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

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

November 1, 2018

INCOMPLETE DRAFT: PLEASE DO NOT CITE OR CIRCULATE THIS VERSION

Abstract

In this paper, we quantify poor market access in rural Africa, 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 40-60% higher 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 travel costs of 6% ad-valorem per kilometer of travel, which is equivalent to 40% 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) doubles adoption and reduces the adoption-remoteness gradient by 18%.

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, and Stanford University 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 all 585 agrovets that sell fertilizer;[2] (3) a retrospective panel of monthly 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 6 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, plus Dar es Salaam.

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 the weighted average of travel time from a set of local regional hubs. We find striking correlations between remoteness 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 reduction in the probability of selling maize. Put differently, we find that input usage in the most remote villages in our study sample are approximately only a third of that in the least remote villages, while maize sales are only 45 percent as high.

[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 of these agrovets (either via a short census or a longer survey). For the remaining agrovets, we impute prices as described later.

While we find clear evidence of reduced market access, and while it is also clear 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 as well. To perform this quantification absent experimental variation, our third contribution is to develop a 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 cost for the farmer to reach a particular agrovet, and by increasing the costs of farmers to reach the local market to sell their harvest.

On the input side, the structure of the model (which is based on Eaton and Kortum, 2002) facilitates a decomposition of choosing an agrovet into 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 also allow us to calibrate (1) local non-price factors that may affect adoption, and (2) the implied trade cost of choosing an agrovet. 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 issues like experience in selling inputs and likelihood of stockouts). The results suggest that transportation costs are large; a non-parametric preferred specification yields estimates of local iceberg cost that are, at the conservative end, 6% ad-valorem per kilometer of travel. This is approximately 40% 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 the risk of 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 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.

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 a reduction in travel times due to road upgrading (Casaburi, Glennerster and Suri 2013). This policy roughly doubles adoption compared with baseline , and also reduces the distance gradient by 18%. 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 2 percentage points, or 6.4%, but yields only 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 mildly larger effect on adoption, the profitability from doing so falls by 0.46 log points with a 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% in reaching output markets, and find a large but slightly smaller change in adoption and a similar 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 far 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, documenting 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 whole host of outcomes other than just prices, including consumption diversity, farm investments, human capital investment, migration, and occupation choice (Aggarwal, 2018a; b; Adukia et al., 2016; Morten and Oliveira 2016; Asher and Novosad 2016; Brooks and Donovan, 2017). In our paper, we focus narrowly on the specific effect of transportation costs on market access (i.e. the actual time and money costs of transportation and the presence of intermediaries and the prices they charge) in isolation, without changing other margins.[5]

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).

[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 (2018), Alder (2017), Adukia et al. (2016), Asher and Novosad (2016), Banerjee et al. (2012), Bird and Straub (2016), Bryan and Morten (2017), Gertler et al. (2014), Ghani et al. (2016), Khanna (2016), Shamdasani (2016), and Storeygard (2016). See Donaldson (2016) for a review.

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

Our paper is most closely related to Atkin and Donaldson (2015), who estimate trade costs in a situation in which an oligopolist intermediary buys products at wholesale prices, transports them to distant markets and sells them directly to consumers. By contrast, we are interested in how trade costs affect the buying decisions of final consumers (in this case, farmers), as well as pricing decisions by retailers. [6]

Finally, much of the trade literature, which has documented larger gains from integration when there are input-output relationships (e.g. Yi, 2001; Costinot and Rodriguez-Clare, 2014; Sotelo, 2016) has only evaluated economies under the assumption of monopolistically competitive or purely competitive sectors at a fairly aggregate level (e.g. international trade by industry).[7] By contrast, our model yields a structural discrete choice problem in which farmers choose whether to adopt, and if so, choose the best agrovet from which to purchase fertilizer from a potentially concentrated group of agro-retailers. [8]

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[9] of Northern Tanzania. The two regions are a combined 57,000 square-kilometers (6% of the land mass of Tanzania), contain 1,183 villages, and had a population of 3.1 million in 2012 (National Bureau of Statistics, 2013). Compared to developed countries, the quality of roads in Kilimanjaro and Manyara is 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

[6] Though not directly comparable since they are at different points in the supply chain, our average ad-valorem "travel costs" of farmers procuring fertilizer, as calculated through interviews with transport providers, turn out to be similar to those of the intermediaries in Atkin and Donaldson (2015). Our costs, however, are calculated over a much shorter trip. Specifically, to find the best travel-adjusted price for fertilizer, our results suggest that for the typical village, the best option is 10km away. In Atkin and Donaldson, ad-valorem estimates are calculated based on the cost difference of a trip to the most remote location (500 miles away) relative to the least remote location (50 miles away), which is approximately a 720km difference.

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

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

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

overall (TanRoads and PMO-RALG, 2014), compared to 68 percent in the US and an OECD average of 134 percent[10]

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

We have four main sources of data we use in this draft: agrovet surveys, farmer surveys, transport surveys, and maize price surveys. All were collected from January 2016 to December 2017 in Kilimanjaro and February to May 2018 in Manyara.[11]

[10]Information compiled from various online 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.

[11]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.

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. agrovets. 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, prices, quantities, and the wholesale costs of acquiring stock from the distributor. The survey took care to differentiate fertilizer types by distributor, brand, and type – thus the level of granularity should be akin to the barcode-level. The survey also included a number of questions about costs of travel to the distributor, as well as some background characteristics.

3.2 Farmer surveys

We conducted farmer surveys in 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[12], 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.[13]

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 GPS location for every village in Kilimanjaro,[14] 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

[12]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.

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

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

to travel to the major towns in Kilimanjaro (Arusha and Moshi), the capital city (Dar-es-Salaam) and the market center. In Manyara, 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 would travel for some distance along a major trunk road, and then turn off to travel for some time on unpaved feeder roads and village roads. 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 distance, travel time, and road quality. We use this data to correlate costs of travel with road quality, and to estimate the percentage of roads which are paved (to inform later counterfactuals).

3.4 Maize prices

To measure maize prices, we visited markets post-harvest in September and October of 2017 in Kilimanjaro and February to May 2018 in Manyara. During the visits, enumerators sampled up to 3 maize sellers per market to document pre- and post-harvest prices for maize during recent seasons. These data allow us to compare prices across markets at the same point in time, though they are not intended to be used in panel analysis.

3.5 Summary statistics 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 2,842 people (see table notes), and is located 11.3 kilometers from the nearest market center. It takes about 40 minutes by vehicle to reach the market and a round-trip costs about $1.92 on average. The average village is over 70 km away from a major hub, and travel there would take almost 3 hours 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 1/3 paved, 1/3 dirt, and 1/3 gravel, and travel times according to google are fairly slow: 30.6 km/hour on rural roads compared to 49.5 km/hr on the main roads.

4 Results

4.1 Defining remoteness

We define remoteness as proximity to major hub towns, where distributors are located. There are a number of potential hubs, but they differ in their importance. To capture this, we weight proximity by the population of the town – the rationale for this is that bigger population centers will have

more distributors than smaller ones. Our measure therefore becomes

$$remoteness_v = \sum_h \frac{d_{hv}}{pop_h} \tag{1}$$

where $pop_h$is the population of hub h and d_{hv} is distance from village v to that hub. For interpretation, we standardize this measure to have mean 0 and standard deviation 1. The distribution of this variable can be seen 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 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)

4.2 Calculating travel cost-adjusted prices

For each village, we assume that farmers are free to travel to any market to buy or sell, but must incur a transportation cost, which we calibrate using information from transport surveys and google distances. 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. We present these results in Table 2. Columns 1-3 show 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 \$1.20 per hour of travel. The remaining columns replicate this on rural roads, where we find higher costs for rural travel.

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

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

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 fertilizer itself. 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 A1). We do this for a 50 kg bag of fertilizer, the modal amount purchased by farmers. We find that a 50 kg bag of fertilizer costs about 75% as much as transporting a person, implying therefore that a farmer must make 2.75 trips to buy a bag (2 for himself and 0.75 for the bag).

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

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

Here, p_m is the price of maize post-harvest for market m, and $c_{mv}$is the cost of traveling from village v to market m. We use a 120 kg bag for this calculation, and assume that the cost of transport is twice that of fertilizer (i.e. proportional to the size).

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. 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 show 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 15% of the mean, and the ratio of this price between the 90th and the 10th percentile ratio is similar about 40% of the mean. Panel B shows the distribution of transportation cost-adjusted fertilizer prices, where the standard deviation in travel cost-adjusted prices is 15% of the mean, and ratio of this price between the 90th and the 10th percentile is equivalent to about 60% of the mean.

4.3 Reduced form analysis

4.3.1 Specification

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

$$m_{fvt} = \theta remoteness_v + \beta X_{fvt} + \epsilon_{fvt} \tag{4}$$

where m_{vt} is a measure of market access (or related outcome) at location v in year t.

The choice of controls varies based on the outcome of interest. For market access measures, 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. Thus for the farmer results, we present with and without a vector of other controls, $\mathbf{X_{fvt}}$. These controls include a host of characteristics from the survey, such as land ownership, income, asset ownership, education and other demographic characteristics, as well as soil information from the FAO-GAEZ.

4.3.2 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 worse off in many ways. Farmers in rural 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. These differences must be accounted for when doing policy counterfactuals.

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 rural areas - yield increases a standard deviation away decline to a 250% increase in yields. Comparing the most and least remote areas (roughly from -2 to 2 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. For this reason, we control for these measures in our main regressions.

Finally, Panel C shows harvest output from the most recent long rains. Total yields are similar by remoteness, but yield per acre is lower in rural areas (since farm size is larger). In particular, 1 standard deviation increase in 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 characteristics. Ultimately this motivates the use of an economic model to conduct counterfactuals.

4.3.3 Access to input markets

Table 4 shows how remoteness correlates 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 60% of purchases are made within 10 km of a farmer's village, and 80% 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 in remote villages, all significant at 1%.

Panel B shows our key measure of access, travel cost-adjusted prices. We find that one standard deviation of remoteness raises prices by about $2.4, equivalent to about 10% of the mean. This implies a difference in prices of approximately 50% between the most and least remote villages in our 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.

To investigate this further, we look at agrovet pricing in Web Appendix Table A3. We note that these results should be taken with some caution, since they are conditional on entry (and so the agrovet location is not random), 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 rural areas (naturally), since they must travel further.

Panel B shows the relationship on prices and markups. Note that these regressions are the fertilizer type level (i.e. product barcode). We find here that rural shops face similar wholesale prices (consistent with the above) but charge higher retail prices. We then calculate markups, inclusive of transport costs. We find limited evidence of higher markups in rural areas – markups are on average about 13% and increase by about 1 percentage point per standard deviation of remoteness (significant at only 10%). From this, we conclude that much of the higher retail prices are a result of higher marginal costs of accessing inputs.

4.3.4 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 B shows travel cost-adjusted prices for maize. Since there are large seasonal price fluctuations in rural Tanzania as in much of rural Africa, 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 do sell sell shortly after harvest. We find that travel cost-adjusted prices are lower in remote areas. As before, we decompose this into the retail price and the travel costs, finding that while retail maize prices rise modestly, transport costs to the closest maize market rise by $4 with each standard deviation in remoteness.

Finally, we show 2 other measures of price, 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. Both of these show no correlation with distance.

4.3.5 Farmer decisions

The results so far show clear evidence of reduced market access on both input and output markets, and clear evidence of higher prices for inputs though more mixed evidence for output. These results should imply lower input usage and maize sales in rural areas. We investigate this in Table 6, where we present results with and without a full set of farmer controls from the surveys (Columns 2-3). 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. The regressions predict approximately 36 percentage point lower usage in the most remote villages, which translates to about 60-70%. The effect for seeds is smaller but still evident.

Similarly, in Panel B, we see strong evidence that sales are lower in remote areas. While the regression predicts that 44% of farmers will sell in the semi-urban areas, this declines to only 20% in the most remote areas. This is overwhelmingly 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. Rural farmers are more likely to buy 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). The proportion that report buying and selling is lower than in other recent work such as Bergquist, Burke and Miguel (2018).[15]

5 Model

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

5.1 Model Preliminaries

Production and Inputs

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

$$Y_i = \widetilde{\theta}_{i0} K_i^{\alpha_0} L_i^{1-\alpha_0} \tag{5}$$

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

$$\begin{aligned} \Pi_{i0} &= \alpha_0(1-\alpha_0)^{\frac{1-\alpha_0}{\alpha_0}} \widetilde{\theta}_{i0}^{\frac{1-\alpha_0}{\alpha_0}} p_i^{\frac{1}{\alpha_0}} w_i^{-\frac{1-\alpha_0}{\alpha_0}} K_i \\ &= \theta_{i0}\pi_{i0} \end{aligned} \tag{6}$$

Here, $\theta_{i0} = \alpha_0(1-\alpha_0)^{\frac{1-\alpha_0}{\alpha_0}} \widetilde{\theta}_{i0}^{\frac{1-\alpha_0}{\alpha_0}}$ and $\pi_{i0} = p_i^{\frac{1}{\alpha_0}} w_i^{-\frac{1-\alpha_0}{\alpha_0}} K_i$. The former term, 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, and

[15] See Web Appendix Figures A3 and A4 for a scatter plot of several of the key outcomes.

also provides a productivity shock, $\widetilde{\theta}_{ijv}$:

$$Y_i = \widetilde{\theta}_{ijv} K^{\alpha} L^{(1-\alpha)\beta} M^{(1-\alpha)(1-\beta)} \tag{7}$$

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

$$\Pi_i = \theta_{ijv} \pi_i r_{ijv}^{-\sigma} \tag{8}$$

where $\sigma \equiv \frac{1-\alpha}{\alpha}(1-\beta)$, $\pi_i = p_i^{\frac{1}{\alpha_0}} w_i^{-\beta\frac{1-\alpha_0}{\alpha_0}} K_i$, and $\theta_{ijv} = \kappa_2 \widetilde{\theta}_{ijv}^{\kappa_1}$.[16] Here, the profitability of fertilizer is a function of the productivity shock, θ_{ijv}, which we will assume is random and with a central moment that is dependent on the source of the fertilizer, the (delivered) price of fertilizer itself, r_{ijv}, and 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, where to purchase fertilizer, which itself is affected by prices for fertilizer at each agrovet location as well as the travel costs to get there and back. Suppose that the set of villages that contain an agrovet is defined as $\mathcal{V}$, where the price charged at location $v \in \mathcal{V}$ by agrovet j is r_{jv}. The per-unit cost to the farmer i, inclusive of transport costs, will be written as $r_{ijv} = r_{jv}\tau_{iv}$, where τ_{iv} is an iceberg trade cost for farmer i in traveling to v and back. Further, we assume that θ_{ijv}, is a random variable that could represent unobserved inputs purchased at agrovet j or location v, or perhaps other networking and information that is acquired at location v that may increase profitability. Whatever the interpretation, we assume that θ_{ijv} is distributed according to Fréchet distribution with location parameter T_{jv} and dispersion parameter ε. Precisely:

$$\Pr\left(\theta_{ijv} < \theta\right) = \exp\left(-T_{jv}\theta^{-\varepsilon}\right) \tag{9}$$

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

$$\Pr\left(\Pi_{ijv} < \pi\right) = \exp\left(-T_{jv}\pi_i^{\varepsilon} r_{ijv}^{-\varepsilon\sigma} \pi^{-\varepsilon}\right) \tag{10}$$

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

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

$$\Pr\left(\Pi_{i0} < \pi\right) = \exp\left(-T_{i0}\pi_{i0}^{\varepsilon}\pi^{-\varepsilon}\right) \tag{11}$$

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:

$$\lambda_{ijv} = \frac{T_{jv}\pi_i^{\varepsilon} r_{ijv}^{-\varepsilon\sigma}}{T_{i0}\pi_{i0}^{\varepsilon} + \sum_{v' \in \mathcal{V}} \sum_{l \in \mathcal{J}_{v'}} T_{lv'}\pi_i^{\varepsilon} r_{ilv'}^{-\varepsilon\sigma}} \tag{12}$$

Decomposing Choice Probabilities

Note that (12) can be broken up into the probability of adoption for i, and the probability i buys from j at location v conditional on buying:

$$\begin{aligned} \lambda_{ijv} &= \underbrace{\frac{\sum_{v' \in \mathcal{V}} \sum_{l \in \mathcal{J}_{v'}} T_{lv'} r_{ilv'}^{-\varepsilon\sigma}}{T_{i0}\left(\frac{\pi_{i0}}{\pi_i}\right)^{\varepsilon} + \sum_{v' \in \mathcal{V}} \sum_{l \in \mathcal{J}_{v'}} T_{lv'} r_{ilv'}^{-\varepsilon\sigma}}}_{\mu_i} \cdot \underbrace{\frac{T_{jv} r_{ijv}^{-\varepsilon\sigma}}{\sum_{v' \in \mathcal{V}} \sum_{l \in \mathcal{J}_{v'}} T_{lv'} r_{ilv'}^{-\varepsilon\sigma}}}_{\lambda_{ijv|adopt}} \qquad (13) \\ &= \mu_i \cdot \lambda_{ijv|adopt} \qquad (14) \end{aligned}$$

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

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

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

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

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

Precisely:

$$\lambda_{ijv|adopt} = \underbrace{\frac{\tau_{iv}^{-\varepsilon_a}\sum_{l\in\mathcal{J}_v} T_{lv}r_{lv}^{-\varepsilon_a}}{\sum_{v'\in\mathcal{V}}\tau_{iv'}^{-\varepsilon_a}\sum_{l\in\mathcal{J}_{v'}} T_{lv'}r_{lv'}^{-\varepsilon_a}}}_{\lambda_{iv|adopt}} \cdot \frac{T_{jv}r_{jv}^{-\varepsilon_a}}{\sum_{l\in\mathcal{J}_v} T_{lv}r_{lv}^{-\varepsilon_a}} \tag{17}$$

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

5.2 Calibrating the Farmer's Problem

Estimating Transport Costs through Location Choice

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

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

$$\phi_v = \sum_{l\in\mathcal{J}_v} T_{lv}r_{lv}^{-\varepsilon_a} \tag{18}$$

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

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

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

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

$$\lambda_{iv|adopt} = \frac{\exp\left(d_v - \varepsilon_a \log\left(\tau_{iv}\right)\right)}{\sum_{v'\in\mathcal{V}}\exp\left(d_{v'} - \varepsilon_a \log\left(\tau_{iv'}\right)\right)} \tag{20}$$

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, with dummy variables for whether location v is outside the village's urban or rural market-catchment areas:

$$-\varepsilon_a \log(\tau_{iv}) = \beta_{dist} dist_{iv}$$

Here, $dist_{iv}$ is the kilometer distance between farmer i and location v.

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

$$-\varepsilon_a \log(\tau_{iv}) = \sum_h \beta_h D^h_{iv} \tag{21}$$

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

$$\lambda_{iv|adopt} = \frac{\exp\left(d_v + \sum_h \beta_h D^h_{iv}\right)}{\sum_{v' \in \mathcal{V}} \exp\left(d_{v'} + \sum_h \beta_h D^h_{iv'}\right)} \tag{22}$$

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 (22) 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 $\varepsilon_a = 5.5$ (which will be supported by later results), the results suggest that iceberg transport costs for fertilizer are 2.6% ad-valorem per kilometer. As technologies may change discretely depending on the distance to each agrovet (walking short distances, taking transit for long distances), our preferred specification using distance bins is presented in column 2 of Table 7. 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 way 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 0.20.[17]

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

$$\frac{\lambda_{0-5km}}{\lambda_{0km}} = \frac{\exp(d_v - 1.61)}{\exp(d_v - 0)} = 0.20$$

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

$$\log(\tau_{iv}) = -\frac{1}{\varepsilon_a}\sum_h \beta_h D_{iv}^h \tag{23}$$

Dividing the coefficient estimates by ε_a give us the log change in trade costs. Given the iceberg assumption, this is also interpreted as the log change in the delivered price. Thus, at the central estimate of $\varepsilon_a = -5.5$, the comparison is equivalent to a 5.4% ad-valorem trade cost over 5km. Using the upper bound of each bin as a conservative measure to calculate ad-valorem trade costs, the ad-valorem equivalent rises modestly over the next few bins - for example, choosing an agrovet 15-20km away, the approximate ad-valorem equivalent trade cost evaluated at the upper cutoff is 6.2% per km.

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 sizeable non-pecuniary costs of traveling to acquire fertilizer.

Calibrating Non-price Attributes of Each Agrovet

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

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

$$\mathbb{E}[v_{jv}] = \sum_i L_i \mu_i \lambda_{ijv|adopt} \mathbb{E}[F_i|adopt\ at\ jv] \tag{24}$$

where $\mathbb{E}[F_i|adopt\ at\ jv]$ is expected fertilizer expenditures by i , conditional on adopting at j, v , and L_i is the village population to use as weights in the demand equation. Noting that $\mu_i = \frac{\mathbb{E}[F_i]}{\mathbb{E}[F_i|adopt]}$, which is the ratio of the unconditional expected expenditures on fertilizer to the expected expenditures on fertilizer, given adoption, we can write:

$$\mathbb{E}[v_{jv}] = \sum_i L_i \lambda_{ijv|adopt} \frac{\mathbb{E}[F_i|adopt\ at\ jv]}{\mathbb{E}[F_i|adopt]} \mathbb{E}[F_i] \tag{25}$$

Using the properties of the Fréchet distribution, it is straightforward to show that $\frac{\mathbb{E}[F_i|adopt\ at\ jv]}{\mathbb{E}[F_i|adopt]} = 1$; that is, the expected expenditures conditional on adoption anywhere is the same as the expected expenditures at some j, conditional on choosing j.[18] Thus, we have:

$$\mathbb{E}\left[v_{jv}\right] = \sum_i \lambda_{ijv|adopt} \mathbb{E}\left[F_i\right] \tag{26}$$

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

$$\mathbb{E}\left[v_{jv}\right] = \sum_i L_i \left(\frac{T_{jv} \tau_{iv}^{-\varepsilon_a} r_{jv}^{-\varepsilon_a}}{\sum_{v' \in \mathcal{V}} \sum_{l \in \mathcal{J}_{v'}} T_{lv'} \tau_{iv'}^{-\varepsilon_a} r_{lv'}^{-\varepsilon_a}} \right) \mathbb{E}\left[F_i\right] \tag{27}$$

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

$$\mathbb{E}\left[v_{jv}\right] = \sum_i L_i \left(\frac{\exp\left(-\sum_h \beta_h D_{iv}^h\right) \eta_{jv}}{\sum_{v' \in \mathcal{V}} \sum_{l \in \mathcal{J}_{v'}} \exp\left(-\sum_h \beta_h D_{iv'}^h\right) \eta_{lv'}} \right) \mathbb{E}\left[F_i\right] \tag{28}$$

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

However, we again run into an 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 16 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.[19]Then, for those villages that report zero adoption in the sample, we assign a small value of $\mathbb{E}\left[F_i\right]$ 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, $\left(\frac{F_i}{K_i}\right)_1$.[20]

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

[18]See technical appendix for a proof

[19]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.

[20]Precisely, imputed (small) values for expected fertilizer expenditures are calculated by: $\mathbb{E}\left[F_i\right] = 0.025 \left(\frac{F_i}{K_i}\right)_1 \cdot K_i$

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 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 (28), under the normalizing assumption that $\sum_v \sum_j \eta_{jv} = 1$.[21] Then, after recovering the η_{jv}'s, we calibrate the value of ε_a using the following regression (based on $\eta_{jv} = T_{jv} r_{jv}^{-\varepsilon_a}$):

$$\log(\eta_{jv}) = -\varepsilon_a \log(r_{jv}) + \underbrace{\beta exper_{jv} + district + u_{jv}}_{\log(T_{jv})} \tag{29}$$

where $\log(T_{jv})$ will be recovered as the components of the regression not equal to the elasticity-adjusted log agrovet price. To account for other local factors that may be correlated with prices and location quality, we include agrovet-level experience at that location, $exper_{iv}$, and district fixed effects. Running this regression, instrumenting for current agrovet prices with one-year lagged prices, yields an estimate of $\varepsilon_a = 5.46$.[22]

Calibrating Adoption

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

$$\mu_i = \frac{\sum_{v' \in \mathcal{V}} \sum_{l \in \mathcal{J}_{v'}} T_{lv'} \tau_{iv'}^{-\varepsilon_a} r_{lv'}^{-\varepsilon_a}}{\Phi_{i0} T_{i0} \left(\frac{\pi_{i0}}{\pi_i}\right)^{\varepsilon} + \sum_{v' \in \mathcal{V}} \sum_{l \in \mathcal{J}_{v'}} T_{lv'} \tau_{iv'}^{-\varepsilon_a} r_{lv'}^{-\varepsilon_a}} \tag{30}$$

where $\Phi_{i0} = T_{i0} \left(\frac{\pi_{i0}}{\pi_i}\right)^{\varepsilon} = T_{i0} p_i^{\varepsilon_p} w_i^{\varepsilon_w}$ $\varepsilon_p = \varepsilon \left(\frac{\alpha - \alpha_0}{\alpha \alpha_0}\right)$ and $\varepsilon_w = \varepsilon \left(\beta \frac{1-\alpha}{\alpha} - \frac{1-\alpha_0}{\alpha_0}\right)$. Using the imputed adoption shares for each village (as described in the preceeding section), we recover Φ_{i0}'s by inverting equation (30).

Beyond calibrating the term related to non-adoption, Φ_{i0}, the elasticities of adoption with respect to price and wages will depend on the values of the parameters of the Cobb-Douglas production functions with and without fertilizer. For assessing basic input counterfactuals and quantifying the impact of remoteness on adoption, these production function estimates are not required. However, they will be required for mark-up calculations to calibrate the agrovet pricing equilibrium; specif-

[21] This normalizing assumption is required since the probabilities within the sum in equation (28) are homogeneous degree zero in η's.

[22] The first stage F-statistic for weak instruments is 304.147, and the Wu-Hausman test statistic for endogeneity is 0.065 (p-value 0.799), suggesting this IV approach is sound.

ically, separating the substitution elasticity ε_a from the native distribution parameter ε. In the appendix, we detail a simple estimator for maize production with and without fertilizer, and using the Tanzanian LSMS, produce estimates for α (0.431) and α_0 (0.570). Also using the LSMS, we use reported wages and labor and fertilizer expenditures to calculate the share of labor in variable factors; β (0.75). Using these estimates, it is straightforward to calculate that ε=15.9. While this may seem high, this essentially means that there is little idiosyncratic variation in quality-adjusted prices at each agrovet, around the T_{jv}'s. Practically, farmers are choosing the lowest quality adjusted price for each agrovet, with minimal other variation that distracts from prices, quality, and transport costs.

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).

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:

$$r_{jv} = \frac{\varepsilon^v_{jv} - 1}{\varepsilon^v_{jv}} c_{jv} \tag{31}$$

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 on 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 0.190 (median = 0.204), which while lower 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 (31), 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 4. 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 29pp, or approximately twice baseline. This counterfactual alone accounts for 18% of the adoption-remoteness relationship at baseline. The results for expenditures are even more pronounced, where expenditures rise approximately 1.4 log points, though it is imporant 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 Moshi. 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 4. Here, adoption rises by about 2pp, or 6.4%, 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 absorbtion by mark-ups, is non-trivial.

Entry

An overarching question throughout the paper has been why agrovet access is worse in rural 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 5 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 higher when entering less remote villages, though adoption effects are modestly higher when entering the more remote villages. The former relationship is paritcularly strong, where a one standard deviation increase in remoteness reduces the profitability of hypothetical entry by 0.463 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 as 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 5. 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 66%. 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 chosing different option. Thus, the noise in the farmer-retailer transport shock as it related to distance will attenuate its effect on the gradient, though still provide a sizeable 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 collection still consist of only one region in one country. Do our results generalize? While we cannot provide a definitive answer, we can provide some suggestive evidence in this section.

7.1 Price Dispersion

To address this, we assembled five secondary datasets[23] across 1,512 locations[24] 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.[25] To quantify price dispersion, we first decompose variation in spatial prices by running the following regression:

$$\log(p_{mcjt}) = \gamma_c + \gamma_j + \gamma_t + \epsilon_{mjt} \tag{32}$$

where p_{mjt} (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 A4. 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[26], to run dyadic regressions to look at price gaps, as follows:

$$\log(\mid p_{mjt} - p_{m'jt} \mid) = \theta \log(c_{mm'}) + \gamma_m + \gamma_{m'} + \gamma_j + \epsilon_{mm'jt} \tag{33}$$

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.[27] Results are presented in Web Appendix Table A5. 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

[23]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.

[24]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.

[25]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.

[26]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).

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

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 A6, we assembled data from the World Bank LSMS-ISA household panel surveys for Ethiopia, Niger, Nigeria, Malawi, Tanzania, and Uganda, to study how remoteness affects fertilizer adoption. In the LSMS, measures of remoteness include distance to the main market, and distance to a population center. Using both measures of remoteness, we find a negative association between remoteness and technology adoption. However, since we cannot associate these adoption decisions with prices or precise measures of transport costs, we plan to continue our full field work in other regions of Tanzania. Further, we plan to do extensive work evaluating the sourcing decisions of output buying intermediaries, and how the presence of output buying intermediaries ultimately affects the decision by farmers to adopt fertilizer.

8 Conclusion

In this paper we collect extremely detailed and precise data on transportation costs, input and output prices, and the intensive and extensive margins of input purchases and output sales from market actors across the entirety of the supply chains for maize and fertilizer in all 1,183 villages in the Kilimanjaro region of Northern Tanzania. This data enables us to document large heterogeneity in market access, and study its implications for prices and for market participation. We find that there is large variation in prices, and the most remote villages are disadvantaged by as much as 50% of the mean relative to the least remote villages. The rates and magnitudes of fertilizer use and maize sales also display a large and significant distance gradient. Five different counterfactuals on the input and output sides suggest an important role for lowering access barriers on the input side for reducing this gradient, and a more qualified one for those on the output side.

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

References

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

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

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

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

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

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

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

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

[9] 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.

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

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

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

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

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

[15] 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.

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

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

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

[19] 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.

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

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

[22] 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.

[23] 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.

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

[25] 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

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

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

[28] 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.

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

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

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

[32] 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

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

A Deriving farmer profits, revenues, and input expenditures

The production function under basic technology is:

$$Y_i = \widetilde{\theta}_{i0} K_i^{\alpha} L_i^{1-\alpha} \tag{34}$$

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

$$\Pi_{i0} = \max_{L_i} \left\{ p_i \widetilde{\theta}_{i0} K_i^{\alpha} L_i^{1-\alpha} - w_i L_i \right\} \tag{35}$$

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:

$$(1-\alpha) p_i \widetilde{\theta}_{i0} K_i^{\alpha} L_i^{-\alpha} = w_i \tag{36}$$

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

$$w_i L_i = (1-\alpha) p_i \widetilde{\theta}_{i0} K_i^{\alpha} L_i^{1-\alpha} = (1-\alpha) R_{i0} \tag{37}$$

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:

$$\begin{aligned} \Pi_{i0} &= \alpha_0 (1-\alpha_0)^{\frac{1-\alpha_0}{\alpha_0}} \widetilde{\theta}_{i0}^{\frac{1}{\alpha_0}} p_i^{\frac{1}{\alpha_0}} w_i^{-\frac{1-\alpha_0}{\alpha_0}} K_i \\ &= \theta_{i0} \pi_{i0} \end{aligned} \tag{38}$$

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:

$$Y_i = \widetilde{\theta}_{ijv} K_i^{\alpha} L_{ijv}^{(1-\alpha)\beta} X_{ijv}^{(1-\alpha)(1-\beta)} \tag{39}$$

The profit maximization problem when using fertilizer is written as:

$$\Pi_{i0} = \max_{L_i, X_{ijv}} p_i \widetilde{\theta}_{ijv} K_i^{\alpha} L_{ijv}^{(1-\alpha)\beta} F_{ijv}^{(1-\alpha)(1-\beta)} - w_i L_{ijv} - r_{ijv} F_{ijv} \tag{40}$$

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

$$c_{ijv} M_{ijv} = (1-\alpha) p_i \widetilde{\theta}_{ijv} K_i^{\alpha} L_{ijv}^{(1-\alpha)\beta} F_{ijv}^{(1-\alpha)(1-\beta)} = (1-\alpha) R_{ijv} \tag{41}$$

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:

$$\Pi_i = \theta_{ijv} \pi_i r_{ijv}^{-\sigma} \tag{42}$$

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 $\theta_{ibjv} = \kappa_2 \widetilde{\theta}_{ibjv}^{\kappa_1}$.[28] 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.

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

B Distributions of Fertilizer Expenditures

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

$$\mathbb{E}\left[rF_i \mid adopt\ at\ j\ in\ v\right] = \mathbb{E}\left[rF_i | adopt\right] \tag{43}$$

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.

$$\begin{aligned}
\Pr\left(\Pi_{ijv} > \pi\right) &= \Pr\left(\theta_{ijv}\pi_i r_{ijv}^{-\sigma} > \pi\right) && (44)\\
&= \Pr\left(\theta_{ijv} > \frac{\pi}{\pi_i} r_{ijv}^{\sigma}\right) && (45)\\
&= 1 - \exp\left(-T_{jv}\pi_i^{\varepsilon} r_{ijv}^{\varepsilon\sigma}\pi^{-\varepsilon}\right) && (46)
\end{aligned}$$

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

$$\Pr\left(\Pi_{ijv} > \pi\right) = 1 - \exp\left(-T_{jv}\gamma_{ijv}\pi^{-\varepsilon}\right) \tag{47}$$

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

$$\Pr\left(\Pi_{i0} > \pi\right) = 1 - \exp\left(-\Phi_{i0}\pi^{-\varepsilon}\right) \tag{48}$$

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

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:

$$\begin{aligned}
\Pr\left(\Pi_i^{max} > \pi\right) &= \Pr\left(\Pi_{ijv} > \pi \text{ for any } jv\right) && (49)\\
&= 1 - \Pr\left(\Pi_{ijv} < \pi\ \forall\ jv\right) && (50)
\end{aligned}$$

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

as:

$$\begin{aligned}
\Pr\left(\Pi_i^{max} > \pi\right) &= 1 - \Pr\left(\Pi_{ijv} < \pi \ \forall \ jv\right) && (51)\\
&= 1 - \prod_{v' \in \mathcal{V}} \prod_{j \in \mathcal{J}_v} \Pr\left(\Pi_{ijv} < \pi\right) && (52)\\
&= 1 - \prod_{v' \in \mathcal{V}} \prod_{j \in \mathcal{J}_v} \exp\left(-\pi^{-\varepsilon}\right) && (53)
\end{aligned}$$

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

$$\Pr\left(\Pi_i^{max} > \pi\right) = 1 - \exp\left(-\Phi_i \pi^{-\varepsilon}\right) \quad (54)$$

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

$$G_i^{max}\left(\pi\right) = \Pr\left(\Pi_i^{max} < \pi\right) = \exp\left(-\Phi_i \pi^{-\varepsilon}\right) \quad (55)$$

with pdf:

$$g_i^{max}\left(\pi\right) = \varepsilon \Phi_i \pi^{-\varepsilon - 1} \exp\left(-\Phi_i \pi^{-\varepsilon}\right) \quad (56)$$

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

$$\begin{aligned}
\Pr\left(\Pi_i > \pi\right) &= \Pr\left(\Pi_{ijv} > \pi \text{ for any } jv \cup \Pi_{i0} > \pi\right) && (57)\\
&= 1 - \Pr\left(\Pi_{ijv} < \pi \ \forall \ jv \ \cap \ \Pi_{i0} < \pi\right) && (58)
\end{aligned}$$

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

$$\begin{aligned}
\Pr\left(\Pi_i > \pi\right) &= 1 - \Pr\left(\Pi_{ijv} < \pi \ \forall \ jv \ \cap \ \Pi_{i0} < \pi\right) && (59)\\
&= 1 - \Pr\left(\Pi_{i0} < \pi\right) \prod_{v' \in \mathcal{V}} \prod_{j \in \mathcal{J}_v} \Pr\left(\Pi_{ijv} < \pi\right) && (60)\\
&= 1 - \exp\left(-T_{i0} \gamma_{i0} \pi^{-\varepsilon}\right) \prod_{v' \in \mathcal{V}} \prod_{j \in \mathcal{J}_v} \exp\left(-T_{jv} \gamma_{ijv} \pi^{-\varepsilon}\right) && (61)
\end{aligned}$$

Using the definitions for Φ_{i0} and Φ_i, this is simplified as:

$$\Pr\left(\Pi_i > \pi\right) = 1 - \exp\left(-\left(\Phi_{i0} + \Phi_i\right) \pi^{-\varepsilon}\right) \quad (62)$$

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

$$G_i\left(\pi\right) = \exp\left(-\left(\Phi_{i0} + \Phi_i\right) \pi^{-\varepsilon}\right) \quad (63)$$

with pdf:

$$g_i(\pi) = \varepsilon(\Phi_{i0}+\Phi_i)\pi^{-\varepsilon-1}\exp\left(-(\Phi_{i0}+\Phi_i)\pi^{-\varepsilon}\right) \quad (64)$$

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:

$$\Pr\left(\Pi_i^{max}<\pi\middle|adopt\right) = \frac{\Pr\left(\Pi_i^{max}<\pi\cap\Pi_i^{max}>\Pi_{i0}\right)}{\Pr\left(\Pi_i^{max}>\Pi_{i0}\right)} \quad (65)$$

This can be re-written as:

$$\begin{aligned}\Pr\left(\Pi_i^{max}<\pi\middle|adopt\right) &= \frac{1}{\Pr\left(\Pi_i^{max}>\Pi_{i0}\right)}\int_0^{\pi}\Pr\left(s>\Pi_{i0}\right)g_i^{max}(s)ds \\ &= \frac{1}{\Pr\left(\Pi_i^{max}>\Pi_{i0}\right)}\int_0^{\pi}\exp\left(-\Phi_{i0}s^{-\varepsilon}\right)\varepsilon\Phi_i s^{-\varepsilon-1}\exp\left(-\Phi_i s^{-\varepsilon}\right)ds \\ &= \frac{1}{\Pr\left(\Pi_i^{max}>\Pi_{i0}\right)}\int_0^{\pi}\varepsilon\Phi_i s^{-\varepsilon-1}\exp\left(-\left(\Phi_{i0}+\Phi_i\right)s^{-\varepsilon}\right)ds \quad (66)\end{aligned}$$

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

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

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

$$\begin{aligned}\Pr\left(\Pi_i^{max}<\pi\middle|adopt\right) &= \int_0^{\pi}\varepsilon\left(\Phi_{i0}+\Phi_i\right)s^{-\varepsilon-1}\exp\left(-\left(\Phi_{i0}+\Phi_i\right)s^{-\varepsilon}\right)ds \quad (67)\\ &= \Pr\left(\Pi_i<\pi\right) \quad (68)\end{aligned}$$

Profits conditional on adoption from j

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

$$\Pr\left(\Pi_{ijv}<\pi\middle|adopt\ from\ j\ in\ v\right) = \frac{\Pr\left(\Pi_{ijl}<\pi\cap\Pi_{ijv}>\Pi_{ij'l}\forall(j',l)\cap\Pi_{ijv}>\Pi_{i0}\right)}{\Pr\left(\Pi_{ijv}>\Pi_{ij'l}\forall(j',l)\cap\Pi_{ijv}>\Pi_{i0}\right)} \quad (69)$$

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

$$\Pr\left(\Pi_{ijv}<\pi\middle|adopt\ from\ j\ in\ v\right) = \frac{1}{\lambda_{ijv}}\int_0^{\pi}Pr\left(s>\Pi_{ij'l}\forall(j',l)\cap s>\Pi_{i0}\right)g_{ijv}(s)ds \quad (70)$$

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

$$\begin{aligned} \Pr\left(s > \Pi_{ij'l} \forall (j', l) \cap s > \Pi_{i0}\right) &= \exp\left(-\Phi_{i0} s^{-\varepsilon}\right) \exp\left(-\widetilde{\Phi}_{ijv} s^{-\varepsilon}\right) \quad (71) \\ &= \exp\left(-\left(\Phi_{i0} + \widetilde{\Phi}_{ijv}\right) s^{-\varepsilon}\right) \quad (72) \end{aligned}$$

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

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

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

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

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

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

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

C 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.

D 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:

$$\mathbb{E}\left[v_{jv}\right] = (1-\beta)\,\frac{1-\alpha}{\alpha}\sum_{i}\lambda_{ijv}\mathbb{E}\left[\Pi_i\right] \tag{73}$$

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

$$\frac{d\mathbb{E}\left[v_{jv}\right]}{dr_{jv}}\frac{r_{jv}}{\mathbb{E}\left[v_{jv}\right]} = \sum s_{ijv}\left(\frac{d\lambda_{ijv}}{dr_{jv}}\frac{r_{jv}}{\lambda_{ijv}} + \frac{d\mathbb{E}\left[\Pi_i\right]}{dr_{jv}}\frac{r_{jv}}{\mathbb{E}\left[\Pi_i\right]}\right) \tag{74}$$

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:

$$\frac{d\mathbb{E}\left[v_{jv}\right]}{dr_{jv}}\frac{r_{jv}}{\mathbb{E}\left[v_{jv}\right]} = -\varepsilon_a\sum_{i}s_{ijv}\left((1-\lambda_{ijv}) + \frac{1}{\varepsilon}\lambda_{ijv}\right) \tag{75}$$

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

$$\varepsilon_v \equiv \frac{d\mathbb{E}\left[v_{jv}\right]}{dr_{jv}}\frac{r_{jv}}{\mathbb{E}\left[v_{jv}\right]} = -\varepsilon_a + \frac{\varepsilon-1}{\varepsilon}\varepsilon_a\sum_{i}s_{ijv}\lambda_{ijv} \tag{76}$$

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	11.3
	(29.80)
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)	1.24
	(1.55)
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]	19.1
	(12.60)
Google estimates	
Travel speed on feeder roads and rural roads - km/hr (Google)	37.0
	(15.5)
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. The number of markets included is 202.

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				0.032***			0.022***		
				(0.007)			(0.005)		
Google maps: hours to destination					3.693***	0.755***		2.715***	0.874***
					(0.263)	(0.070)		(0.256)	(0.076)
Number of villages				1145	1036	1036	1149	1149	1027
Number of observations				1145	1036	1036	1149	1149	1027

Notes: Data is constructed from interviews with transportation operators in market centers. 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. Standard errors clustered by markets for Panel A and reported in parentheses.

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

[1]Cost is for one-way trip for a given route.

Table 3. Remoteness and farmer characteristics

	(1) Mean	(2) Coefficient from regression on standardized distance from regional hubs
Panel A. Demographic and background characteristics		
Age	49.76	-0.98*
	(15.23)	(0.52)
Female	0.45	-0.02
		(0.02)
Married	0.76	0.00
		(0.01)
Household size	4.95	0.26**
	(2.78)	(0.11)
Years of education	6.58	-0.31***
	(3.56)	(0.11)
Home has thatch roof	0.17	0.03
		(0.02)
Has cell phone	0.89	-0.03***
		(0.01)
Has bank account	0.15	-0.05***
		(0.01)
Has mobile money account	0.77	-0.08***
		(0.02)
Acres of land	5.46	1.37**
	(13.89)	(0.57)
Has market business	0.28	-0.05***
		(0.01)
Annual total income from non-farming (USD)	408.9	-74.72**
	(772.60)	(30.25)
Panel B. Production Capacity (in kg/acre)[1]		
FAO-GAEZ production capacity for low input level	779.4	82.81***
	(275.90)	(20.22)
FAO-GAEZ production capacity for high input level	3319	-269.79***
	(835.60)	(55.87)
FAO-GAEZ production difference between high and low	2540	-352.60***
	(725.60)	(46.79)
Panel C. Harvest Output		
Total harvest output in 2016 long rains (kg)	928.7	-16.62
	(1360.00)	(51.18)
Harvest output per acre	455.3	-85.11***
	(384.30)	(17.46)
Value of harvest output at average regional post-harvest price	198	-19.10**
	(237.90)	(9.26)

Notes: N = 2,845 farmers in 246 villages. In Column 1, standard deviations are in parentheses. In Column 2, 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) Mean	(2) Coefficient from regression on standardized distance from regional hubs
Panel A. Summary measures of access to input retailers		
Has at least 1 agrovet within 10 km of village	0.75	-0.14***
which sells fertilizer or seeds		(0.01)
Number of agrovets within 10 km of village	7.79	-2.93***
which sells fertilizer or seeds	(8.96)	(0.25)
Distance to nearest agrovet	6.79	3.17***
which sells fertilizer or seeds	(15.15)	(0.47)
Panel B. Travel-cost adjusted prices faced by farmers		
Minimum travel-cost adjusted price	24.53	2.42***
for 50 kg of Urea[1]	(5.61)	(0.16)
Decomposition of price between retail price and cost of transportation		
Retail price at the location with the lowest	19.61	1.10***
travel-cost adjusted price (USD)	(2.49)	(0.07)
Cost of travel to obtain minimum travel-cost	4.914	1.33***
adjusted price (USD)	(5.35)	(0.16)

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. In Column 2, standard errors in parentheses.

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

[1]We assume farmers buy a 50 kg bag in one trip (enough for 1 acre), 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)
	Mean	Coefficient from regression on standardized distance from regional 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***
		(0.01)
Number of maize sellers within 10 km of village	1.89	-1.06***
	(2.48)	(0.07)
Distance to the nearest output market with maize sellers (km)	8.67	5.65***
	(14.17)	(0.42)
Panel B. Maximum imputed travel-cost adjusted price if farmers were to sell in a local market		
Market survey: maximum travel-cost adjusted price	29.48	-3.62***
immediately after 2017 harvest (USD)	(9.12)	(0.26)
Decomposition of price between retail price and cost of transportation		
Retail price at the location with the highest	39.34	0.80***
travel-cost adjusted price (USD)	(3.17)	(0.09)
Cost of travel to obtain the highest travel-cost	9.87	4.42***
adjusted price (USD)	(8.99)	(0.24)
Panel B2. Price available within village by maize-buying intermediaries immediately after last season's harv		
Farmer surveys: average "going price" in local	30.38	0.16
village immediately after previous harvest[2]	(8.60)	(0.62)
Farmer surveys: average village sales price	25.86	0.38
after previous harvest[2]	(6.24)	(0.29)

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. In Column 2, standard errors in parentheses.

*, **, 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 only available from the farmer surveys (2,171 farmers in 137 villages).

Table 6. Remoteness and input market access and adoption

	(1)	(2)	(3)
	Mean	Coefficient from regression on standardized distance from regional hubs	
		No controls	Controls for soil and farmer characteristics
Panel A: Input usage			
Used chemical fertilizer in 2017 long rains	0.39	-0.17***	-0.09***
		(0.03)	(0.03)
Quantity of chemical fertilizer used (kg)	19.82	-12.91***	-6.43***
	(31.60)	(2.12)	(1.70)
Used improved seeds in 2017 long rains	0.64	-0.08***	-0.05**
		(0.02)	(0.02)
Quantity of improved seeds used (kg)	5.82	-1.48***	-1.25***
	(7.23)	(0.34)	(0.40)
Panel B. Maize sales			
Sold maize after previous long rains	0.323	-0.09***	-0.06**
		(0.02)	(0.03)
Total quantity sold (kg)	388.1	-97.86***	-112.16**
	(1142.00)	(35.11)	(48.86)
Sales to agents at home			
Agent visited homestead	0.31	-0.14***	-0.09***
		(0.03)	(0.03)
Sold maize to an agent after previous long rains	0.17	-0.07***	-0.04**
		(0.02)	(0.02)
Quantity sold to agents (kg)	142	-46.39***	-39.11**
	(433.70)	(13.80)	(18.61)
Sales at market			
Sold maize at a market after previous long rains	0.06	-0.03***	-0.03***
		(0.01)	(0.01)
Quantity sold at market (kg)	34.42	-14.61***	-15.26**
Panel C. Maize purchases	(197.10)	(5.32)	(7.72)
Farmer ever buys maize	0.48	0.11***	0.08***
		(0.02)	(0.02)
Quantity purchased in typical year (kg)	152.3	75.67***	65.05***
	(315.50)	(15.79)	(17.92)
Net buying			
Farmer buys maize but sells none	0.37	0.11***	0.08***
		(0.02)	(0.03)
Farmer sells maize and buys none	0.24	-0.09***	-0.06***
		(0.02)	(0.02)
Farmer buys and sells maize	0.08	-0.00	0.00
		(0.01)	(0.01)
Net buyer (quantity bought > quantity sold)	0.44	0.10***	0.07**
		(0.02)	(0.03)
Net seller (quantity bought < quantity sold)	0.26	-0.08***	-0.06**
		(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. Standard errors, clustered at village level, in parentheses in Columns 2-3. *, **, and *** indicate significance at 10%, 5%, and 1%.

Table 7. Multinomial logit of agrovet choice

	(1)	(2)
	Agrovet Chosen	
Kilometers to agrovet	-0.139	
	(0.007)	
Dummies for agrovet distance bin:		
between (0,5] km		-1.610***
		(0.374)
between (5,10] km		-2.774***
		(0.383)
between (10,15] km		-4.332***
		(0.396)
between (15,20] km		-4.904***
		(0.388)
between (20,30] km		-5.917***
		(0.383)
between (30,40] km		-6.626***
		(0.398)
between (40,50] km		-8.606***
		(0.494)
between (50,100] km		-10.017***
		(0.479)
over 100 km		-12.965***
		(0.779)

Notes: 448 farmers, 112 observed locations. Omitted group is agrovet locatred in respondent's village. Standard errors in parentheses. *, **, and *** indicate significance at 10%, 5%, and 1%.

Figure 1. Map of Survey Region and Villages

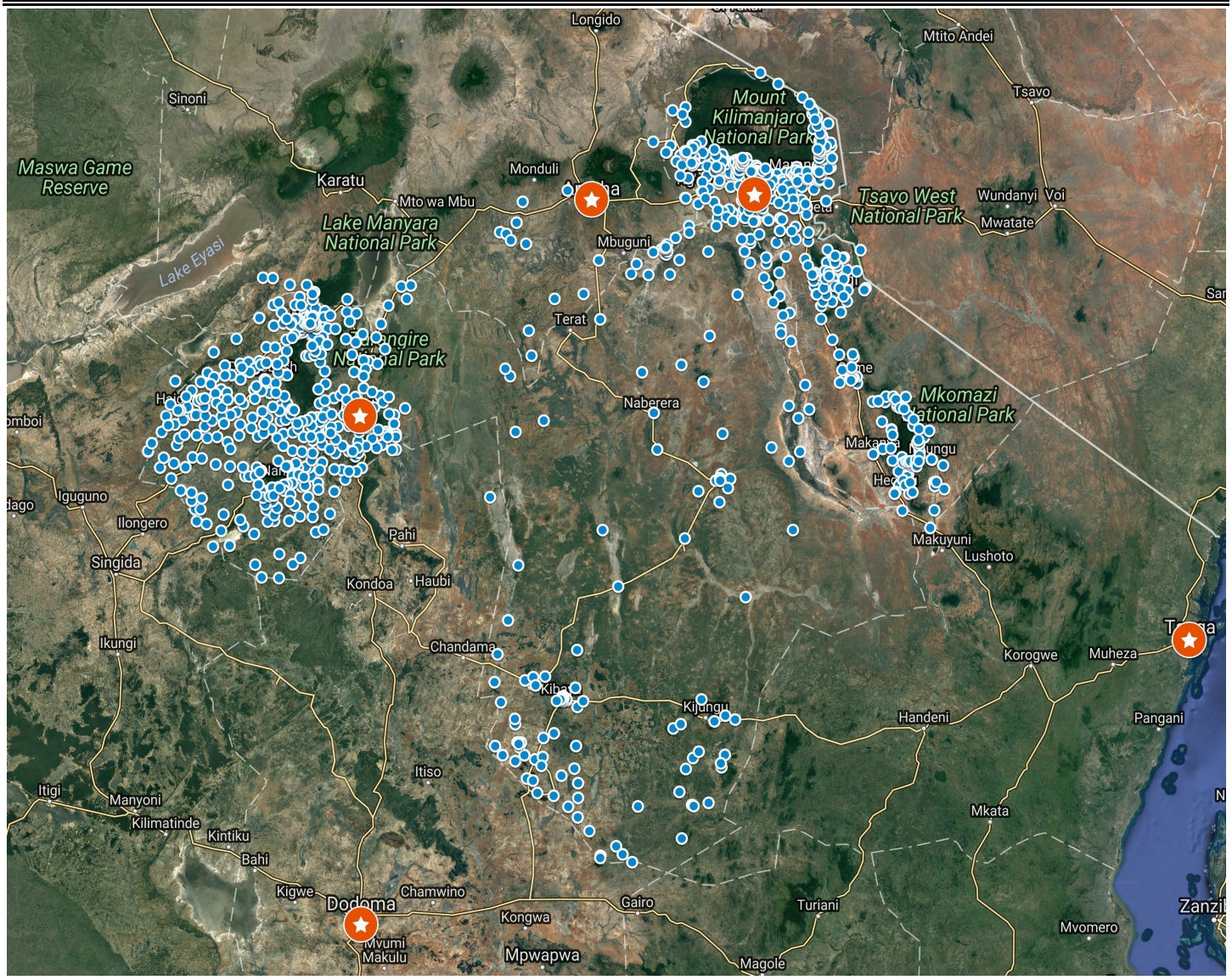
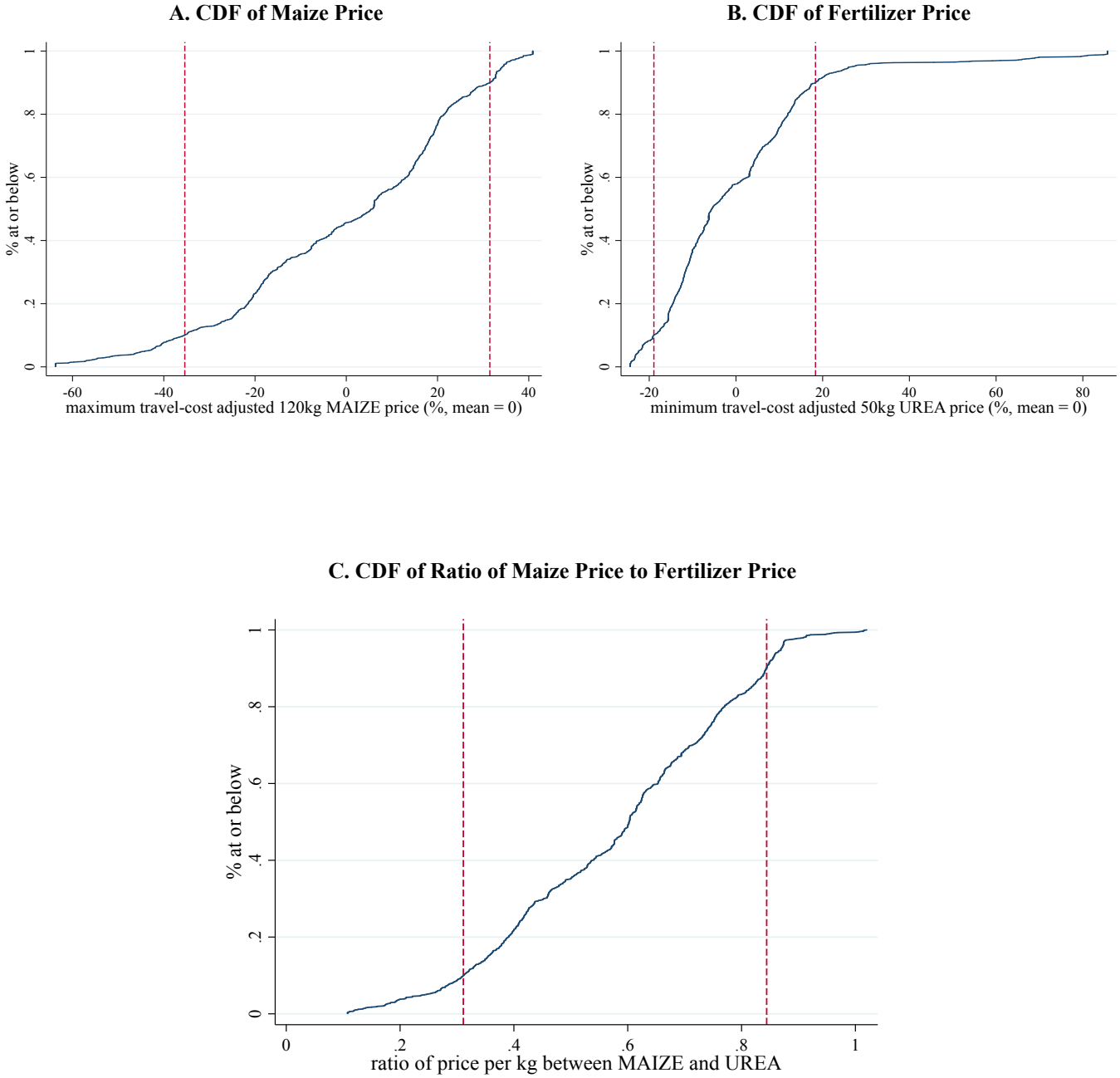
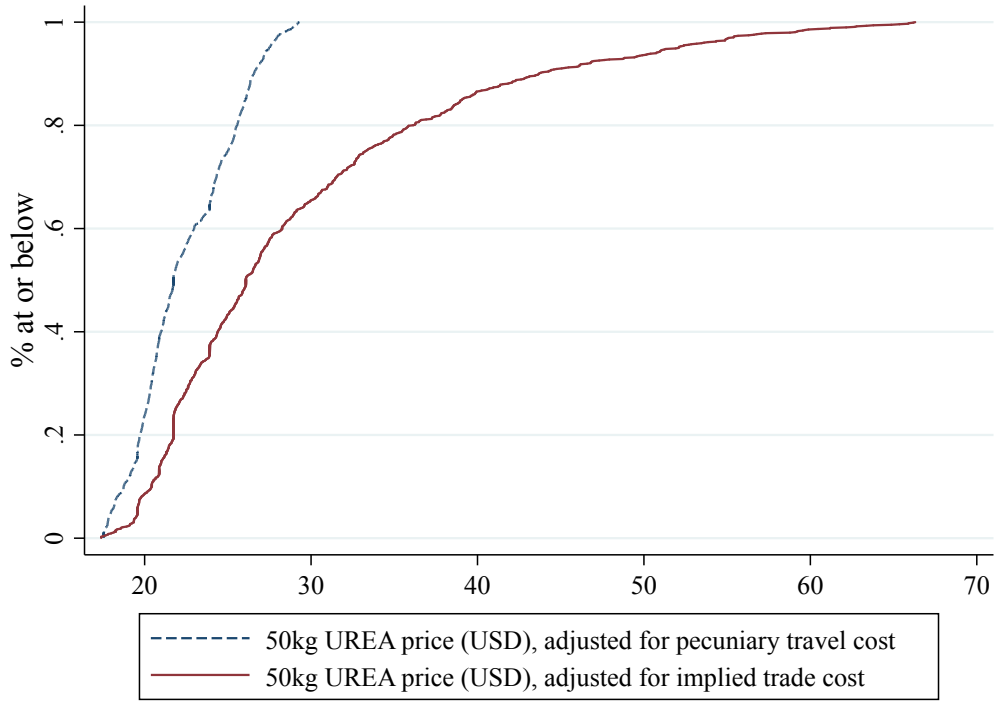


Figure 2. CDF of travel-cost adjusted prices across villages



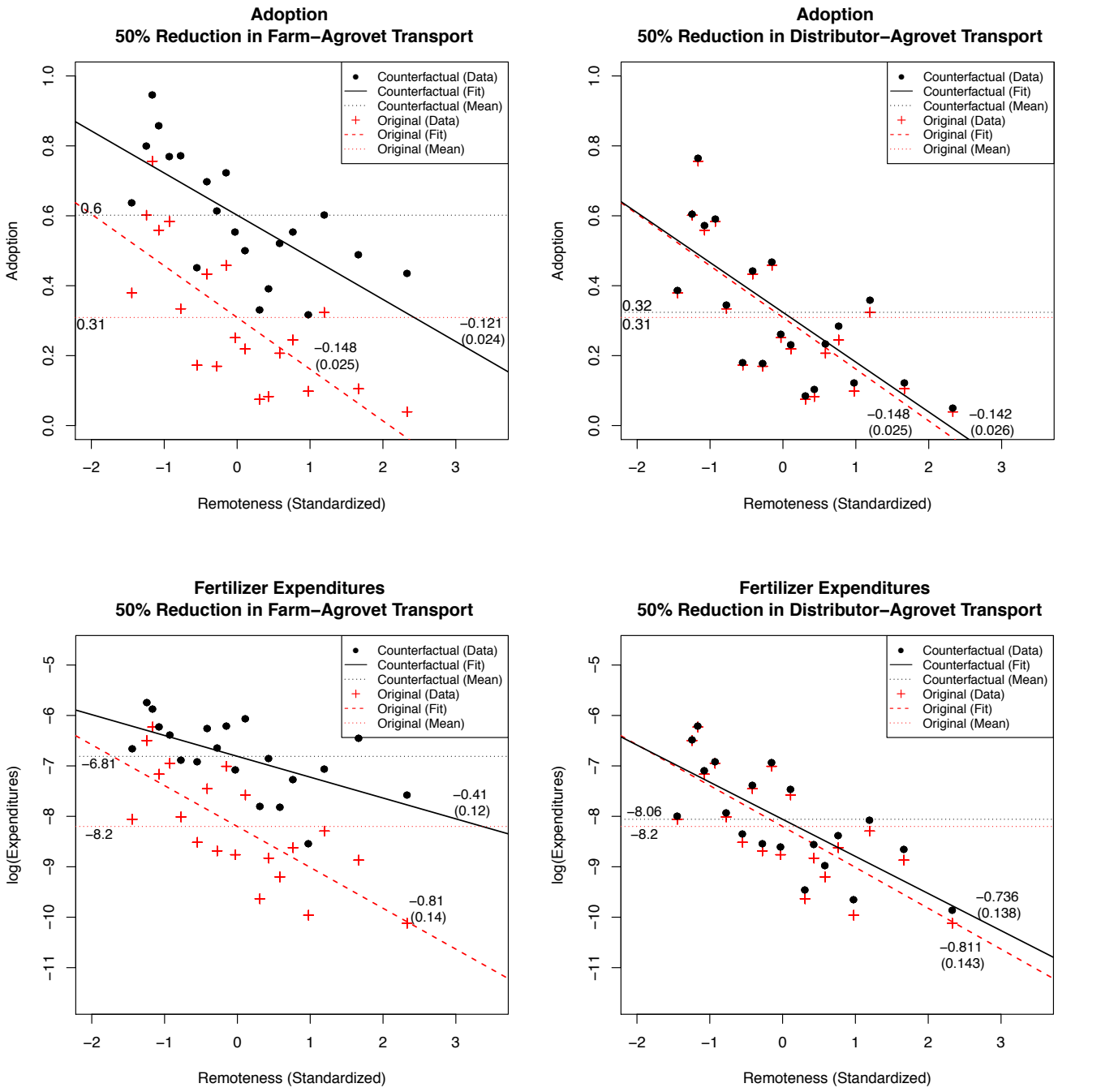
Notes: Each observation represents a village. Travel-cost adjusted prices are calculated through observed prices from an agrovet survey, a maize price survey at markets and transport cost information collected from interviews with transport operators.

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



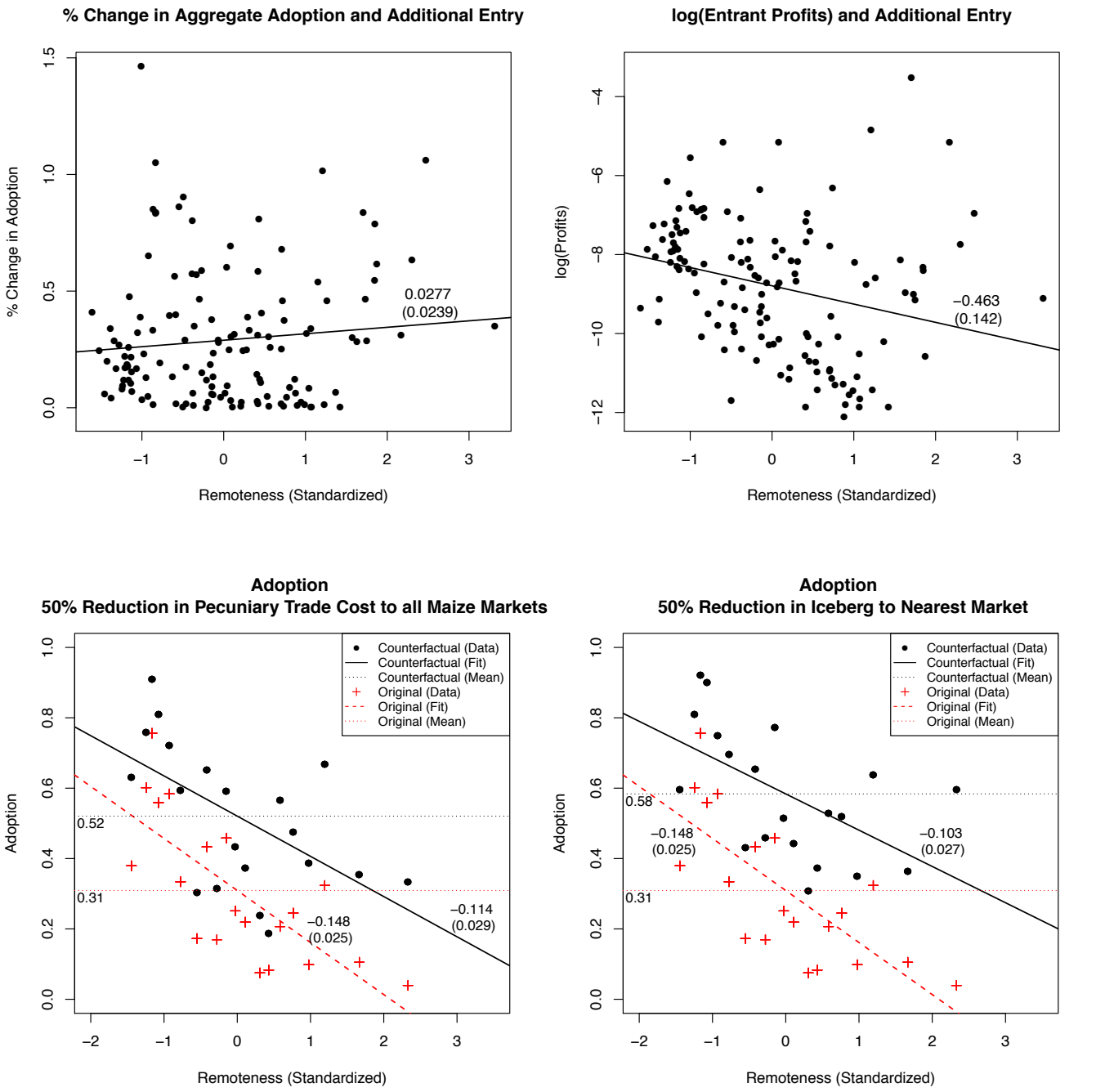
Notes: Each observation represents a village. Travel-cost adjusted prices are calculated through observed prices from an agrovet survey, a maize price survey at markets and transport cost information collected from interviews with transport operators.

Figure 4. Input Access Counterfactuals



Notes: See text for discussion of counterfactuals.

Figure 5. 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 actual input purchase or/and output selling. Standard errors clustered by village and reported in parentheses.
*, **, and *** indicate significance at 10%, 5%, and 1% respectively.

Web Appendix Table A3. Remoteness and fertilizer retailer sales, prices, and markups

	(1)	(2)
	Mean	Coefficient from regression on standardized distance from regional hubs
Panel A. Agrovet shop-level (N=507)		
Sells fertilizer	0.87	-0.03
	(0.34)	(0.02)
Number of varieties of fertilizer	1.76	-0.15*
	(1.58)	(0.09)
Quantity of fertilizer sold last year (kg)	5995	-473.41
	(12721)	(757.30)
Sells seeds	0.72	0.04
	(0.45)	(0.02)
Number of varieties of seeds	1.2	0.10
	(1.26)	(0.07)
Quantity of seeds sold last year (kg)	2194	903.59
	(8008)	(557.99)
Cost of transport from wholesaler (per 50 kg)	0.64	0.30***
	(0.69)	(0.04)
Panel B. Prices and markups (Agrovet shop-variety level, N=938)		
Retail price for 50 kilograms	25.21	0.65***
	(5.21)	(0.22)
Wholesale price for 50 kilograms	21.43	0.16*
	(4.14)	(0.09)
Markup (percentage points)[1]	13.42	0.86
	(10.25)	(0.62)

Notes: In Column 1, standard deviations are in parentheses. In Column 2, standard errors in parentheses. 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 A4. Robustness of Travel-cost Adjusted Prices

	(1) Mean	(2) Coefficient from regression on standardized distance from regional hubs	(2) Coefficient from regression on standardized distance from regional 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	24.41	22.59***	2.85***
for 50 kg of Urea[1]	(5.83)	(0.21)	(0.19)
Decomposition of price between retail price and cost of transportation			
Retail price at the location with the lowest	19.72	22.59***	1.32***
travel-cost adjusted price (USD)	(2.53)	(0.21)	(0.08)
Cost of travel to obtain minimum travel-cost	4.683	22.59***	1.53***
adjusted price (USD)	(5.47)	(0.21)	(0.20)
A2. Output Side: Travel-cost adjusted maize prices if farmers were to sell in a local market			
Market survey: maximum travel-cost adjusted price	28.88	28.71***	-4.58***
immediately after 2017 harvest (USD)	(9.73)	(0.32)	(0.33)
Decomposition of price between retail price and cost of transportation			
Retail price at the location with the highest	39.10	39.13***	0.89***
travel-cost adjusted price (USD)	(3.20)	(0.11)	(0.12)
Cost of travel to obtain the highest travel-cost	10.22	10.42***	5.47***
adjusted price (USD)	(9.78)	(0.30)	(0.31)
Panel B. Robustness to Assigning Best Prices to Missing Agrovets[1]			
Input Side: Travel-cost adjusted fertilizer prices faced by farmers			
Minimum travel-cost adjusted price	24.45	22.59***	2.36***
for 50 kg of Urea[1]	(5.64)	(0.21)	(0.16)
Decomposition of price between retail price and cost of transportation			
Retail price at the location with the lowest	19.65	22.59***	1.09***
travel-cost adjusted price (USD)	(2.46)	(0.21)	(0.07)
Cost of travel to obtain minimum travel-cost	4.801	22.59***	1.27***
adjusted price (USD)	(5.33)	(0.21)	(0.16)

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. In Column 2, standard errors in parentheses.
*, **, and *** indicate significance at 10%, 5%, and 1%.

[1]We assignd a 10 percentile and a 90 percentile of the fertilizer price distribution to each of the missing agrovets and chose the price that made the relationship weaker.

Web Appendix Table A5. Input and output market price dispersion across countries

	(1) Secondary Datasets	(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: 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.

[1]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.

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

Web Appendix Table A6. 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 A7. Adoption in LSMS-ISA surveys

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

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

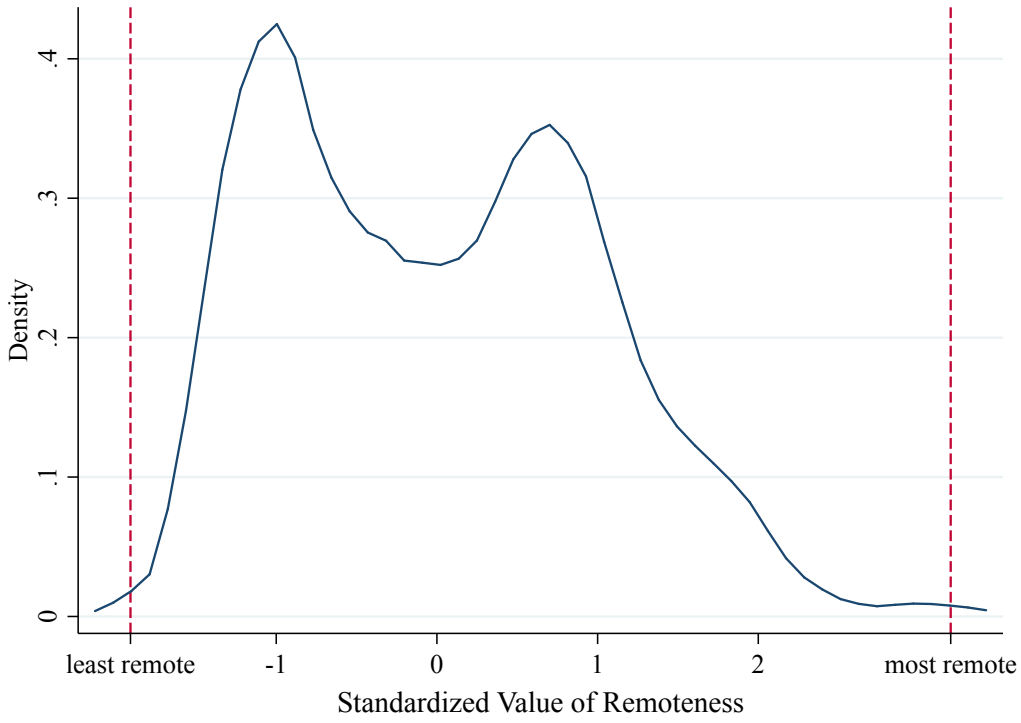
***, **, and * indicate significance at 1%, 5%, and 10%.

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

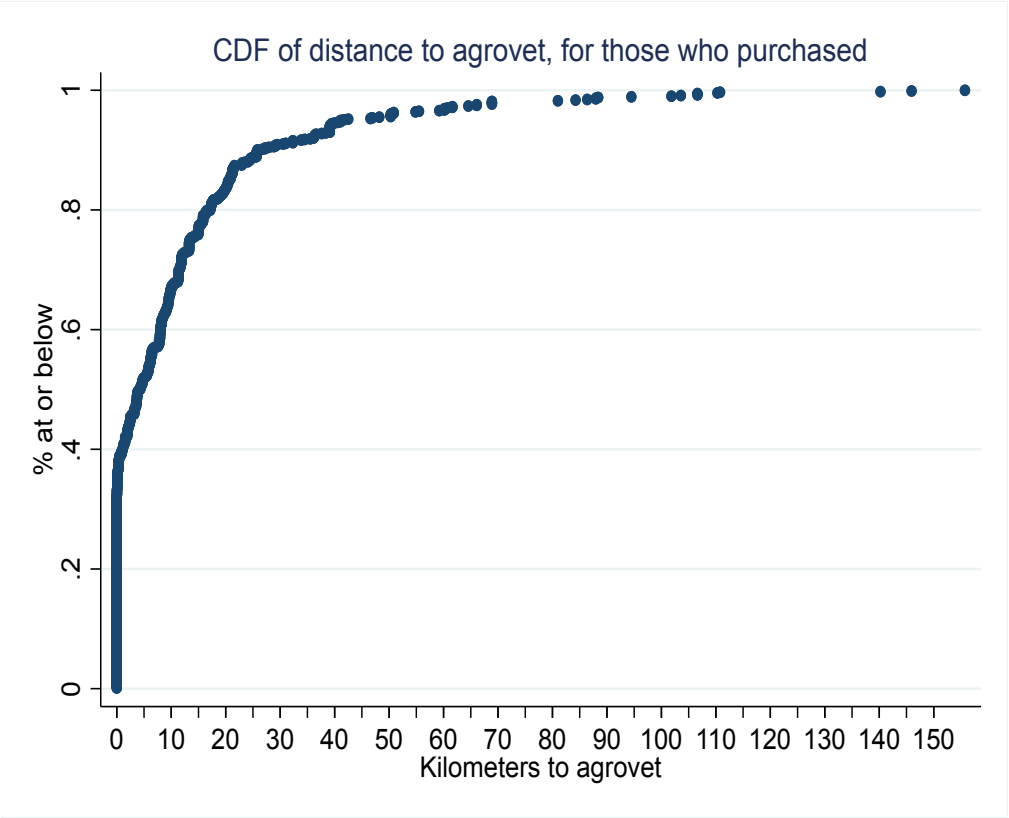
	(1)		(2)
	Dependent variable: log(Harvest/Acres)		
log(Labor/Acres)	0.419***	0.433***	0.430***
	(0.042)	(0.042)	(0.042)
log(Labor/Acres) x Used Fertilizer?	0.124*	0.124*	0.149*
	(0.075)	(0.074)	(0.077)
Used Fertilizer?	(0.33)	(0.33)	
	(0.300)	(0.298)	
District-Year fixed effects		X	
District-Year-Fertilizer 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

Web Appendix Figure A1. Distribution of Remoteness



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



Notes: Each point represents a farmer.