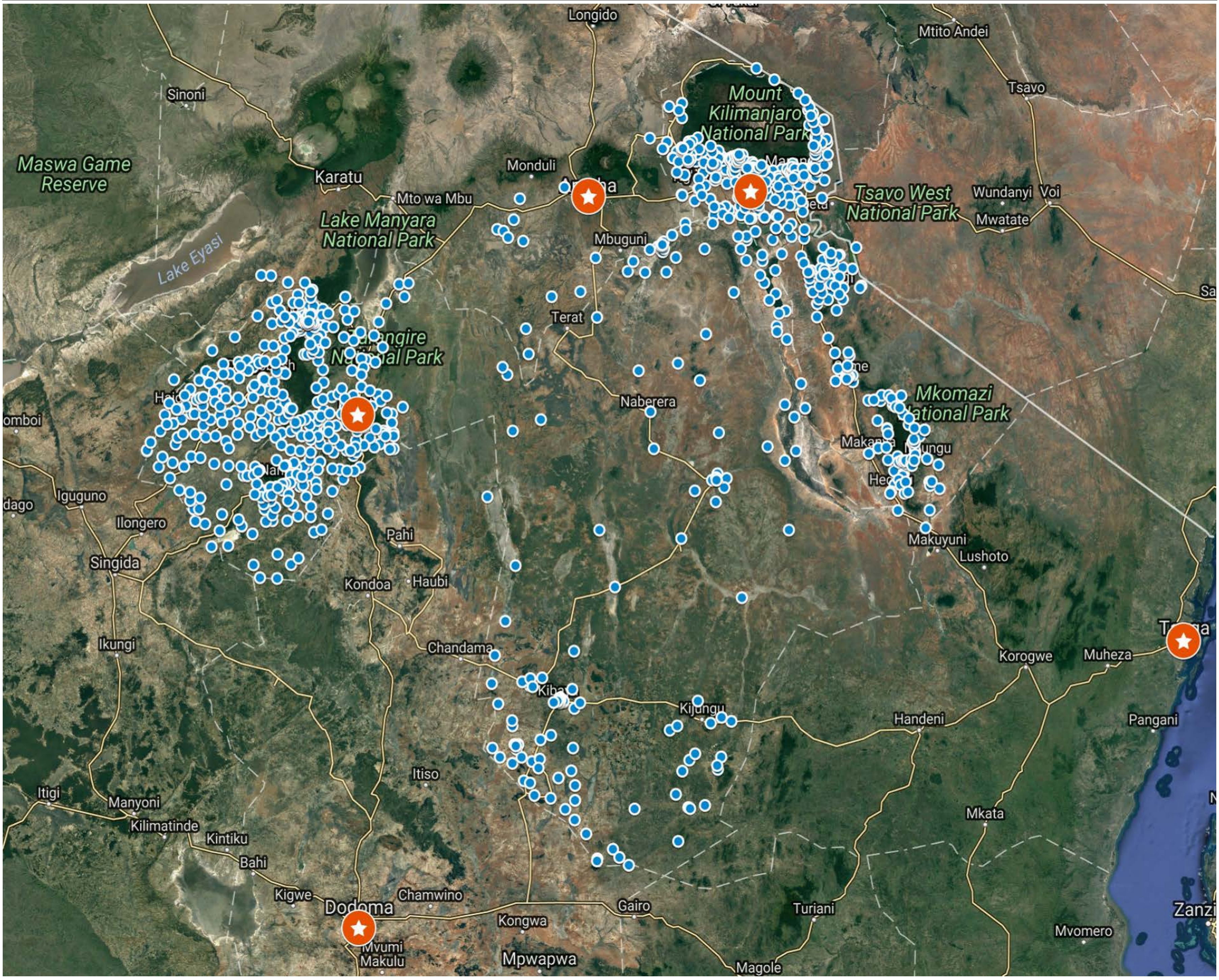
Notes: N = 2,845 farmers in 246 villages. See text for sampling details. Standard deviations are in parentheses in Column 1. Columns 2-5 show regression coefficients from separate regressions of the dependent variable on a measure of remoteness (equations 3 and 4 in the paper). See text for further discussion of these measures. In those columns, standard errors in parentheses, clustered at the village level.
*, **, and *** indicate significance at 10%, 5%, and 1%.

Table 7. Multinomial logit of agrovet choice

	(1)	(2)	(3)
	Agrovet Chosen		AVE/KM
Kilometers to agrovet	-0.171***		3%
	(0.009)		
Dummies for agrovet distance bin:			
between (0,5] km		-1.380***	5.7%
		(0.372)	
between (5,10] km		-2.914***	5.8%
		(0.379)	
between (10,15] km		-4.331***	5.8%
		(0.380)	
between (15,20] km		-5.367***	4.5%
		(0.400)	
between (20,30] km		-5.875***	4.3%
		(0.378)	
between (30,40] km		-7.602***	3.9%
		(0.449)	
between (40,50] km		-8.685***	3.5%
		(0.495)	
between (50,100] km		-10.625***	2.1%
		(0.560)	
over 100 km		-14.253***	-
		(0.992)	

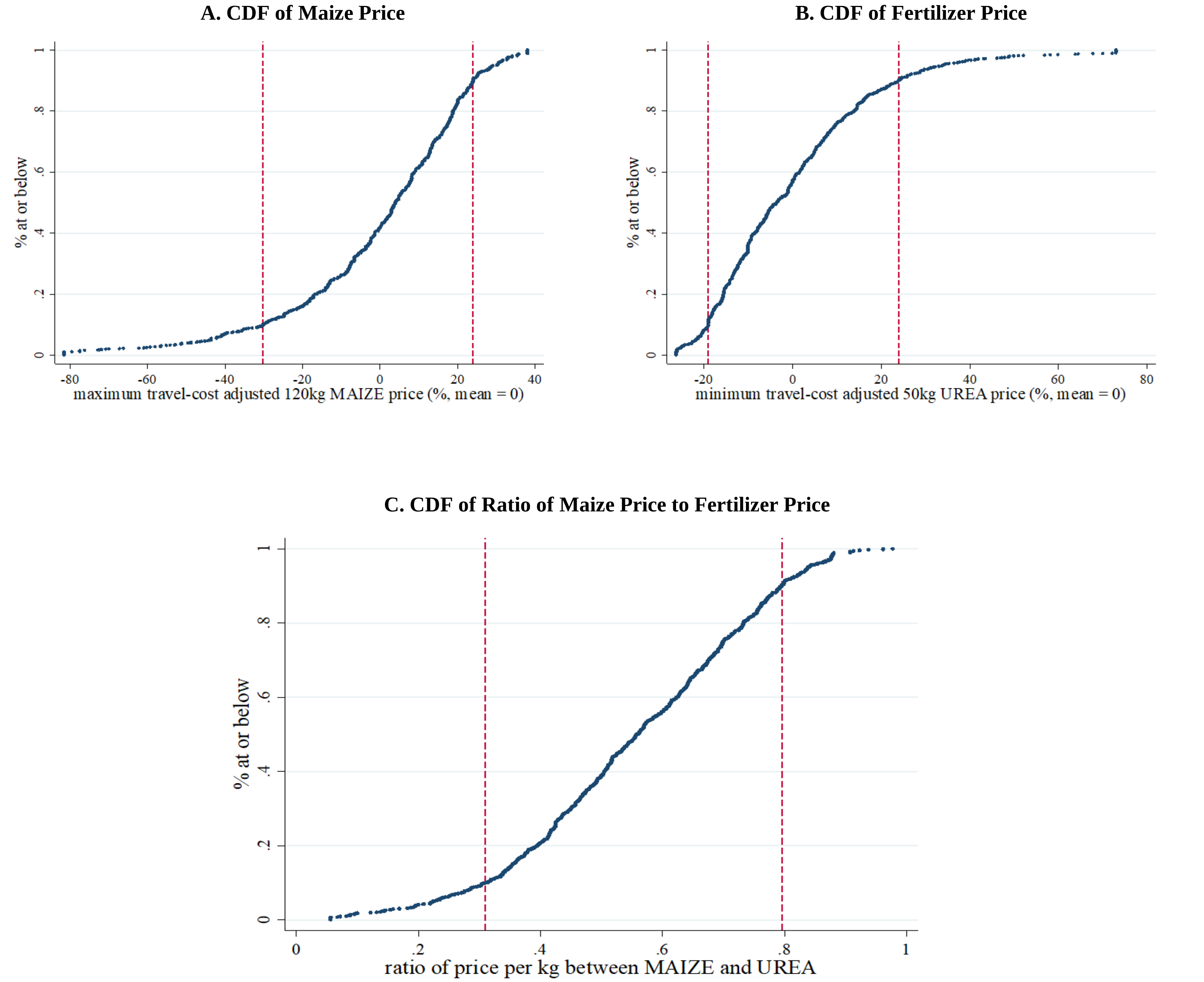
Notes: N = 519 farmers, 119 observed locations. Omitted group is agrovet located in respondent's village. Ad-valorem equivalent per kilometer is calculated at the upper bound of each bin, and assumes that the trade cost compounds each kilometer. Standard errors in parentheses. *, **, and *** indicate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

Figure 1. Map of Survey Region and Villages



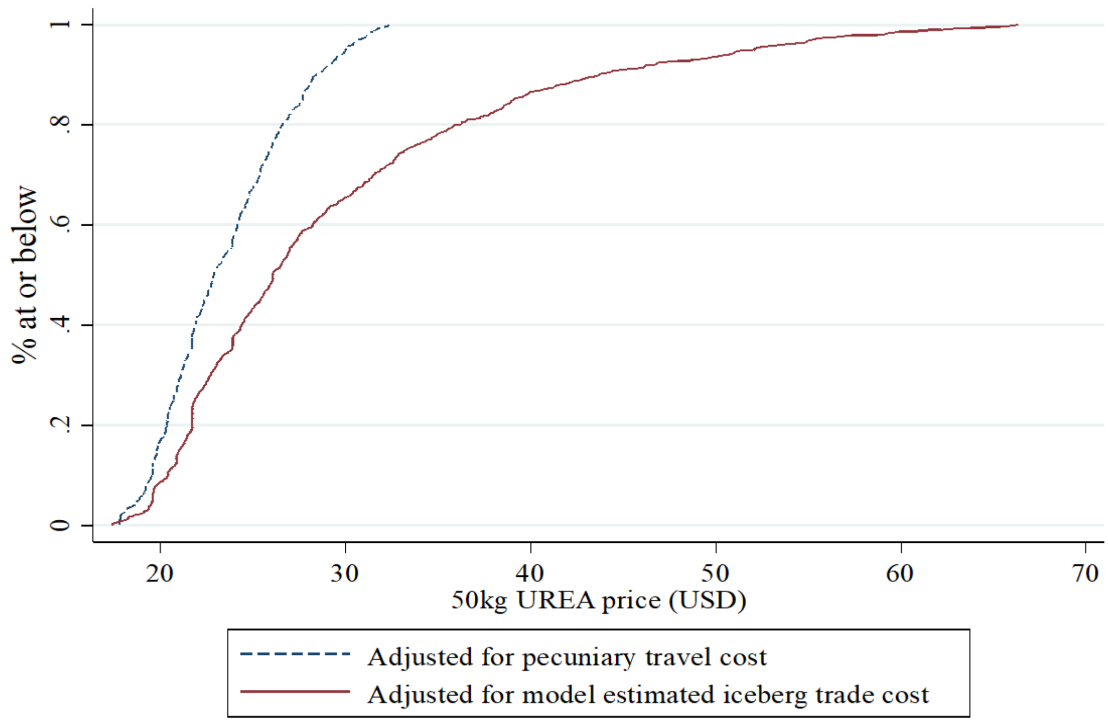
Notes: Blue dots represent all villages in the Kilimanjaro and Manyara Regions. The star signs represent the five major hubs that are used to construct our market access proxies in Section 4.1. They are Moshi, Arusha, Babati, Dodoma, and Tanga.

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. Vertical dotted lines represent a 10 percentile and a 90 percentile of the distribution.

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 and transport cost information collected from interviews with transport operators. The CDFs are trimmed at a 95 percentile for a better visual comparison.

Figure 4. Market Access vs. Remoteness

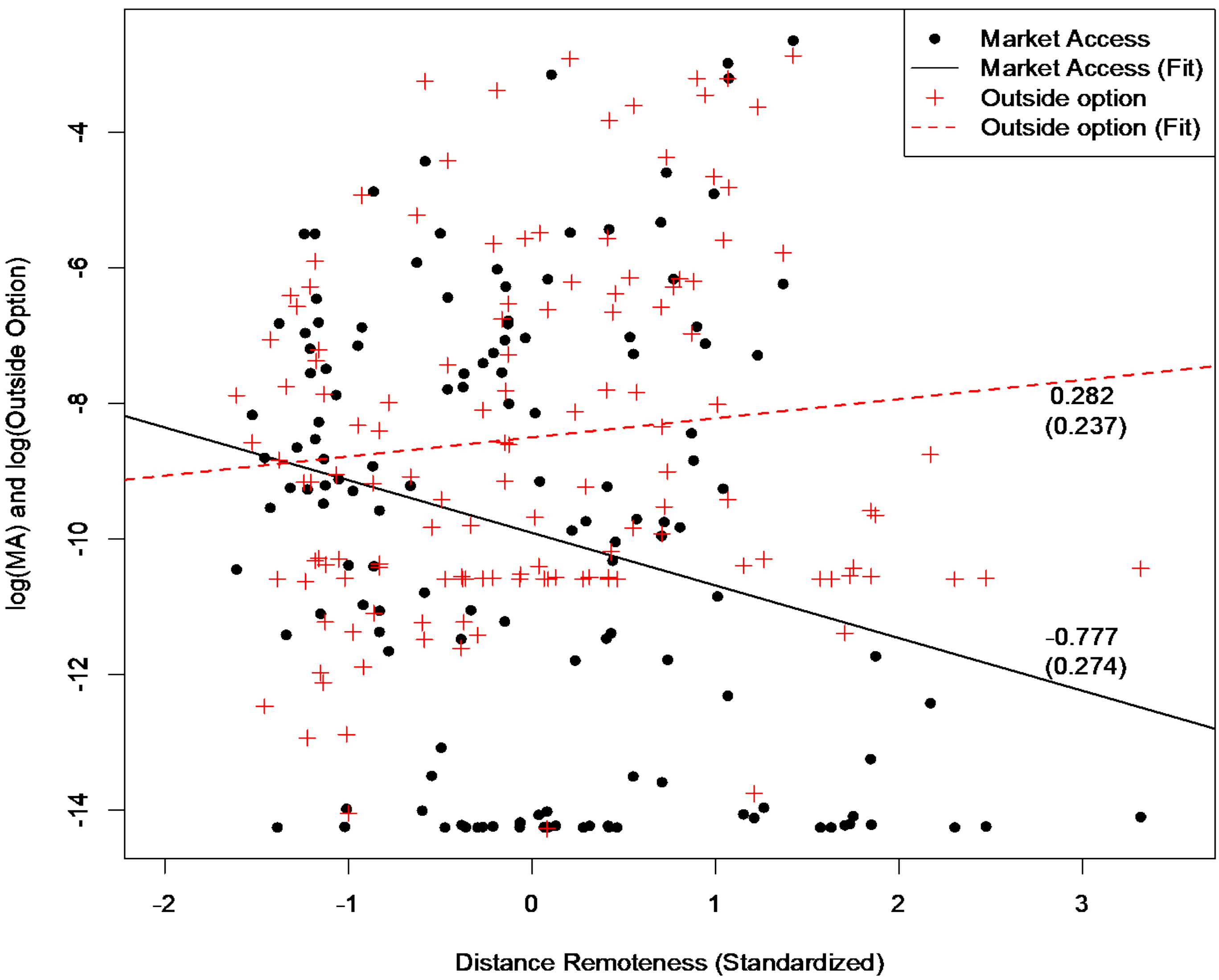
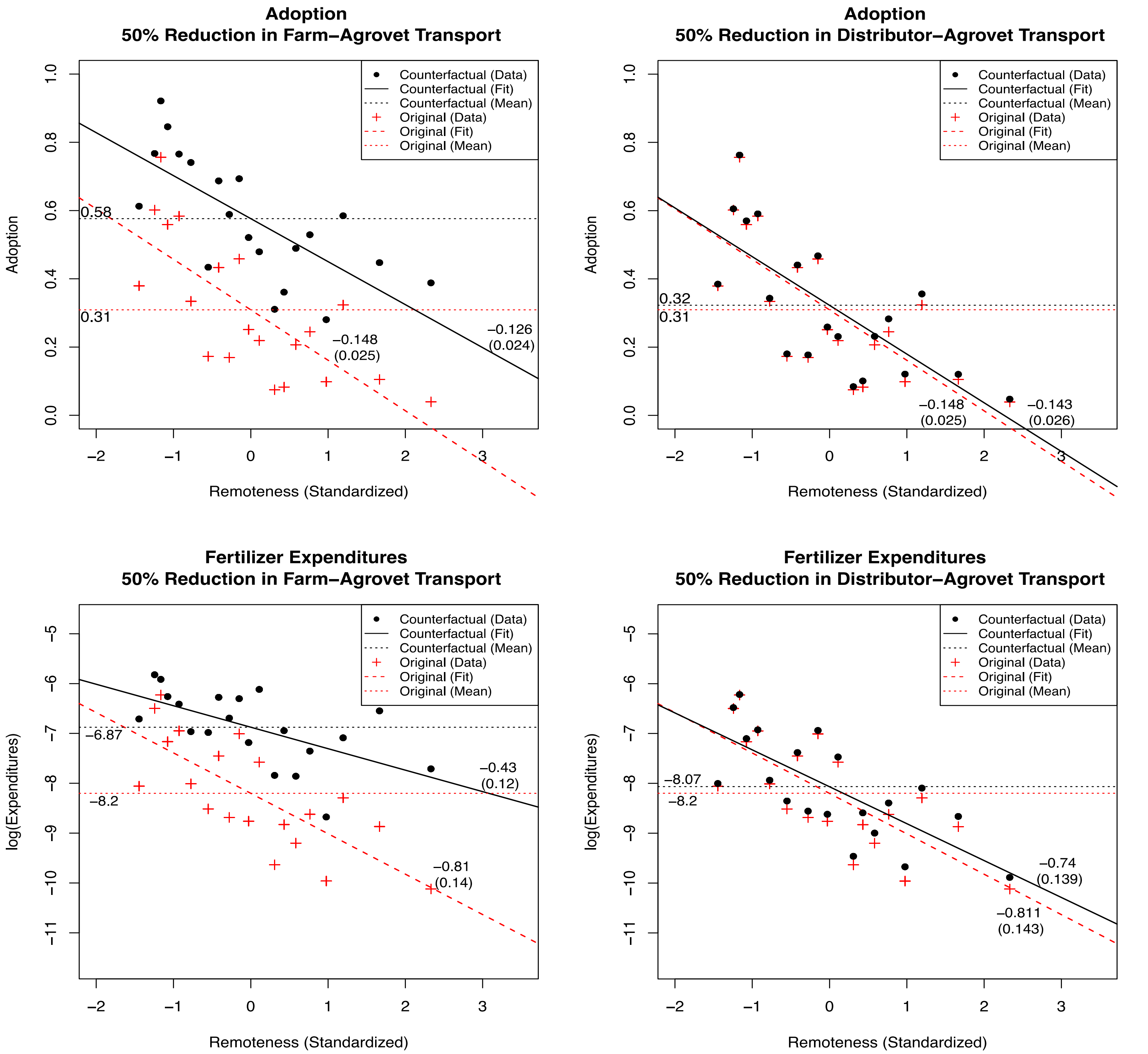
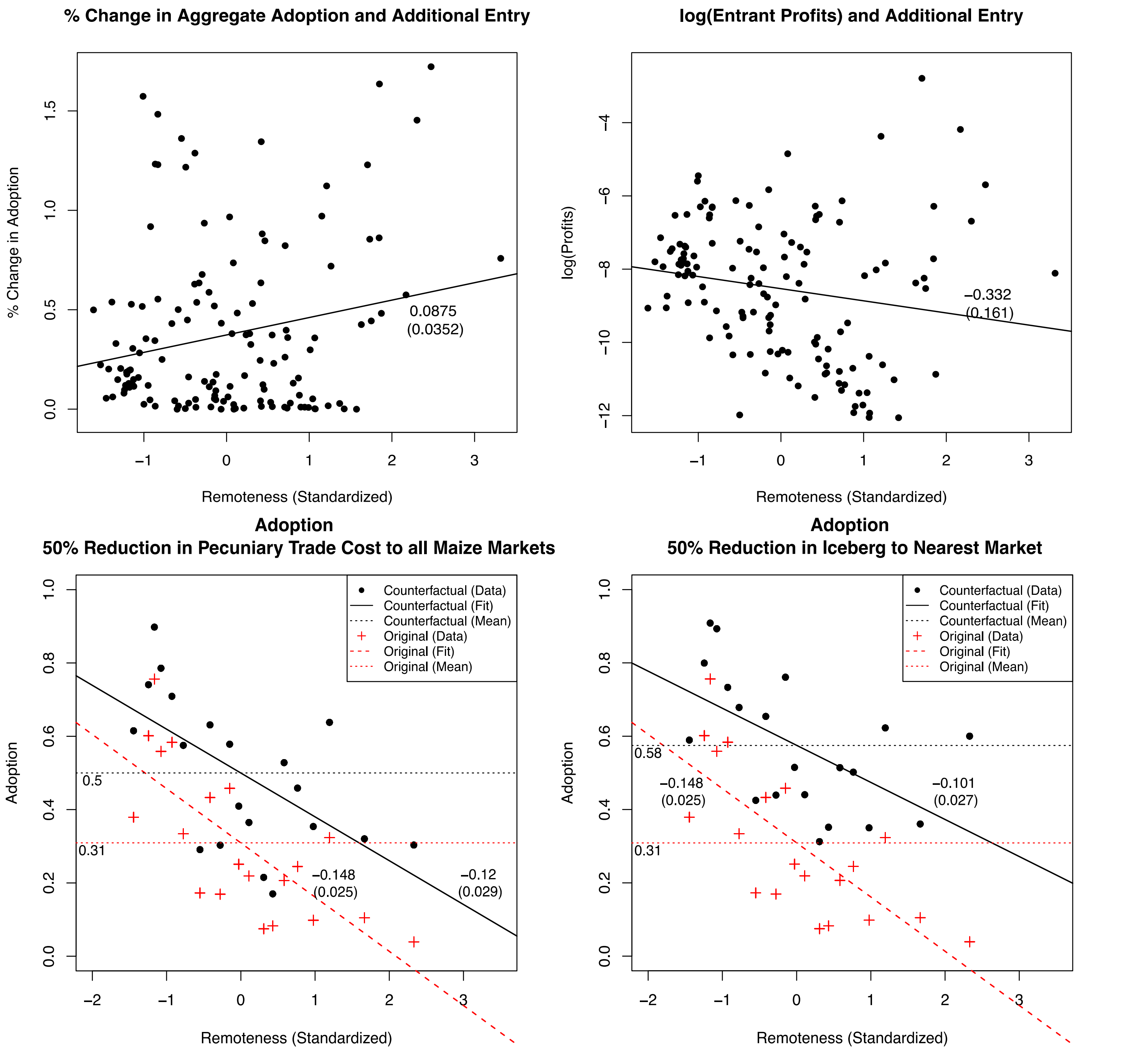


Figure 5. Input Access Counterfactuals



Notes: See text for discussion of counterfactuals.

Figure 6. Entry and Output Access Counterfactuals



Notes: See text for discussion of counterfactuals.

Web Appendix Table A1. Survey Compliance Rates

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

Notes: See text of details of surveys.

Web Appendix Table A2. Costs of transporting fertilizer and transporting farmer, by distance

	(1)	(2)	(3)	(4)
	Cost of transporting fertilizer from agrovet in destination village (standardized to 50 kg)		Cost of farmer traveling himself to agrovet	
Google maps: kilometers to destination	0.036***		0.047***	
	(0.020)		(0.007)	
Google maps: hours to destination		1.276***		1.831***
		(0.677)		(0.260)
Number of villages	73	73	119	119
Number of observations	341	341	988	988

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

Web Appendix Table A3. Locations of Input and Output Distributors

Panel A. Locations of Agro-Input Distributors

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

Panel B. Locations of Output Distributors

B1. 2017 Maize Store Census

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

B2. 2016 Maize Store Census

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

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

Web Appendix Table A4. Summary Statistics of Market Access Proxies

	(1)
	Remoteness measured by distance
Remoteness measured by elasticity-adjusted travel costs to hubs	0.842***
	(0.016)
Dependent variable mean before standardization	304.19
Dependent variable sd before standardization	31.56
Independent variable mean before standardization	-0.11
Independent variable sd before standardization	0.04
Observations	1,135

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

Web Appendix Table A5. Remoteness and fertilizer retailer sales, prices, and markups

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

Notes: In Column 1, standard deviations are in parentheses. Columns 2 and 3 show regression coefficients from separate regressions of the dependent variable on a measure of remoteness (equations 3 and 4 in the paper). See text for further discussion of these measures. Regressions in Panel B includes type and brand fixed effects.
*, **, and *** indicate significance at 10%, 5%, and 1%.
[1]Markup accounts for cost of transport to wholesaler.

Web Appendix Table A6. Robustness of Travel-cost Adjusted Prices

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

Notes: Data is from the universe of villages in Kilimanjaro and Manyara region (N = 1183). The unit of observation is the village. Travel costs imputed from transport surveys and Google maps. In Column 1, standard deviations are in parentheses. Columns 2 and 3 show regression coefficients from separate regressions of the dependent variable on a measure of remoteness (equations 3 and 4 in the paper). See text for further discussion of these measures.

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

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

Web Appendix Table A7. Elasticity Estimation for Calibration

	(1)	(2)	(3)	(4)
	Dependent variable: log(Eta) (from Calibration)			
log(Price)	-5.566***	-5.532***	-4.731***	-4.911**
	(0.990)	(2.001)	(0.960)	(1.959)
log(Experience)			0.938***	0.973***
			(0.154)	(0.260)
	OLS	IV	OLS	IV
R-squared	0.37	0.35	0.43	0.4
First-stage F Statistic		246.35		242.68
Wu-Hausman		0.33		0.19
Observations	374	242	374	242

Notes: District fixed effects used in all regressions. ***, **, and * indicate significance at 1%, 5%, and 10%.

Web Appendix Table A8. Production Function Estimates with and without fertilizer

	(1)	(2)	(3)
	Dependent variable: log(Harvest/Acres)		
log(Labor/Acres)	0.419***	0.433***	0.430***
	(0.042)	(0.042)	(0.042)
log(Labor/Acres) x Used Fertilizer?	0.124*	0.124*	0.149*
	(0.075)	(0.074)	(0.077)
Used Fertilizer?	-0.327	-0.332	
	(0.300)	(0.298)	
District-Year fixed effects		X	
District-Year-Fertilizer Use fixed effects			X
Plot fixed effects	X	X	X
Observations	3,395	3,395	3,395
Plots	2,554	2,554	2,554

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

Web Appendix Table A9. Input and output market price dispersion across countries

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

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

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

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

Web Appendix Table A10. 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	0.000	0.03***	0.000	0.000	0.01***	0.000	0.010
	(0.002)	0.000	(0.010)	(0.002)	0.000	(0.015)	(0.002)	0.000	(0.014)
Log (travel time)		0.03***	0.03***		0.04***	0.04**		0.01***	0.000
		(0.002)	(0.011)		(0.003)	(0.017)		(0.002)	(0.016)
Products	All	All	All	Maize	Maize	Maize	Fertilizer	Fertilizer	Fertilizer
Dependent variable mean	0.21	0.21	0.21	0.20	0.20	0.20	0.11	0.11	0.11
Dependent variable sd	0.20	0.20	0.20	0.17	0.17	0.17	0.13	0.13	0.13
Observations	4,752,196	4,752,196	4,752,196	675,880	675,880	675,880	38,364	38,364	38,364
Number of locations	1335	1335	1335	1335	1335	1335	1335	1335	1335
Countries	49	49	49	43	43	43	18	18	18
Panel B. Northern 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 Web Appendix Table A3 and text for discussion of datasets.
Two-way clustered standard errors in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10%.

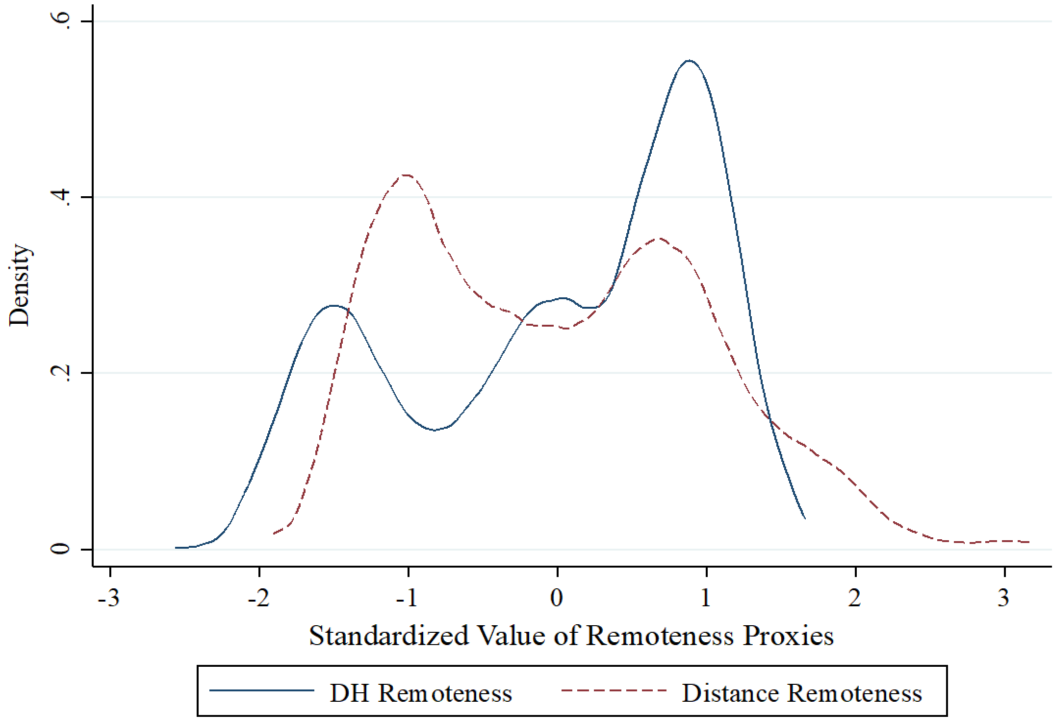
Web Appendix Table A11. Adoption in LSMS-ISA surveys

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

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

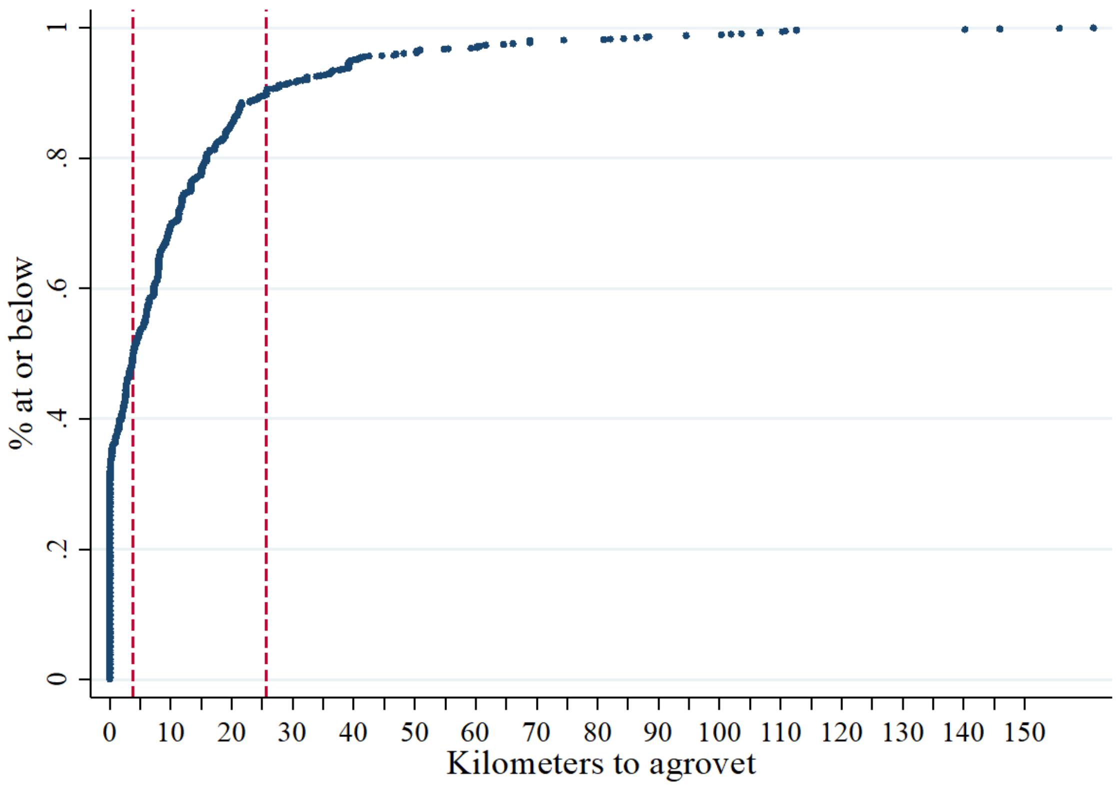
***, **, and * indicate significance at 1%, 5%, and 10%.

Web Appendix Figure A1. The Distribution of Remoteness Proxies



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

Web Appendix Figure A2. CDF of Distance Farmers Travel to Purchase Inputs



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