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

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

February 26, 2020

Abstract

We collect data on prices, travel costs and farmer decisions to quantify market access and its impact on agricultural productivity in 1,183 villages in two regions of Tanzania. Villages at the bottom of the travel cost-adjusted price distribution face 40-55% less favorable prices than those at the top. A one standard deviation increase of village-level remoteness is associated with 20-25% lower input adoption and output sales. A spatial model of input adoption conservatively estimates that farmers behave as if travel costs are 4% ad-valorem per kilometer. Counterfactuals suggest that halving travel costs would double adoption and reduce the adoption-remoteness gradient by 39%.

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

Keywords: market access, inputs, technology adoption, transport costs, roads, output sales

*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 Chatterjee, Denise Horner, Alain de Janvry, Betty Sadoulet and seminar participants at Ashoka University, Cal Poly San Luis Obispo, UC Berkeley, the University of Colorado, Georgetown SFS Qatar, ISB, ISI Delhi ACEGD 2017, Johns Hopkins SAIS, LSE, the University of Maryland, the NBER Trade and Agriculture Conference, the NBER Transforming Rural Africa Conference, the NEUDC, the University of Pittsburgh, Stanford,the University of Virginia, UCSC, the University of West Virginia and the WEAI for helpful comments. We acknowledge funding from the IGC (grant 1-VRC-VTZA-VXXXX-89307), the National Science Foundation (SES 1658942), Innovations for Poverty Action, ISB and UCSC. The data collection protocol 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 and, by extension, input adoption and agricultural productivity, there is little research to quantify its effect.

In this paper, we rigorously document market access for farmers in two regions – Kilimanjaro and Manyara – of Northern Tanzania, which together comprise 6 percent of the land area and population of the country. Our data collection exercise spans the entire supply chain of maize (output) as well as of fertilizer (input) in all 1,183 villages in these two regions, including (1) surveys with a random sample of 2,845 farmers in 246 randomly selected villages; (2) surveys with 532 agro-input retailers ("agrovets"), effectively spanning the universe of input retail locations; (3) a retrospective panel of buying and selling prices of maize from a sample of maize-sellers in each of the 226 markets in the area; (4) surveys with transportation operators which measure road quality, travel times, and travel costs; and (5) driving times and distances from Google Maps API.

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 and output prices is equivalent to about 50% of the mean.

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 the remoteness of any location is proxied by two measures: (a) the population-weighted distance from a set of 5 major urban centers, and (b) the elasticity-weighted trade cost from the same set of hubs (calculated analogously to Donaldson and Hornbeck, 2016). We find that a standard deviation increase in remoteness is associated with a 9-20 percentage point reduction in the probability of using fertilizer and a 4-9 percentage point reduction in the probability of selling maize. These effect sizes are meaningful: input usage in the most remote villages is only a third of that in the least remote villages, while maize sales are only half as high.

While we find clear evidence of reduced market access in more remote villages, and while it is intuitive that this reduced access will affect the choice set and decisions of farmers, it is not possible to quantify these effects in the reduced form alone, since remote villages and villagers may differ from proximate ones in other econometrically unobservable ways not directly related to access to markets. To evaluate the effect of market access on input adoption, our third contribution is to

[1] Transportation infrastructure is particularly underdeveloped in Africa. 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.

develop a quantitative spatial model of fertilizer adoption, in which the decision to adopt fertilizer is based on local output prices, innate farmer productivity, the distribution of delivered input prices and retailer quality, and idiosyncratic shocks. Transportation costs affect the distribution of prices by increasing the costs for farmers to reach a particular agrovet to buy inputs, as well as costs to reach the local market to sell their harvest.

On the input side, the structure of the model (which is similar to Eaton and Kortum, 2002) facilitates a decomposition of choosing an agrovet into three components: (1) the decision whether to adopt; (2), the decision of which location to buy from; and (3) the decision of which retailer to pick within that location. Farmer surveys record (1) and (2) and thus allow us to calibrate local factors that may affect adoption, as well as the implied trade costs incurred while sourcing from each agrovet location. To estimate trade costs, we derive a structural multinomial logit specification that estimates the implied iceberg trade costs to each location as a function of distance. The results suggest that transportation costs are large: our preferred specification yields estimates of local iceberg costs that are approximately 4% ad-valorem per kilometer of travel, which translates to an average of approximately 30% when buying from the closest agrovet. When comparing this estimate to data collected in our surveys, pecuniary costs make up approximately 43% of this overall travel cost, suggesting that there are significant non-pecuniary costs of travel (which may include the opportunity cost of the time to travel, risk-aversion related to potential stock-outs, or information frictions). After estimating trade costs, we use the model to build a market-clearing condition for fertilizer for each agrovet, which is a function of the expected spatial distribution of fertilizer expenditures by each farmer and the probability that a farmer at each location adopts at a given agrovet. We balance these market clearing conditions by finding a vector of agrovet "amenities" that exactly rationalize the market-shares of each agrovet. After doing so, we are able to calculate a precise measure of market access for fertilizer.[2] The estimated measure of market access for fertilizer falls approximately 50% per standard deviation increase in remoteness.

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 similar to the expected reduction in travel time if roads were upgraded (Casaburi, Glennerster and Suri 2013). This policy roughly doubles adoption relative to baseline, and also reduces the remoteness gradient by 39% for a binary measure of using fertilizer, and 59% for total fertilizer expenditures. We also leverage our detailed transport surveys to assess a counterfactual in which transport improvements are targeted toward main roads and rural roads separately. While improving both types of roads increases adoption, it is only the improvement of main roads which reduces the remoteness gradient (surprisingly, improving rural roads only has no effect on the gradient). The reason for this is that there are not many retailers located in remote areas, and so farmers in remote areas typically must travel on main roads to reach a retailer. We also study hypothetical entry counterfactuals, and we find that agrovet entry in remote areas has a larger effect

[2]This is similar in spirit to Redding and Venables 2004; Redding and Sturm 2007; Head and Mayer 2011; and Donaldson and Hornbeck 2016.

on adoption, but entry into more remote areas is less profitable. Finally, we halve transportation costs for reaching output markets, and find a slightly smaller change in adoption, though a slightly larger reduction in the remoteness gradient (relative to input market access).

This paper sits at the intersection of trade and development economics. On the development side, we contribute to a literature examining why sub-Saharan Africa has lagged behind the rest of the developing world in agricultural technology adoption. Many studies find evidence of large *yield* increases due to 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. 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.

Our paper is related to a growing literature about the effect of transportation infrastructure improvements on development outcomes and on the spatial distribution of economic activity,[3] which includes outcomes other than just prices, such as consumption, farm and human capital investments, migration, and occupational choice. In our paper, we focus narrowly on the specific effect of transportation costs on goods market access (i.e. transportation costs and the presence of intermediaries and the prices they charge) in isolation, without changing other margins.[4]

Our work is related to a voluminous trade literature which attributes spatial price differentials to three primary components – marginal trade costs (e.g. Donaldson, 2018; Eaton and Kortum, 2002; Keller and Shiue, 2007; Sotelo, 2019), spatially varying mark-ups (Atkin and Donaldson, 2015; Asturias et al., 2019), and the organization of intermediaries (Allen and Atkin, 2016; Dhingra and Tenreyro, 2017; Bergquist and Dinerstein, 2019; Casaburi and Reed, 2019; Chatterjee, 2019). Our work is particularly related to Atkin and Donaldson (2015), who estimate trade costs in a setting where an intermediary buys products at wholesale prices, transports them to distant markets, and sells directly to consumers. In contrast, we are interested in how trade costs affect the decisions of producers (in this case, farmers) regarding buying intermediate goods through their access to retailers and output markets.

Other general equilibrium trade models also assess the link between trade costs, price gaps and

[3] A partial listing of papers includes Aggarwal (2018), Aggarwal (2019), Alder (2019), Adukia et al. (2019), Asher and Novosad (2019), Bird and Straub (2016), Brooks and Donovan (2019), Bryan and Morten (2019), Gertler et al. (2019), Ghani et al. (2016), Khanna (2019), Morten and Oliveira (2016), Shamdasani (2016), and Storeygard (2016). See Donaldson (2016) for a review.

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

technology adoption, though most existing work focuses on trade between larger cities and markets, typically using trucks or trains (Atkin and Donaldson 2015; Porteous 2019, 2020) whereas our paper focuses on last-mile costs to farmers (which are usually incurred using smaller private vehicles or simply on foot). Consequently, the per-unit travel costs we study are much larger than estimated in prior work. For example, we measure the cost of transport at $4.74 per ton-km, compared to $0.29 per ton-km in Porteous, while our estimated ad-valorem travel costs to retailers is 30% over just 6.8 kilometers, whereas Atkin and Donaldson (2015) estimate ad-valorem costs of 10-20% over a distance of approximately 720 km. Our paper is also differentiated by developing a new approach to estimating implied iceberg costs for farmers to reach retailers, as revealed by sourcing decisions that are measured through surveys.

The paper proceeds as follows. Section 2 provides background and context on our study region. Section 3 explains the data, and documents summary statistics. 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 is used for running policy counterfactuals in Section 6. Section 7 concludes.

2 Background on input and output markets and study regions

This study took place in the Kilimanjaro and Manyara regions[5] 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 poor: the paved road density is 2.2% in Kilimanjaro (i.e. 2.2 kilometers of paved roads per 100 square kilometers of area), 0.15% in Manyara, and 0.7% in Tanzania overall (TanRoads and PMO-RALG, 2014), compared to 68% in the US and 134% across the OECD.[6]

The main crop grown in this area is maize. There are two growing seasons in this area: a longer, more productive "long rains" season, 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 do not plant in the short rains. 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 and almost all of what is used is imported via the port at Dar es Salaam (FAO, 2016; Hernandez and Torero, 2011), and then transported throughout the country over surface roads. 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

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

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

then must travel further inland to reach farmers in the destination country. Urea is the most widely used and sold fertilizer, with about a third of the farmers using it and 84% of the agrovets selling it, followed by DAP, which half of the agrovets sell and 10% of the farmers use. For the agrovets in our sample, procurement almost universally happens at wholesalers located in major cities or towns (agrovets that procure from hubs account for about 94% of market revenue), and the vast majority of agrovets (90%) travel to the wholesalers themselves for procurement.

On the output side, farmers can sell to itinerant buyers – "agents" – who visit the farmgate soon after harvest and therefore buy at the low post-harvest prices. Alternatively, a farmer could travel to a market with their maize and find a buyer, or engage in informal sales near their homestead.

3 Data and summary statistics

We have four main sources of data: 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.

3.1 Agrovet surveys

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

3.2 Farmer surveys

We conducted farmer surveys in 246 randomly selected villages in three waves. The first wave covered 115 villages in Kilimanjaro in early 2016, the second wave 97 villages in Kilimanjaro in 2017, and the third wave 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 household and demographic information. Though the exact questions varied from survey to survey, the general format was 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 through a random walk procedure[7], while in rounds 2-3 households were pre-identified from a listing exercise

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

conducted with village leaders. In Wave 1, we sampled only 5 households per village for budgetary reasons, while in Waves 2-3 we selected 18 households per village. We find no qualitative difference in results from the two methods, and thus we pool all surveys together in the analysis.[8]

3.3 Measuring transport costs

One of the primary contributions of this work is to carefully document transport costs incurred by farmers. We measured transportation costs in several ways. First, we collected the GPS location for every village,[9] from which we calculated driving times and distances using the Google API (via the statistical program R). Second, we conducted surveys of transportation operators in every village in our sample, which were either motorbike taxis ("Boda Bodas"), or consumer van taxis ("Dala Dalas"). In each village, we asked up to 3 operators how much it cost to travel to the major towns in Kilimanjaro (Arusha and Moshi), the capital city (Dar es Salaam), and importantly, the market center as defined for the sampling procedure.[10]

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

3.4 Maize prices

To measure maize prices, we first identified the local market for each village.[11] These markets are typically located some distance from farmers, and market activity occurs on pre-specified days. Enumerators visited these markets in September and October of 2017 in Kilimanjaro and February to May of 2018 in Manyara. During these visits, enumerators sampled up to 3 maize sellers per market and collected pre- and post-harvest selling 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 a panel analysis.

next homestead after five minutes of walking, whichever came first.

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

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

[10] In Manyara, we also asked about trip costs and times to Babati, Dodoma, and Tanga.

[11] This was done by visiting ward offices (the ward is the second-lowest administrative level, just above the village) and asking the ward officer to list the market that people from each village frequented.

3.5 Summary statistics on villages

A map of Kilimanjaro and Manyara is shown in Figure 1. Summary statistics on villages are provided in Table 1. The average village has 480 households (see table notes), and is located 6.5 kilometers from the nearest market center. It takes about 40 minutes of driving to reach the market and return, and a round-trip costs about $1.90 on average. The average village is over 70 km away from the nearest major hub, and a round-trip to the hub would take about 3 hours and cost $6.[12] Some villages are extremely remote – the standard deviation of time to a hub is about 2 hours.

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

4 Results

In this section, we examine dispersion in input and output prices, and document the relationship between remoteness and prices, adoption, yields, and other related outcomes.

4.1 Travel cost-adjusted price dispersion

For each village, we assume that farmers are free to travel to any agrovet/market to buy inputs or sell output, but must incur a transportation cost, which we calibrate using information from transport surveys and Google distances. Specifically, using Google API, we calculate the route from every village to every agrovet/market. This route will involve either (1) traveling only on local roads over a relatively short distance, or (2) using local roads to connect to trunk roads. We calibrate the costs of local and trunk roads using our transport operator surveys, and information collected by enumerators during their own travel. We present these results in Table 2. Columns 1-3 show the costs of traveling from market centers to hub towns, which involves primarily traveling on trunk roads. We find a cost of about $0.021 per km, or $1.26 per hour of travel. The remaining columns present figures for rural roads (i.e. villages to market). As expected, we find higher costs for rural travel: $0.088 per km, or $2.61 per hour of travel. We use these estimates to calibrate costs from every village to every retailer, depending on whether the route includes only local roads or involves travel along trunk roads as well.[13]

With these costs, we calculate a travel cost-adjusted price of fertilizer for every village in two

[12] The average income of a farmer from all non-farming and farming sources is about $610 (Table 3).

[13] The specific methodology is as follows. For travel from a village to another village using only local roads, we do not have direct survey measures of travel costs, and so we imputed the travel costs using the Google distance between the two villages and the average travel cost per km on a rural road. For non-local travel which involves connecting to a larger trunk road, we calculate costs more directly. In this case, getting from an origin village to a destination involves going from the origin to the trunk road, traveling along the trunk road for some distance, and then turning off the trunk road to travel to the destination location. Our surveys contain a direct measure of the cost of going from any village to the main road (via the market), since we asked transport operators this question. Most markets are directly on the main road; for those that are not on the road, we calibrate any remaining distance using the cost of local travel.

ways. First, we define the minimum travel cost-adjusted price that is available to villagers as follows:

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

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

Equation (1) assumes that farmers have information on prices for every location, that they are free to travel to any location, and that they choose the lowest price from this menu. While we argue that this is the appropriate benchmark, some readers might argue that farmers make decisions using a simpler decision rule. While it is impossible to characterize all possible alternative decision rules, the most extreme possibility is that farmers only travel locally and prices in all other locations are irrelevant (i.e. costs beyond the nearest retailers are effectively infinite). Accordingly, we also conduct our analysis using the following travel cost adjusted price to the *nearest* agrovet.

$$r_v^{nearest} = r_{nearest} + c_{nearest,v} \tag{2}$$

On the output side, we construct the *maximum* travel cost-adjusted selling price for maize using a similar approach:

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

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 transporting the bag is proportional to the weight. Thus, a trip to the market and back to sell 120kg of maize requires 3.7 trips (2 for the farmer and 1.7 for the bag). Finally, as motivated above, we also calculate the price if farmers only transact at the nearest maize market.

$$p_v^{nearest} = p_{nearest} - c_{nearest,v} \tag{4}$$

We calculate these prices for every village-agrovet and village-market pair. Figure 2 plots CDFs of village-level best prices of inputs and output, adjusting for travel costs, and shows tremendous heterogeneity in prices across villages. In Panel A, the difference of the travel cost-adjusted price for maize between the 90th and the 10th percentile is about 54% of the mean, while the standard deviation is about 23% of the mean. In Panel B, the 90-10 difference for the travel cost-adjusted price for fertilizer is about 43% of the mean, and the standard deviation is about 19% of the mean.

To give a sense of the variation in profitability in using fertilizer, Panel C of Figure 2 calculates the ratio of the best travel-cost-adjusted maize price (per kg) to the best travel-cost-adjusted urea price (per kg). The 90-10 gap is 88% of the mean and the standard deviation is 34% of the mean. Web Appendix Figure A1 shows analagous figures for prices at the nearest location, and figures look similar.

4.2 Reduced form analysis

In this subsection, we explore the relationship of these best prices (along with other outcomes like input usage, selling behavior, and access to retailers and markets) to the remoteness of the market.

4.2.1 Specification

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

$$m_{vt} = \beta_r \cdot R_v + \epsilon_{vt} \tag{5}$$

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

For the measures of village-level market conditions estimated in (5), 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 farmer-specific characteristics. Therefore, these are estimated as:

$$m_{fvt} = \beta_r \cdot R_v + \beta_X X_{fvt} + \epsilon_{fvt} \tag{6}$$

where subscript f refers to farmer and X_{fvt} is a vector of other controls. These controls include information from the survey, such as land ownership, income, assets, education and other demographic characteristics, as well as soil information from the FAO-GAEZ. All farmer-level results are presented both with and without these controls.

4.2.2 Defining remoteness

To measure the remoteness of each village v, we focus on its proximity to selected "hubs" that are within or near the study regions: Arusha, Babati, Dodoma, Moshi, and Tanga. These locations are chosen because distributors for both maize and fertilizer are commonly located here.[14]

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

[14] Appendix Table A3 presents input- and output-distributor locations, showing that nearly all of them are located in the towns of Arusha, Moshi, and Babati. We extend this list to also include the regionally important cities of Tanga and Dodoma, and our complete set of hubs are marked with "stars" in Figure 1. We do not include Dar es Salaam in the remoteness measure as its high relative population leads it to overwhelm all the other hubs.

$$remoteness_v = \sum_h d_{hv} pop_h \tag{7}$$

where pop_h is the (relative) population of hub h (i.e. the population of that hub divided by the population of all hubs) and d_{hv} is distance from village v to that hub. In this measure, relative population is used as a proxy for the importance of each city in terms of availability of goods and average prices. Unlike Donaldson and Hornbeck, we use distance to measure proximity to hubs, rather than calculate the ad-valorem costs of travel and estimate a distance elasticity. We do this because it simplifies the construction of the measure substantially.[15]

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

$$MA_v = \sum_h \tau_{hv}^{-\theta} pop_h \tag{8}$$

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

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

where $cost_{hv}$ is the estimated cost to get from village v to hub h, 2.69 is the number of one-way trips required to travel to a destination and return with a 50kg bag of fertilizer (see section 4.1), and $avgprice$ is the average price of fertilizer in the sample (measured at agrovets). We choose fertilizer as the benchmark good to measure ad-valorem costs, since it is the focus of the paper, and also because agrovets commonly report traveling to hubs to stock fertilizer. To measure the elasticity term, $-\theta$, we appeal to estimation later in the paper where the substitution elasticity across agrovets is estimated to be approximately -7.5, though the results are robust to other estimates.

We standardize both measures to have mean 0 and standard deviation 1 (and put a negative sign in front of MA_v, so that it measures remoteness rather than market access). The distributions of these variables are illustrated in Appendix Figure A2. Both measures feature two modes, and for both measures 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).

4.2.3 Summary statistics and correlations with remoteness

Table 3 presents summary statistics, and shows how these variables vary with remoteness. From Panel A, we see a number of differences: farmers in more remote areas are less educated, own fewer assets, have less access to finance, and earn less income from sources outside of farming. These

[15] Despite this simplification, we show in the technical appendix that there exists a first-order approximation that links the two measures.

farmers also tend to have larger families and larger farms.

Panel B shows production capacity, based on GIS data from the FAO-GAEZ database, which provides information on counterfactual yields with and without inputs. At the mean, the FAO estimates that using inputs would more than quadruple yields. There is some evidence that more remote areas have lower returns to inputs – a 1 standard deviation increase in remoteness is associated with 14% lower increases in yields. However, we note that yield increases remain very large even in the most remote areas – for the most remote villages (3 standard deviations away), yields with inputs are still 140% higher than without inputs. We control for these measures in our main regressions, and develop a strategy to absorb factors like these within the spatial model. Panel C shows harvest output from the most recent long rains. While the relationship between total yields and remoteness depends on the measure used, yield per acre is lower in areas located farther away from hubs. In particular, a standard deviation increase in either measure of remoteness is associated with a reduction in harvest output per acre of about 20%. This is consistent with lower input usage in rural areas, or with differences in other factors such as soil quality.

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

4.2.4 Access to input markets

Table 4 shows how the two remoteness measures correlate with access to input markets. We first tabulate access to retailers within 10 km, which is a distance that is reasonably traveled by farmers.[16] Panel A shows several measures, including a dummy for retailer presence within 10 km, the number of retailers within 10 km, and the minimum distance to a retailer. On each measure, we find clear evidence of reduced access to retailers in more remote villages, all significant at 1%.

Panel B1 of Table 4 shows our preferred measure of input market access, the minimum travel cost-adjusted price that is available to farmers. We find that one standard deviation of remoteness raises prices by \$2.33-2.41, 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 sample. We then decompose this price difference into differences in the retail price itself, and those in the travel cost. We find that the retail prices (at the optimal location) and transportation costs (to the optimal location) are approximately equal in their contribution toward the increase in minimum delivered prices.[17]

[16] Appendix Figure A3 shows a CDF of the distance farmers travel to access inputs, conditional on purchase. We find that approximately 70% of purchases are made within 10 km of a farmer's village, and 85% within 20 km.

[17] In Appendix Table A6, we present 2 robustness checks. First, since we only surveyed retailers within the regional boundaries, we have no information on retailers in neighboring regions. It is possible, therefore, that there exist lower-priced retailers just across the border, causing us to potentially overstate travel cost-adjusted prices. To address this, in Panel A1, we drop all villages within 10 km of regional boundaries – the results are actually stronger. Second, while we had high survey completion rates among agrovets (91% – see Appendix Table A1), we nevertheless do not have the universe of retail options. This suggests that retail price heterogeneity may be understated. To address this, we conduct a bounding exercise in Appendix Table A6, Panel B, where we estimate the distribution of prices within regions. We then assign prices in the tails of this distribution (the 10th or 90th percentile) to missing agrovets in a

Panel B2 of Table 4 presents our secondary measure of access, the travel cost-adjusted price at the nearest shop. By definition, the travel cost-adjusted price is higher than in B1 (by about 10%); in particular, because farmers do not shop around, the retail price is higher and the travel cost is lower. As before, we find roughly equal contributions of each to the remoteness gradient.

These numbers give us a sense of the (pecuniary) ad-valorem equivalent transport costs for buying fertilizer. For the minimum price (Panel B1), transport costs are about 22% of the optimal purchase; for the nearest retailer (Panel B2), they are about 13%. While these costs are already substantial, when we examine farmer behavior more carefully in Section 5.2, we find that farmers behave as if costs are even larger than these pecuniary costs. [18]

4.2.5 Access to output markets

Table 5 performs a similar analysis, but on the output side. As before, Panel A shows that more remote villages are less likely to have a market within 10 km, and the nearest market where maize is sold is located farther away. Panel B1 shows travel cost-adjusted prices for maize. Since there are large seasonal price fluctuations in rural Tanzania (as in much of rural Africa),[19] we use a price for the single point in time which is most relevant for farmers: immediately post-harvest (our surveys show that most farmers who sell do so shortly after harvest). We find that across both remoteness measures, travel cost-adjusted prices of output 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 with remoteness, transport costs to their best maize market rise by $3.9 with each standard deviation in remoteness, overwhelming the increase in the price of maize. In Panel B2, we repeat the analogous exercise from the input market to evaluate the impact of remoteness when farmers simply choose the *closest* weekly maize market to sell their harvest. By definition, average travel cost adjusted sales prices are lower, and empirically the magnitude is large (about 50%). As in Panel B1, we find that this price declines with remoteness, and in fact the point estimate is similar. However, the decomposition between the retail price and the travel cost is very different: for the nearest price, the retail price falls and the travel cost rises.

Finally, we show one other measure of price, in this case measured at the village level. First,

way that attenuate our regression results – for example, in remote areas, we assign agrovets low prices. This exercise lowers the coefficient marginally, but the qualitative results are unchanged.

[18] How much price heterogeneity can be explained by retailer pricing behavior? While causal identification is challenging (since entry is endogenous), we provide descriptive evidence in Appendix Table A5. From Panel A, we find some evidence that more remote shops sell different products (further complicating inference) – remote shops are less likely to sell fertilizer but more likely to sell seeds. We find strong evidence that retailers face higher costs of procuring supply from wholesalers; in this setting, retailers typically travel themselves to wholesalers to purchase inventory, and so it is intuitive that these procurement costs are higher in remote areas. In Panel B, we examine wholesale and retail prices. We find evidence that remote retailers charge higher prices but they also face higher wholesale prices (perhaps because competition is weaker among available wholesalers). Ultimately we find that mark-ups are no higher in remote areas. This descriptive evidence suggests that pricing behavior is likely a secondary factor in explaining higher prices.

[19] Aggarwal, Francis and Robinson (2018) document an average price increase of about 46% over the season for the years 2006-16 in Kisumu market in neighboring Kenya; Bergquist, Burke, and Miguel (2019) document increases in the range of 15-30% for a sample of markets in the east African region.

in Panel B3, we report coefficients from farmers' self-reported "going price" of maize after the last harvest, regressed on measures of remoteness. Consistent with the above, the going price in the village is decreasing in remoteness. This is intuitive if maize agents are traveling from the larger population centers (which are used to construct our remoteness measures), and offering lower selling prices to compensate for the higher costs of travel. Overall, whether searching for the best market, or selling locally, the returns from selling maize are clearly lower in more remote regions.

4.2.6 Farmer decisions

The results so far show clear evidence of reduced market access in more remote areas for both inputs and output, and of higher prices for inputs, lower (travel cost-adjusted) prices for output, and lower "going" prices for output within the village. These results lead us to expect lower input usage and maize sales in more remote areas. We investigate this in Table 6, where we present results with and without a full set of farmer controls. In Panel A, we present the extensive and intensive margin of input use, for both seeds and fertilizer. In all specifications, these relationships are strong (significant at 1%) and large. We find that use of fertilizer is 9-20 percentage points lower in villages 1 standard deviation away, and that of hybrid seeds is 5-11 percentage points lower. Since the distance between the least and most remote regions is about 5 standard deviations, the regressions predict at least 45 percentage point lower usage of fertilizer in the most remote villages, which translates to about 80% of the mean in the least remote areas. The effect for seeds is smaller but still evident.

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

Consistent with this, Panel C shows buying behavior. Remote farmers are more likely to buy maize and to be net buyers of maize. Interestingly, we find a lot of heterogeneity in net buying behavior - 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 at all).

5 Model

In this section, we quantify the impact of access to input and output markets by developing a spatial model of fertilizer adoption. In the model, we develop a rigorous framework of retailer choice, while including other factors that affect adoption but are not related to input access. Ultimately, using a model specified measure of input access that we estimate, we run counterfactuals that study the role of transportation costs in the adoption decision.

5.1 Model Preliminaries

Production and Inputs

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

$$Y_{i0} = \widetilde{\theta}_{i0} K_i^{\alpha_0} L_{i0}^{1-\alpha_0} \tag{9}$$

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

$$\begin{aligned} \Pi_{i0} &= \alpha_0(1-\alpha_0)^{\frac{1-\alpha_0}{\alpha_0}} \widetilde{\theta}_{i0}^{\frac{1-\alpha_0}{\alpha_0}} p_i^{\frac{1}{\alpha_0}} w_i^{-\frac{1-\alpha_0}{\alpha_0}} K_i \\ &= \theta_{i0}\pi_{i0} \end{aligned} \tag{10}$$

where $\theta_{i0} = \alpha_0(1-\alpha_0)^{\frac{1-\alpha_0}{\alpha_0}} \widetilde{\theta}_{i0}^{\frac{1-\alpha_0}{\alpha_0}}$ and $\pi_{i0} = p_i^{\frac{1}{\alpha_0}} w_i^{-\frac{1-\alpha_0}{\alpha_0}} K_i$.[20] 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 has both labor and fertilizer as variable inputs, while maintaining the basic Cobb-Douglas assumption (for tractability). When using fertilizer, farmers not only have a choice of how much fertilizer to buy, but also which retailer to choose. Supposing that farmer i buys fertilizer from agrovet j in location v, the production function is written as:

$$Y_{ijv} = \widetilde{\theta}_{ijv} \left(\theta_i K_i\right)^{\alpha} L_{ijv}^{(1-\alpha)\beta} M_{ijv}^{(1-\alpha)(1-\beta)} \tag{11}$$

Note we are assuming that the exponents on capital and labor may be different for the technology with fertilizer, which as we will show below, allows for output prices to affect adoption decisions (while maintaining the analytical simplicity of a basic Cobb-Douglass technology).[21] Further, when using fertilizer, there are two additional productivity terms to consider. The first is the known local productivity of using fertilizer, θ_i, which in the production function, scales the effective amount of land for farmer i. The second is a productivity shock for farmer i, $\widetilde{\theta}_{ijv}$, that potentially varies by the agrovet j and location v where the fertilizer was purchased. We discuss this particular productivity shock when solving for optimal retailer choice.

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

$$\Pi_{ijv} = \theta_{ijv}\pi_i r_{ijv}^{-\sigma} \tag{12}$$

[20] We assume for tractability that all farmers internalize a market price for maize and labor.

[21] The Cobb-Douglas framework to model agricultural production shares similarities with recent work in Chatterjee (2019), Gollin and Udry (2019), though unlike the latter, we do not model allocations across different plots within the household.

where $\sigma \equiv \frac{1-\alpha}{\alpha}(1-\beta)$, $\pi_i = \theta_i p_i^{\frac{1}{\alpha_0}} w_i^{-\beta\frac{1-\alpha_0}{\alpha_0}} K_i$, and $\theta_{ijv} = \kappa_2 \widetilde{\theta}_{ijv}^{\kappa_1}$.[22] Here, the profitability of fertilizer is a function of the productivity shock, θ_{ijv}, the (delivered) price of fertilizer itself, r_{ijv}, and deterministic profits based on local factors and technology, π_i.

Input and Agrovet Choice

Farmers choose whether to purchase fertilizer, and if so, how much and from where. These decisions are affected by prices for fertilizer at each agrovet location, the productivity shock received in buying from a particular location, and the round-trip travel costs. Suppose that the set of villages that contain an agrovet is defined as $\mathcal{V}$, where the price charged at location $v \in \mathcal{V}$ by agrovet j is r_{jv}. The per-unit cost to the farmer i, inclusive of transport costs, will be written as $r_{ijv} = r_{jv}\tau_{iv}$, where τ_{iv} is an iceberg trade cost for farmer i in traveling to v and back. The assumption of iceberg trade costs will facilitate a decomposition of the model that aids estimation and calibration.

We assume that θ_{ijv} is a random variable that measures the benefit of i purchasing at agrovet j in location v. These latter benefits could represent other inputs purchased in location v (hybrid seeds, for example), availability of extension services at location v, or perhaps other networking and information that is acquired at location v that may affect profitability. Further, it may represent the probability of getting bad or adulterated inputs at a given retail location, or given the functional form of (12), measurement error in the price at a retail location. Whatever the interpretation, for analytical convenience we assume that θ_{ijv} is distributed according to a Fréchet distribution with location parameter T_{jv} and dispersion parameter ε. Precisely:

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

That is, while each farmer may get a random draw from this distribution, its central moments are specific to the retail location itself. Using this distributional assumption, the unconditional distribution of profits for farmer i buying from agrovet j in location v is written as:

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

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

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

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

[22] κ_1 and κ_2 are constant functions of model parameters.

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

$$\lambda_{ijv} = \frac{T_{jv}\pi_i^{\varepsilon} r_{ijv}^{-\varepsilon_a}}{T_{i0}\pi_{i0}^{\varepsilon} + \sum_{v' \in \mathcal{V}} \sum_{l \in \mathcal{J}_{v'}} T_{lv'}\pi_i^{\varepsilon} r_{ilv'}^{-\varepsilon_a}} \tag{13}$$

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

$$\mu_i = \frac{\sum_{v' \in \mathcal{V}} \sum_{l \in \mathcal{J}_{v'}} T_{lv'} r_{ilv'}^{-\varepsilon_a}}{T_{i0}\left(\frac{\pi_{i0}}{\pi_i}\right)^{\varepsilon} + \sum_{v' \in \mathcal{V}} \sum_{l \in \mathcal{J}_{v'}} T_{lv'} r_{ilv'}^{-\varepsilon_a}} \equiv \frac{\Phi_i}{\Phi_{i0} + \Phi_i} \tag{14}$$

In (14), we define the two terms that fully characterize the adoption decision for each farmer i. First, we define farmer i's *market access* to inputs as $\Phi_i = \sum_{v' \in \mathcal{V}} \sum_{l \in \mathcal{J}_{v'}} T_{lv'} r_{ilv'}^{-\varepsilon_a}$, which after imposing the iceberg assumption and simplifying, can be written as

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

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

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

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

5.2 Calibrating the Farmer's Problem

In linking the model to the data, we will make use of the novel cross-sectional surveys as described in sections three and four, and match this data to adoption and location-choice probabilities as specified by the model. To make clear what we have, and what we do not, we note that (13) can be broken up into the probability of adoption for i, μ_i; the probability i buys somewhere at location v conditional on adopting at all, $\lambda_{iv|adopt}$; and finally, conditional on adopting from an agrovet at

location v, the probability that agrovet j is chosen, $\lambda_{j|adopt\,at\,v}$:

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

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

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

Estimating Transport Costs through Location Choice

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

$$\lambda_{iv|adopt} = \frac{\tau_{iv}^{-\varepsilon_a}\phi_v}{\sum_{v'\in\mathcal{V}}\tau_{iv'}^{-\varepsilon_a}\phi_{v'}} \tag{15}$$

To estimate equation (15), we need a dataset that identifies when each farmer i chooses location v to purchase fertilizer. Thus, defining $\mathcal{I}$ as the set of farmers who adopt, and $\mathcal{V}$ as the set of locations with an agrovet, we construct a $\mathcal{I}\times\mathcal{V}$ dataset of bilateral visit indicators. For each bilateral combination, we will also measure the routed distance in kilometers between the farmer's village and the potential purchase location, $dist_{iv}$ (similar to Section 3).

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

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

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

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

$$\lambda_{iv|adopt} = \frac{\exp\left(d_v + \sum_b \beta_b D_{iv}^b\right)}{\sum_{v' \in \mathcal{V}} \exp\left(d_{v'} + \sum_b \beta_b D_{iv'}^b\right)} \tag{16}$$

Equation (16), and its linear alternative, can be estimated by McFadden's alternative-specific conditional logit. The results from doing so are presented in Table 7. In the first column, we present the linear specification of distance. Assuming $\varepsilon_a \approx 7.5$ (which will be supported by later results), the results suggest that iceberg transport costs for fertilizer are 2.3% 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, while Column 3 reports the ad-valorem equivalent per kilometer when evaluated at the farthest distance that defines each bin, and with trade costs compounded each kilometer.[23] The estimates suggest costly travel for farmers acquiring fertilizer. To interpret the coefficients, we take two approaches. In the first, we compare two locations with the same "return" from fertilizer, d_v, and then focus on the reduction in probability if one is (0,5] km away rather than 0 km away (in the same village). In this case, the probability that one chooses the location (0,5] km away compared to 0 km away (in the home village) for idiosyncratic reasons that overcome trade costs is 0.25.[24]Alternatively, we can interpret the results as log changes in trade costs via $\log(\tau_{iv}) = -\frac{1}{\varepsilon_a}\sum_b \beta_b D_{iv}^b$, where dividing the coefficient estimates by ε_a gives us log trade costs. Given the iceberg assumption, this is also interpreted as the log change in the delivered price. Thus, at the central estimate of $\varepsilon_a \approx 7.5$, the comparison is equivalent to approximately 4% ad-valorem trade cost per km for the first three bins (up to 15km). When evaluated to the typical closest agrovet (6.7 km), the ad-valorem equivalent trade cost is about 30%. Beyond this, the ad-valorem equivalent per kilometer falls modestly, which is consistent with our transport surveys and also the likelihood that longer distances require a more efficient means of travel (though still at a high overall cost, on average).

An advantage of using revealed adoption choices to estimate trade costs is that the iceberg estimates described above may include both pecuniary and non-pecuniary costs of travel. Again, when including both, we estimate that the ad-valorem equivalent to the closest agrovet is about 30%. In section 4.2.4, we summarize the ad-valorem equivalent travel cost of buying from the nearest retailer, estimating 13%. Thus, the pecuniary cost is roughly 43% of the overall cost of

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

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

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

travel, suggesting significant non-pecuniary costs of acquiring fertilizer.

Finally, similar to Section 4.2, we calculate best trade-cost-adjusted prices for agrovets for all villages in the region, using the binned estimates of iceberg costs as described above. These results are presented in Panel D of Figure 2. Here, there is significantly more heterogeneity in best trade cost adjusted prices for fertilizer, again suggesting sizable non-pecuniary costs of traveling to acquire fertilizer. Precisely, at the median best travel cost adjusted price, the delivered price using our ad-valorem estimates is \$28.7 for a 50kg bag, which is approximately 25% higher than the median best-delivered price when using the pecuniary costs from transport surveys.

Model Calibration

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

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

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

where $\mathbb{E}\left[F_i | adopt\ at\ jv\right]$ is expected fertilizer expenditures by i, conditional on adopting at jv, and L_i is the village population to use as weights in the demand equation. As this conditional expectation is not observed in any practical way, we will appeal to the structure of the model to simplify to an unconditional expectation for fertilizer expenditures by i. Precisely, using the properties of the Fréchet distribution, it is straightforward to show that $\mathbb{E}\left[F_i | adopt\ at\ jv\right] = \mathbb{E}\left[F_i | adopt\right]$; that is, the expected expenditures conditional on adoption anywhere is the same as the expected expenditures at some j, conditional on choosing j.[25] Noting further that $\mu_i \mathbb{E}\left[F_i | adopt\right] = \mathbb{E}\left[F_i\right]$, and imposing the definition of $\lambda_{ijv|adopt}$, we get:

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

Finally, we can combine the agrovet-specific non-price attributes and the price into an "agrovet-

[25] See technical appendix for a proof.

effect" ($\eta_{jv} \equiv T_{jv} r_{jv}^{-\varepsilon_a}$), and also impose the specification for transportation costs, to get:

$$\mathbb{E}\left[v_{jv}\right] = \sum_i L_i \left(\frac{\exp\left(-\sum_b \hat{\beta}_b D_{iv}^b\right) \eta_{jv}}{\sum_{v' \in \mathcal{V}} \sum_{l \in \mathcal{J}_{v'}} \exp\left(-\sum_b \hat{\beta}_b D_{iv'}^b\right) \eta_{lv'}} \right) \mathbb{E}\left[F_i\right] \tag{17}$$

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

However, we again run into a number of issues as funding was not sufficient to survey more farmers and more villages. There are two issues to consider. First, while we surveyed approximately 18 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. Village adoption at the bounds complicates the calibration of the overall adoption decision (given the logit functional form). 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.[26] 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 the 1st percentile value of fertilizer expenditures per acre of land across the entire sample of farmers who adopt, $\left(\frac{F_i}{K_i}\right)_{1st}$.[27]

Focusing on the sampling of villages, if we assume that the farmer sample captures the entire geography of demand, there will exist agrovets in other locations that appear more remote than they actually are since no farmers were surveyed in that location. This will cause a bias in estimates of η_{jv} 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 (17), under the normalizing assumption that $\sum_v \sum_j \eta_{jv} = 1$.[28]

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

[26] 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.

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

[28] This normalizing assumption is required as the probabilities within the sum in equation (17) are homogeneous degree zero in η_{jv}. Eckert (2019, Lemma p. 28) uses a similar technique to infer services trade across locations in the US, and also provides a uniqueness proof for such a system of non-linear equations.

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

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

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

The estimated (log) values of $\hat{\Phi}_i$ and $\hat{\Phi}_{i0}$ are regressed on remoteness in Panel B of Appendix Table A7, where the relationship between standardized remoteness and market access, $\hat{\Phi}_i$, is significantly negative - a one standard deviation increase in remoteness leads to a 0.77 reduction in log market access. In contrast, there is a positive but statistically insignificant relationship between standardized remoteness and the outside option, $\hat{\Phi}_{i0}$.

5.3 Estimating Elasticities for Counterfactuals

We have fully calibrated the farmer's decision to adopt as a function of quality-adjusted access to markets, and shown that market access is poorer for remote villages. The results also suggest a slightly higher outside option to using fertilizer in remote markets that may be due to local suitability for fertilizer or other market conditions. To evaluate the role of trade costs in market access and output prices in the outside option, we further estimate the fundamental parameters of the productivity distribution and the production functions, with and without fertilizer.

To begin, we estimate the composite elasticity of substitution between agrovet options, ε_a, which is a function of the native Fréchet dispersion parameter, and the land (α) and labor (β) shares in the production function with fertilizer. Taking logs of η_{jv}, we get:

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

As η_{jv} is calibrated using revenue-expenditure market-clearing conditions, there is an endogeneity problem in estimating ε_a. To address this endogeneity, we appeal to the urban and trade literature (eg. Melitz and Ottaviano, 2007; Combes et al. 2012) in which larger markets induce greater entry and competition, leading to more efficient sellers and lower prices. Accordingly, we instrument for current agrovet prices using a significantly lagged (2011) population of the market catchment area that defines the location of the agrovet.[29] The regression using this instrument as well as district fixed effects yields an estimate $\varepsilon_a = 7.5$, which we use in later counterfactuals. A full set of estimates under OLS and IV is presented in Panel A of Appendix Table A7.

[29] We experimented with other instruments that attempt to leverage the insights from Berry, Levinsohn, and Pakes (1995) and Hausman (1996), though each has its own drawbacks in this context. BLP requires defining the relevant product characteristics for the unit defining price, and summing up the characteristics of competitors. However, the relevant set of competitor characteristics is unclear, and also some agrovets are located in isolated areas in which they are the only retailer. The Hausman technique uses the prices of the same product in other markets as instruments, but in this context, the retailers are often buying from the same distributors, rendering this instrument inappropriate.

Next, we estimate production parameters that are embedded in Φ_{i0}. Recall that $\Phi_{i0} = T_{i0}\left(\frac{\pi_{i0}}{\pi_i}\right)^{\varepsilon} = T_{i0}p_i^{\varepsilon_p}w_i^{\varepsilon_w}$, where $\varepsilon_p = \varepsilon\left(\frac{\alpha-\alpha_0}{\alpha\alpha_0}\right)$ and $\varepsilon_w = \varepsilon\left(\beta\frac{1-\alpha}{\alpha} - \frac{1-\alpha_0}{\alpha_0}\right)$; thus, the effects of any output price shock are a function of the relative importance of land in the production function. Further, β is important for the wage elasticity of adoption, as well as for decomposing ε_a into the native dispersion parameter ε and production parameters. In Appendix Table A8, we detail a simple estimator for maize production with and without fertilizer, and using data from the Tanzania Living Standards Measurement Study and Integrated Surveys on Agriculture (LSMS-ISA) produce estimates for α (0.431) and α_0 (0.570). Also using the LSMS-ISA, we use reported wages and labor and fertilizer expenditures to calculate the share of labor in variable factors; β (0.75).[30] Using these estimates, it is straightforward to calculate that ε=21.9. 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.

5.4 Agrovet Pricing and Markups

For a farmer, adoption is a function of a quality-adjusted delivered price for fertilizer at each agrovet, as well as other terms that represent the relative incentives to abstain from using fertilizer. When evaluating trade shocks, we could hold fertilizer prices fixed. However, while this may be fine for local shocks, for a large trade shock, such as a roads program, allowing retail prices and mark-ups to change is more realistic. We now derive the pricing problem for agrovets, and describe the calibration for mark-ups (similar to Berry, 1994).

The first order condition for an oligopolist is a mark-up over marginal cost:

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

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

$$r_{jv} = \frac{\varepsilon_{jv}^v - 1}{\varepsilon_{jv}^v}c_{jv} \tag{18}$$

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

$$\varepsilon_{jv} = -\varepsilon_a + \frac{\varepsilon - 1}{\varepsilon}\varepsilon_a\sum_i s_{ijv}\lambda_{ijv} \tag{19}$$

The elasticity equation in (19) provides clear intuition regarding the spatial distribution of

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

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.

We use the agrovet-specific elasticity to solve for the revealed marginal cost of selling fertilizer by using equation (18). The predicted markups have a mean of 14.8% (median = 13.8%), which is similar to what we find in the reduced form (13%). This is notable because estimated mark-ups do not use any marginal cost information measured for retailers.

6 Counterfactuals

In this section, we use the calibrated and estimated parameters to evaluate counterfactuals on input and output market access. To implement the counterfactuals, we solve for a new vector of fertilizer prices that solves the first order conditions for pricing in (18), while taking into account equilibrium changes in the farmer's 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. Our general hypothesis is that farmers are disadvantaged if agrovets are not close-by. We study these issues in two ways: (1) a reduction in transport costs, both on rural and main roads; and (2) agrovet entry.

Reducing Farmer-to-Agrovet 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 from farmer to retailer through a hypothetical roads improvement program. Such a cost reduction can also be motivated by speeds on trunk roads in Kilimanjaro being approximately 50% lower than US speeds. Figure 3 displays the results of this counterfactual on adoption rates (top-left panel) and log fertilizer expenditures (top-right panel) 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 unbinned raw village data.

In the top left panel, we find a large adoption effect of 36pp, which is more than twice baseline, and which alone accounts for 39% of the baseline adoption-remoteness relationship. Expenditure counterfactuals are even more pronounced, where average log expenditures rise approximately 1.5, though from an extremely low base in many cases. Nevertheless, the log-expenditure-remoteness gradient is cut by 59% in this counterfactual scenario. Thus, we conclude that holding local factors fixed, poor access to input markets contributes substantially to the reduced adoption levels in remote areas.

Rural roads vs. Main roads

The above counterfactual evaluates a 50% reduction in iceberg transport costs across all roads to reach agrovets, but does not distinguish between main and rural roads. This distinction may be important for a number of reasons. First, main roads may be congested, though paved, while rural roads may be uncongested but also of poor quality. Thus, both types of road improvements may be necessary to quantify. Germane to the discussion of access to retailers, critical for each farmer is whether one has to travel on rural roads to reach an agrovet, or must connect on a main road to reach an agrovet. This may be a particular problem in areas with limited entry, where farmers connect from one market area to the next using a combination of rural and main roads.

To evaluate the impact of adjusting costs on rural and main roads individually, we make an additional assumption about the iceberg trade cost, and then leverage the detail in our transport operator surveys to implement the counterfactual. We assume that the iceberg cost itself is an additive component of rural and main road costs: $\tau = 1 + t_r + t_m$. By doing so, we can derive that the iceberg cost, after reducing transport costs on rural roads by 50%, is $\tau' = 1 + \frac{t_r}{2} + t_m = 1 + (\tau - 1)\left(1 - \frac{1}{2}s_r\right)$, where $s_r \equiv \frac{t_r}{t_r + t_m}$ is the share of transportation costs incurred on rural roads. Our transport cost surveys facilitate calculating this share measure from every village to every agrovet. A similar approach can be used to isolate the impact of main road costs.

Using this approach, counterfactuals for cutting main road costs by 50%, and rural road costs by 50%, are presented in the middle and bottom rows respectively of Figure 3. Both counterfactuals increase adoption, but interestingly, the effect is larger when reducing main road costs. Further, rural transport costs have no appreciable effect on the remoteness gradient. The intuition for this result is that in remote areas, farmers tend to be farther from the nearest village with an agrovet, and that this travel is necessary and via a main road. Thus, to increase access to inputs, especially for rural areas, improvements in main roads are central to this policy goal. Indeed, there also appears to be a complementarity between the two effects, where the combined counterfactual from the top row of Figure 3 is 36pp, while the sum of the specific road-type counterfactuals is 32pp. Overall, road improvements increase adoption, but especially so for remote markets.

Distributor-Agrovet Costs

Next, we evaluate how the costs for retailers to source inputs from distributors affect the adoption decision. We document in our reduced-form analysis that sourcing costs rise significantly with remoteness. So, in another counterfactual that is related to input access, we halve distributor-agrovet transportation costs. The results are presented in the top panels of Figure 4. Adoption rises by about 1pp, or 4%, and yields a 2.7% reduction in the remoteness-adoption gradient.

Entry

An overarching question throughout the paper has been why agrovet access is worse in remote areas, and in particular, why agrovets enter intensely in other areas. While we do not present an

empirical model of entry (as in papers such as Seim, 2006), we do run a simple counterfactual to examine profitability of entry and any corresponding effects on adoption rates. 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 bottom panels of Figure 4 aggregate adoption (after entry) and entrant profits as a function of the remoteness of the village in which the entry took place. Clearly, profits are lower when entering more remote villages, though adoption effects are higher when entering the more remote villages (though the latter is not significant). The former relationship is particularly strong, where a one standard deviation increase in remoteness reduces the profitability of hypothetical entry by 39%. Thus, while access to agrovets in remote areas would appear to 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

We show in the reduced form that 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, 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, we assume that farmers optimally choose the best market to sell, while accounting for the costs of transportation. We measure the baseline best net-selling price, and then recalculate this price after halving transportation costs. 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 are presented in the middle left panel of Figure 4. We see a similar reduction in the adoption-remoteness gradient as the input market counterfactuals, and an adoption effect that is about 0.28, or 90%. Thus, this counterfactual has a smaller effect on adoption when compared with a 50% cut in farmer-retailer transport costs, but a similar effect on the gradient.[31]

With this same caveat in mind, 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

[31] However, care must be taken in comparing the two. In the case of the farmer-retailer transport costs, the transport cost cut is interacted with prices and calibrated agrovet quality terms, which adds noise to the shock. That is, while a farmer might be more likely to travel to any agrovet, the transport shock is not concentrated on the agrovet that is closest or that the farmer will necessarily choose. In contrast, the best net selling price is determined by a simple calculation of the maximum net price, with no probability of choosing different option. Thus, the noise in the farmer-retailer transport shock as it relates to distance will attenuate its effect on the gradient, though still provide a sizable effect on adoption.

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 middle-right panel of Figure 4. Here, adoption almost doubles, and the adoption-remoteness gradient falls by about 40%.

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.

7 Conclusion

We collect detailed data on transportation costs, input and output prices, input usage and maize sales along the supply chains for maize and fertilizer in all 1,183 villages in the Kilimanjaro and Manyara regions of Tanzania. We find that there is meaningful price dispersion, especially when accounting for travel costs. Access to retailers for inputs and buyers for output is much lower in remote regions, and consequently farmers in remote villages are much less likely to use fertilizer or sell output. Counterfactuals suggest that lowering transportation costs via road upgrading would substantially reduce the gradient between input usage and remoteness.

An important question is whether our results generalize to other settings. To provide some suggestive evidence on this question, we use secondary datasets and a dataset of prices we collected in our study area to examine how patterns compare between Northern Tanzania and other African countries.[32] First, we examine how price dispersion in Northern Tanzania compares to a set of 1,512 markets in 56 African countries. Using two approaches, we find that the degree of observed price dispersion in Northern Tanzania is comparable to other countries. Second, we use data from World Bank LSMS-ISA panel surveys to study how remoteness affects fertilizer adoption in other African countries.[33] Using both measures of remoteness available in the dataset (distance to the main market, and distance to a population center), we find a negative association between remoteness and technology adoption. Finally, we compile some statistics on the state of road infrastructure in other countries in the East African region (Table A12), and find that Tanzania is about average. The evidence therefore suggests that Northern Tanzania is not atypical of the region.

The results of our counterfactual simulations as well as the presence of similar patterns in other countries lead directly to the question of policy implications. Many African countries have experimented with input subsidies and these have had large adoption effects by directly lowering the delivered price of fertilizer even though the transport cost may have been unaffected (depending on retailer entry response to the program). However, most farmers fail to graduate out of the subsidy, perhaps in part because the market access issues remain unresolved, and therefore, inputs continue to be unprofitable at market prices. Our findings suggest that policies that lastingly affect input and output prices faced by farmers can have sustained effects. 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.

[32] For details, see Appendix F and Tables A10 and A11.

[33] The countries included here are Ethiopia, Niger, Nigeria, Malawi, Tanzania, and Uganda.

References

[1] Adukia, Anjali, Sam Asher, and Paul Novosad (2020). "Educational Investment Responses to Economic Opportunity: Evidence from Indian Road Construction." *American Economic Journal: Applied Economics* 12 (1) 348-376.

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

[3] Aggarwal, Shilpa (2019). "The Long Road to Health: Healthcare Utilization Impacts of a Road Pavement Policy in Rural India." Unpublished

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

[5] Alder, Simon (2019). "Chinese Roads in India: The Effect of Transport Infrastructure on Economic Development." Unpublished.

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

[7] Allen, Treb and David Atkin (2016). "Volatility and the Gains from Trade." Unpublished.

[8] Asher, Sam and Paul Novosad (2019). "Rural Roads and Local Market Development." Forthcoming, *American Economic Review*.

[9] Asturias, Jose, Manuel Garcia-Santana, and Roberto Ramos (2017). "Competition and the welfare gains from transportation infrastructure: Evidence from the Golden Quadrilateral of India." *Journal of the European Economic Association* 17 (6): 1881-1940.

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

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

[12] Bergquist, Lauren and Michael Dinerstein (2019). "Competition and Entry in Agricultural Markets: Experimental Evidence from Kenya." Unpublished.

[13] Bergquist, Lauren, Marshall Burke, and Edward Miguel (2019). "Sell Low and Buy High: Arbitrage and Local Price Effects in Kenyan Markets." *Quarterly Journal of Economics* 134 (2): 785-842.

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

[15] Berry, Steven, James Levinsohn, and Arield Pakes (1995). "Automobile prices in market equilibrium". *Econometrica* 63 (4): 841-890.

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

[17] Brooks, Wyatt and Kevin Donovan (2019). "Eliminating Uncertainty in Market Access: The Impact of New Bridges in Rural Nicaragua." Unpublished.

[18] Bryan, Gharad and Melanie Morton (2019). "The Aggregate Productivity Effects of Internal Migration: Evidence from Indonesia." *Journal of Political Economy* 127 (5): 2229-2268.

[19] Casaburi, Lorenzo, Rachel Glennerster, Tavneet Suri (2013). "Rural Roads and Intermediated Trade: Regression Discontinuity Evidence from Sierra Leone." Unpublished.

[20] Casaburi, Lorenzo and Tristan Reed (2019). "Interlinked Transactions and Competition: Experimental Evidence from Cocoa Markets." Unpublished.

[21] Chatterjee, Shoumitro (2019). "Market Power and Spatial Competition in Rural India." Unpublished.

[22] Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, DiegoPuga, and Sébastien Roux (2012). "The productivity advantages of large cities: Distinguishing agglomeration from firm selection." *Econometrica* 80 (6), 2543-2594

[23] 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.

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

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

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

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

[28] 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.

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

[30] Eckert, Fabian (2019). "Growing Apart: Tradable Services and the Fragmentation of the US Economy", Working Paper

[31] FAO (2016). *FAOSTAT*. Food and Agriculture Organization of the United Nations, Rome, Italy.

[32] Gertler, Paul, Tadeja Gracner, Marco Gonzalez-Navarro, and Alex Rothenberg (2019). "Road Quality and Local Economic Activity: Evidence from Indonesia's Highways." Unpublished.

[33] 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.

[34] Gollin, Douglas, and Christopher Udry (2019). "Heterogeneity, Measurement Error and Misallocation: Evidence from African Agriculture." Unpublished.

[35] Hausman, JA (1996). "Valuation of new goods under perfect and imperfect competition". In *The Economics of New Goods*. University of Chicago Press: 209-248

[36] 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.

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

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

[39] Khanna, Gaurav (2019). "The Road Oft Taken: The Route to Spatial Development." Unpublished.

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

[41] Morten, Melanie and Jaqueline Oliveira (2018). "The Effects of Roads on Trade and Migration: Evidence from a Planned Capital City." Unpublished.

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

[43] Porteous, Obie (2020). "Trade and Agricultural Technology Adoption: Evidence from Africa." *Journal of Development Economics* 144.

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

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

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

[47] Shamdasani, Yogita (2019). "Rural Road Infrastructure & Agricultural Production: Evidence from India." Unpublished.

[48] Sotelo, Sebastian (2019). "Domestic Trade Frictions and Agriculture." Unpublished.

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

[50] 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.

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

[52] Udry, Christopher and Santosh Anagol (2006). "The Return to Capital in Ghana." *AER Papers and Proceedings* 96 (2): 388-393..

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

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

[55] 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

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

Table 1. Summary statistics on villages

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

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

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

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

Table 2. Calibrating Travel Costs

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Main roads (Market centers to hub towns) (Transport Operator Surveys)			Rural roads (villages to market)					
				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	0.02***								
	(0.00)								
Google maps: hours to destination		1.26***	1.00***						
		(0.03)	(0.03)						
Number of markets	201	201	201						
Number of observations	900	900	893						
Panel B. Costs from villages									
Google maps: kilometers to destination				0.12***			0.09***		
				(0.01)			(0.01)		
Google maps: hours to destination					3.54***	0.72***		2.61***	0.84***
					(0.27)	(0.07)		(0.25)	(0.08)
Number of villages				1127	1033	1036	1133	1133	1027
Number of observations				1127	1033	1036	1133	1133	1027

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

Table 3. Remoteness and farmer characteristics

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

Notes: N = 2,845 farmers in 246 villages. In Column 1, standard deviations are in parentheses. Columns 2 and 3 show regression coefficients from separate regressions of the dependent variable on a measure of remoteness (equations 5 and 6 in the paper). See text for further discussion of these measures. In those columns, standard errors in parentheses, clustered at the village level.
*, **, and *** indicate significance at 10%, 5%, and 1%.
[1]Regressions for production capacity are at village level.

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

	(1)	(2)	(3)
	Mean	(Standardized) coefficient from regression of dependent variable on remoteness measure based on (population-weighted):	
		Distance to hubs	Elasticity-adjusted travel costs to hubs
Panel A. Summary measures of access to input retailers			
Has at least 1 agrovet within 10 km of village	0.75	-0.14***	-0.13***
which sells fertilizer or seeds		(0.01)	(0.01)
Number of agrovets within 10 km of village	7.79	-2.93***	-4.18***
which sells fertilizer or seeds	(8.96)	(0.18)	(0.21)
Distance to nearest agrovet	6.79	3.17***	2.46***
which sells fertilizer or seeds	(15.15)	(0.73)	(0.58)
Distance to the second nearest village with an agrovet	15.52	5.38***	5.93***
which sells fertilizer or seeds	(23.97)	(1.02)	(0.86)
Panel B1. Travel-cost adjusted prices faced by farmers			
Minimum travel-cost adjusted price	24.19	2.33***	2.41***
for 50 kg of Urea (USD)[1]	(4.66)	(0.14)	(0.12)
Decomposition of price between retail price and cost of transportation			
Retail price at the location with the lowest	19.82	1.09***	1.27***
travel-cost adjusted price (USD)	(2.63)	(0.07)	(0.06)
Cost of travel to obtain minimum travel-cost	4.372	1.24***	1.14***
adjusted price (USD)	(4.39)	(0.14)	(0.12)
Panel B2. Travel-cost adjusted prices at the nearest agro-input shop			
Travel-cost adjusted price at the nearest input seller	26.55	2.37***	2.14***
for 50 kg of Urea (USD)[1]	(6.10)	(0.20)	(0.17)
Decomposition of price between retail price and cost of transportation			
Retail price at the nearest input seller (USD)	23.35	1.30***	1.29***
	(3.39)	(0.10)	(0.08)
Cost of travel to the nearest input seller (USD)	3.21	1.07***	0.85***
	(4.58)	(0.17)	(0.14)

Notes: The unit of observation is the village. Data is from the universe of villages in Kilimanjaro and Manyara regions (N = 1,183). Travel costs imputed from transport surveys and Google maps. In Column 1, standard deviations are in parentheses. Columns 2 and 3 show regression coefficients from separate regressions of the dependent variable on a measure of remoteness (equations 5 and 6 in the paper). See text for further discussion of these measures. In those columns, standard errors in parentheses, clustered at the village level.
*, **, and *** indicate significance at 10%, 5%, and 1%.

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

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

	(1)	(2)	(3)
	Mean	(Standardized) coefficient from regression of dependent variable on remoteness measure based on (population-weighted):	
		Distance to hubs	Elasticity-adjusted travel costs to hubs
Panel A. Summary measures of access to output markets			
Has at least 1 maize seller within 10 km of village	0.67	-0.16***	-0.16***
		(0.01)	(0.01)
Number of maize sellers within 10 km of village	1.89	-1.06***	-1.48***
	(2.48)	(0.07)	(0.08)
Distance to nearest output market with maize sellers (km)	8.67	5.65***	4.39***
	(14.17)	(0.71)	(0.43)
Panel B1. Maximum imputed travel-cost adjusted price if farmers were to sell in a local market			
Market survey: maximum travel-cost adjusted price	30.30	-3.08***	-3.05***
immediately after 2017 harvest (USD)[1]	(7.24)	(0.22)	(0.19)
Decomposition of price between retail price and cost of transportation			
Retail price at the location with the highest	39.34	0.80***	0.23**
travel-cost adjusted price (USD)	(3.17)	(0.08)	(0.09)
Cost of travel to obtain the highest travel-cost	9.05	3.88***	3.28***
adjusted price (USD)	(7.06)	(0.21)	(0.18)
Panel B2. Travel-cost adjusted otput price at the nearest maize selling market			
Travel-cost unadjusted 120 kg bag of maize price	20.83	-3.26***	-3.16***
immediately after 2017 harvest (USD)[1]	(8.98)	(0.29)	(0.24)
Decomposition of price between retail price and cost of transportation			
Retail price at the nearest maize selling market (USD)	26.67	-1.37***	-1.77***
	(5.95)	(0.19)	(0.15)
Cost of travel to the nearest maize selling market (USD)	5.840	1.89***	1.39***
	(5.88)	(0.21)	(0.17)
Panel B3. Price available within village by maize-buying intermediaries immediately after last season's harvest			
Farmer surveys: average "going price" in local	25.86	-1.31**	-2.60***
village immediately after previous harvest[2]	(6.24)	(0.52)	(0.48)

Notes: The unit of observation is the village. Data is from the universe of villages in Kilimanjaro and Manyara regions (N = 1,183). Travel costs imputed from transport surveys and Google maps. In Column 1, standard deviations are in parentheses. Columns 2 and 3 show regression coefficients from separate regressions of the dependent variable on a measure of remoteness (equations 5 and 6 in the paper). See text for further discussion of these measures. In those columns, standard errors in parentheses, clustered at the village level.
*, **, and *** indicate significance at 10%, 5%, and 1%.

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

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

Table 6. Remoteness and input market access and adoption

	(1)	(2)	(3)	(4)	(5)
		(Standardized) coefficient from regression of dependent variable on remoteness measure based on (population-weighted):			
	Mean	Distance to hubs		Elasticity-adjusted travel costs to hubs	
		No controls	Controls for soil and farmer characteristics	No controls	Controls for soil and farmer characteristics
Panel A: Input usage					
Used chemical fertilizer in previous long rains	0.39	-0.17***	-0.09***	-0.20***	-0.13***
		(0.03)	(0.03)	(0.03)	(0.03)
Quantity of chemical fertilizer used (kg)	19.84	-13.06***	-6.46***	-14.42***	-9.33***
	(31.63)	(2.15)	(1.74)	(1.91)	(1.88)
Used improved seeds in previous long rains	0.66	-0.07***	-0.05**	-0.11***	-0.10***
		(0.02)	(0.02)	(0.02)	(0.03)
Quantity of improved seeds used (kg)	6.29	-1.30***	-1.21***	-1.09***	-1.03**
	(8.21)	(0.36)	(0.44)	(0.32)	
Panel B. Maize sales					
Sold maize after previous long rains	0.32	-0.09***	-0.06**	-0.07***	-0.04*
		(0.02)	(0.03)	(0.02)	(0.02)
Total quantity sold (kg)	388.1	-97.86***	-112.16**	-5.90	-19.23
	(1142.00)	(35.11)	(48.86)	(39.86)	(47.71)
Sales to agents at home					
Agent visited homestead	0.31	-0.14***	-0.09***	-0.12***	-0.07*
		(0.03)	(0.03)	(0.03)	(0.04)
Sold maize to agent after previous long rains	0.17	-0.07***	-0.04**	-0.05***	-0.02
		(0.02)	(0.02)	(0.01)	(0.02)
Quantity sold to agents (kg)	142	-46.39***	-39.11**	-13.14	4.06
	(433.70)	(13.80)	(18.61)	(12.51)	(19.12)
Sales at market					
Sold maize at market after previous long rains	0.06	-0.03***	-0.03***	-0.02***	-0.02**
		(0.01)	(0.01)	(0.01)	(0.01)
Quantity sold at market (kg)	34.42	-14.61***	-15.26**	-9.53	-11.26
	(197.10)	(5.32)	(7.72)	(5.98)	(7.40)
Panel C. Maize purchases					
Farmer ever buys maize	0.48	0.11***	0.08***	0.11***	0.09***
		(0.02)	(0.02)	(0.02)	(0.02)
Quantity purchased in typical year (kg)	152.3	75.67***	65.05***	90.29***	77.49***
	(315.50)	(15.79)	(17.92)	(16.06)	(14.37)
Net buying					
Farmer buys maize but sells none	0.37	0.11***	0.08***	0.11***	0.08***
		(0.02)	(0.03)	(0.02)	(0.02)
Farmer sells maize and buys none		-0.09***	-0.06***	-0.08***	-0.06***
		(0.02)	(0.02)	(0.01)	(0.02)
Farmer buys and sells maize	0.08	-0.00	0.00	0.01	0.02
		(0.01)	(0.01)	(0.01)	(0.01)
Net buyer (quantity bought > quantity sold)	0.32	-0.09***	-0.06**	0.11***	0.08***
		(0.02)	(0.03)	(0.02)	(0.03)
Net seller (quantity bought < quantity sold)	0.17	-0.07***	-0.04**	-0.07***	-0.05*
		(0.02)	(0.02)	(0.02)	(0.03)

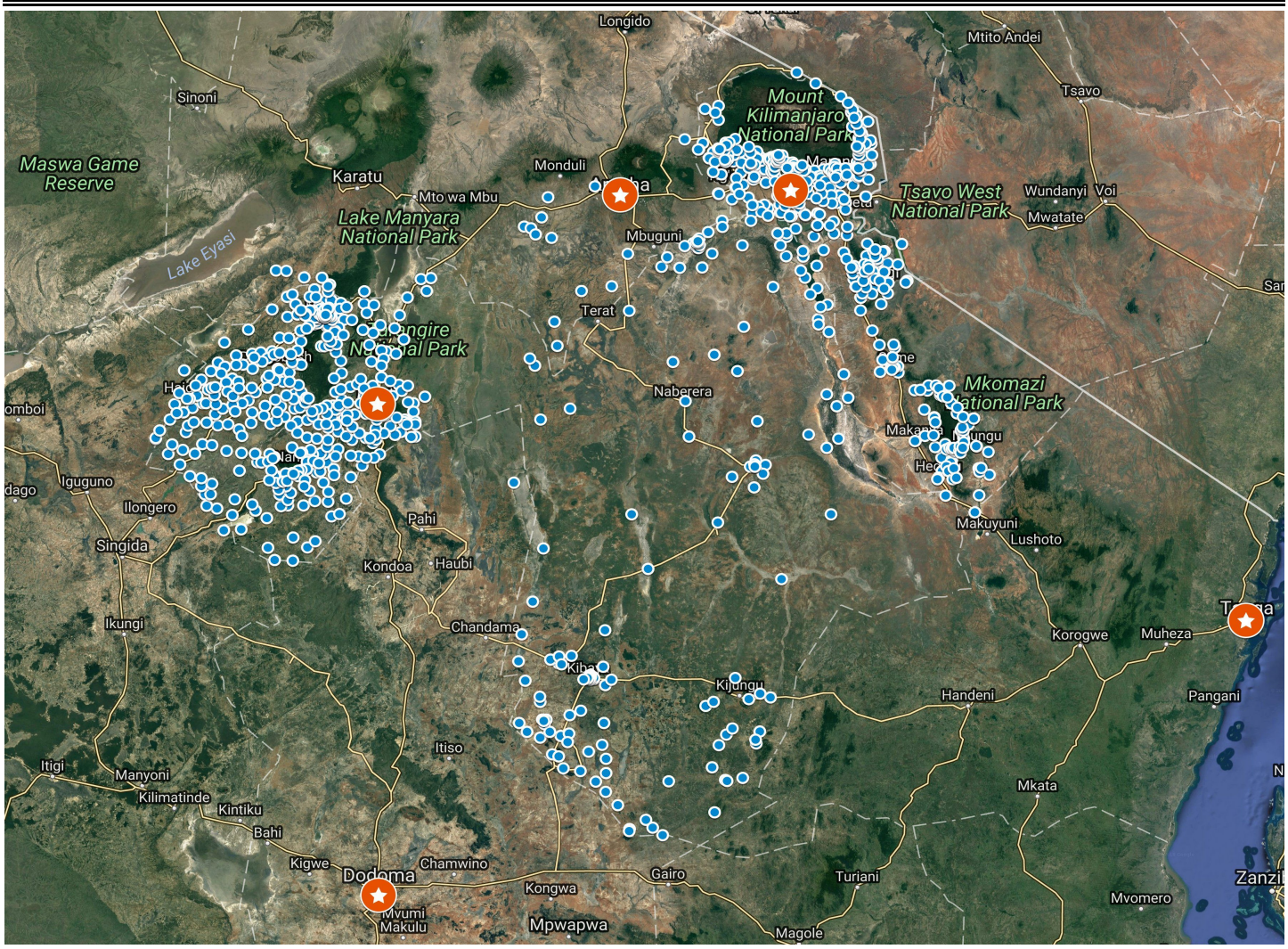
Notes: N = 2,845 farmers in 246 villages. See text for sampling details. Standard deviations are in parentheses in Column 1. Columns 2-5 show regression coefficients from separate regressions of the dependent variable on a measure of remoteness (equations 5 and 6 in the paper). See text for further discussion of these measures. In those columns, standard errors in parentheses, clustered at the village level.
*, **, and *** indicate significance at 10%, 5%, and 1%.

Table 7. Multinomial logit of agrovet choice

	(1)	(2)	(3)
	Agrovet Chosen		AVE/KM
Kilometers to agrovet	-0.171***		2.3%
	(0.009)		
Dummies for agrovet distance bin:			
between (0,5] km		-1.380***	3.7%
		(0.372)	
between (5,10] km		-2.914***	4.0%
		(0.379)	
between (10,15] km		-4.331***	3.9%
		(0.380)	
		-5.367***	3.6%
		(0.400)	
between (20,30] km		-5.875***	2.6%
		(0.378)	
between (30,40] km		-7.602***	2.6%
		(0.449)	
between (40,50] km		-8.685***	2.3%
		(0.495)	
between (50,100] km		-10.625***	1.4%
		(0.560)	
over 100 km		-14.253***	-
		(0.992)	

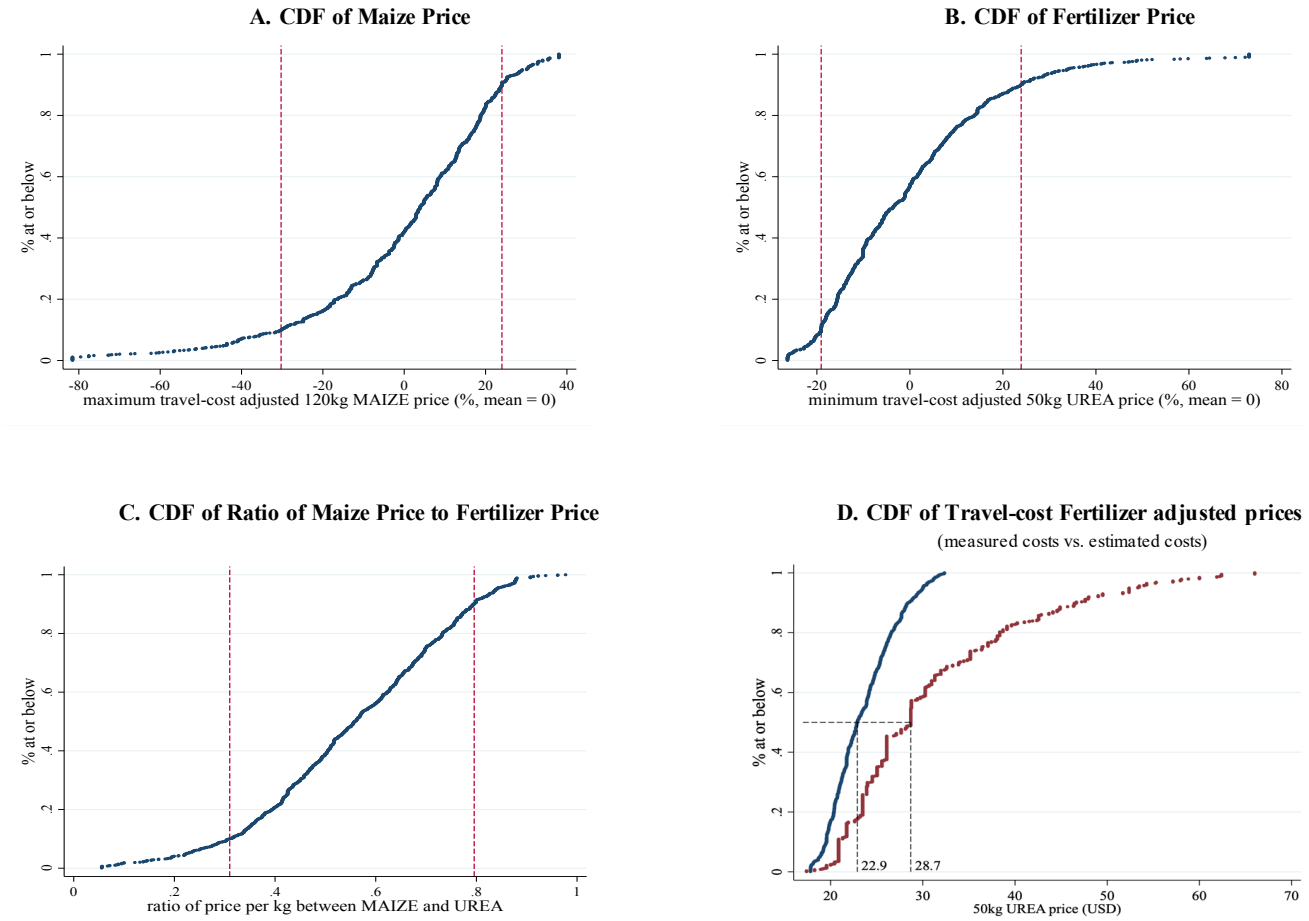
Notes: N = 519 farmers, 119 observed locations. Omitted group is agrovet located in respondent's village. Ad-valorem equivalent per kilometer is calculated at the upper bound of each bin, and assumes that the trade cost compounds each kilometer. Standard errors in parentheses. *, **, and *** indicate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

Figure 1. Map of Survey Region and Villages



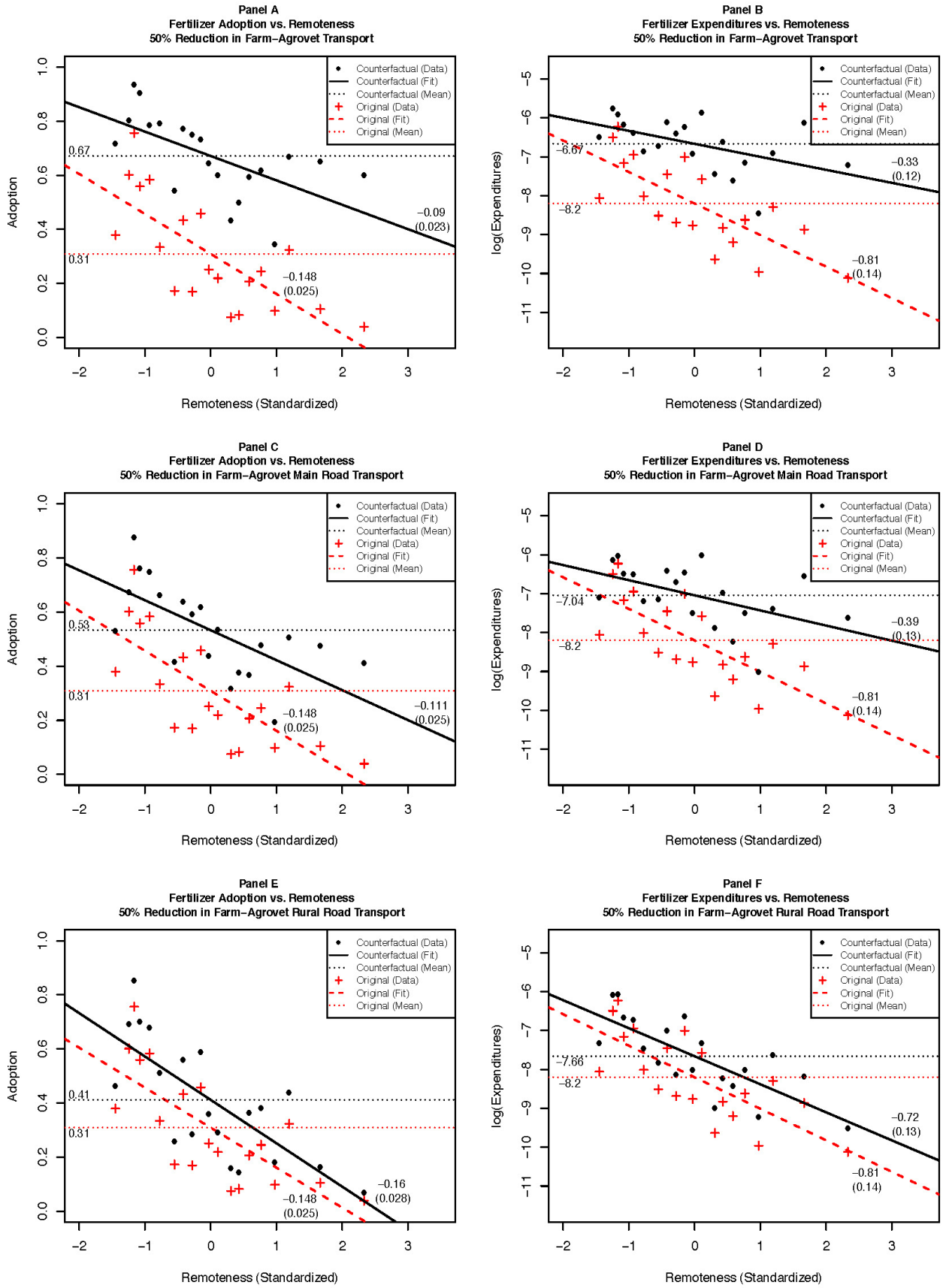
Notes: Blue dots represent all villages in the Kilimanjaro and Manyara Regions. The star signs represent the five major hubs that are used to construct our market access proxies in Section 4.1. They are Moshi, Arusha, Babati, Dodoma, and Tanga.

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



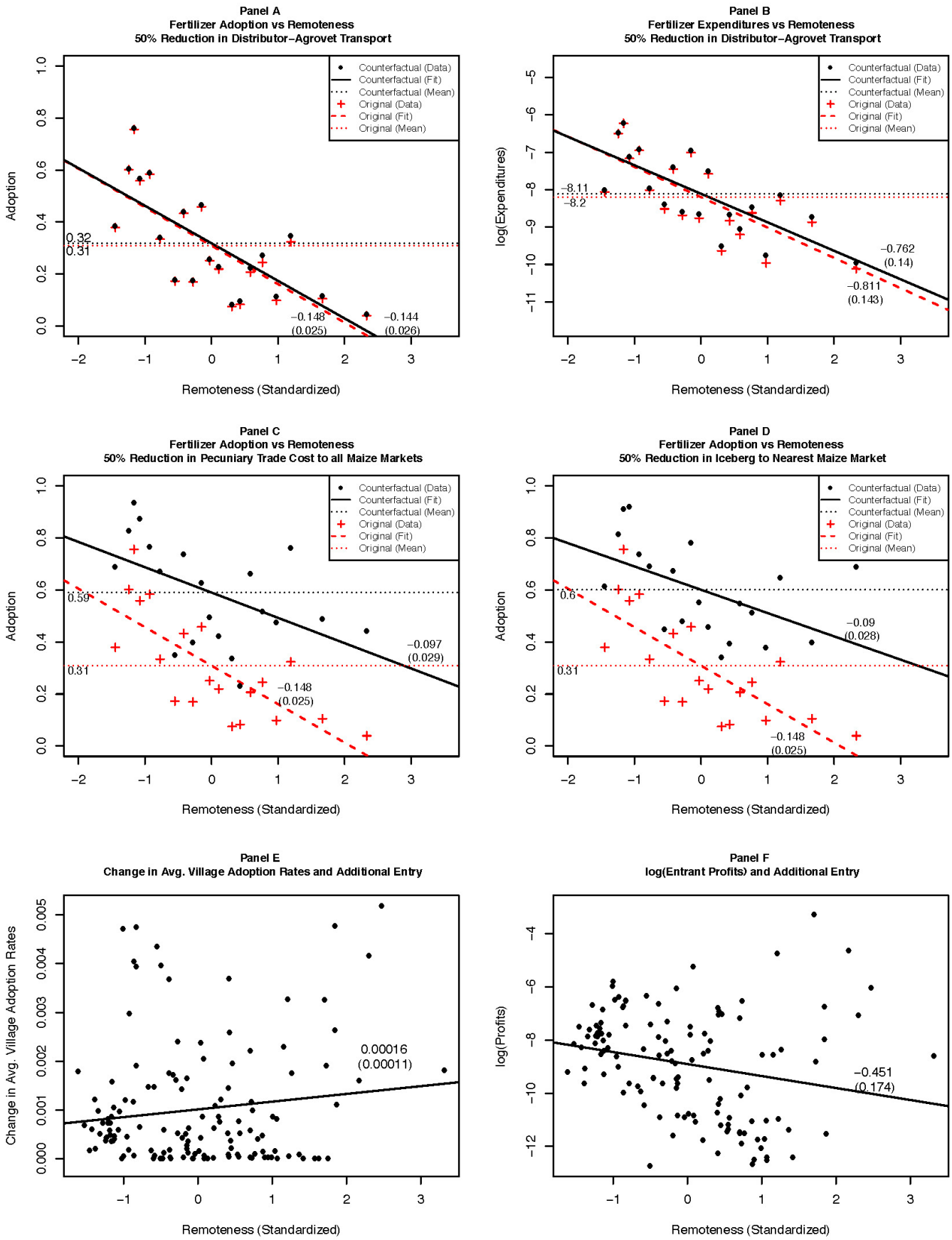
survey at markets and transport cost information collected from interviews with transport operators. In Panels A-C, the vertical dotted lines represent the 10th and 90th percentile. In Panel D, the vertical lines represent the median.

Figure 3. Input Access Counterfactuals



Notes: See text for discussion of counterfactuals.

Figure 4. Distributor, Maize Price, and Entry Counterfactuals



Notes: See text for discussion of counterfactuals.

Appendix A: Remoteness and Market Access

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

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

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

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

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

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

Collecting terms:

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

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

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

Standardizing this equation gives us:

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

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

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

The production function under basic technology is:

$$Y_i = \widetilde{\theta}_{i0} K_i^{\alpha} L_i^{1-\alpha} \tag{20}$$

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:

$$\Pi_{i0} = \max_{L_i} \left\{ p_i \widetilde{\theta}_{i0} K_i^{\alpha} L_i^{1-\alpha} - w_i L_i \right\} \tag{21}$$

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:

$$(1-\alpha) p_i \widetilde{\theta}_{i0} K_i^{\alpha} L_i^{-\alpha} = w_i \tag{22}$$

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

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

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:

$$\begin{aligned} \Pi_{i0} &= \alpha_0 (1-\alpha_0)^{\frac{1-\alpha_0}{\alpha_0}} \widetilde{\theta}_{i0}^{\frac{1}{\alpha_0}} p_i^{\frac{1}{\alpha_0}} w_i^{-\frac{1-\alpha_0}{\alpha_0}} K_i \\ &= \theta_{i0} \pi_{i0} \end{aligned} \tag{24}$$

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

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

$$Y_i = \widetilde{\theta}_{ijv} \left(\theta_i K_i\right)^{\alpha} L_{ijv}^{(1-\alpha)\beta} X_{ijv}^{(1-\alpha)(1-\beta)} \tag{25}$$

The profit maximization problem when using fertilizer is written as:

$$\Pi_{i0} \quad = \quad \max_{L_i, X_{ijv}} p_i \widetilde{\theta}_{ijv} \left(\theta_i K_i\right)^{\alpha} L_{ijv}^{(1-\alpha)\beta} F_{ijv}^{(1-\alpha)(1-\beta)} - w_i L_{ijv} - r_{ijv} F_{ijv} \tag{26}$$

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

$$c_{ijv} M_{ijv} = (1-\alpha) p_i \widetilde{\theta}_{ijv} \left(\theta_i K_i\right)^{\alpha} L_{ijv}^{(1-\alpha)\beta} F_{ijv}^{(1-\alpha)(1-\beta)} = (1-\alpha) R_{ijv} \tag{27}$$

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:

$$\Pi_i = \theta_{ijv} \pi_i r_{ijv}^{-\sigma} \tag{28}$$

where $\sigma \equiv \frac{1-\alpha}{\alpha}(1-\beta)$, $\pi_i = p_i^{\frac{1}{\alpha}} w_i^{-\beta \frac{1-\alpha}{\alpha}} K_i$, and $\theta_{ijv} = \kappa_2 \widetilde{\theta}_{ijv}^{\kappa_1}$.[34] 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.

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

Appendix C: Distributions of Fertilizer Expenditures

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

$$\mathbb{E}\left[rF_i \mid adopt\ at\ j\ in\ v\right] = \mathbb{E}\left[rF_i | adopt\right] \tag{29}$$

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 (2002), 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.

$$\begin{aligned}
\Pr\left(\Pi_{ijv} > \pi\right) &= \Pr\left(\theta_{ijv}\pi_i r_{ijv}^{-\sigma} > \pi\right) && (30)\\
&= \Pr\left(\theta_{ijv} > \frac{\pi}{\pi_i} r_{ijv}^{\sigma}\right) && (31)\\
&= 1 - \exp\left(-T_{jv}\pi_i^{\varepsilon} r_{ijv}^{\varepsilon\sigma}\pi^{-\varepsilon}\right) && (32)
\end{aligned}$$

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

$$\Pr\left(\Pi_{ijv} > \pi\right) = 1 - \exp\left(-T_{jv}\gamma_{ijv}\pi^{-\varepsilon}\right) \tag{33}$$

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

$$\Pr\left(\Pi_{i0} > \pi\right) = 1 - \exp\left(-\Phi_{i0}\pi^{-\varepsilon}\right) \tag{34}$$

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

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

$$\begin{aligned}
\Pr\left(\Pi_i^{max} > \pi\right) &= \Pr\left(\Pi_{ijv} > \pi \text{ for any } jv\right) && (35)\\
&= 1 - \Pr\left(\Pi_{ijv} < \pi\ \forall\ jv\right) && (36)
\end{aligned}$$

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

as:

$$\begin{aligned}
\Pr\left(\Pi_i^{max} > \pi\right) &= 1 - \Pr\left(\Pi_{ijv} < \pi \ \forall \ jv\right) && (37)\\
&= 1 - \prod_{v' \in \mathcal{V}} \prod_{j \in \mathcal{J}_v} \Pr\left(\Pi_{ijv} < \pi\right) && (38)\\
&= 1 - \prod_{v' \in \mathcal{V}} \prod_{j \in \mathcal{J}_v} \exp\left(-\pi^{-\varepsilon}\right) && (39)
\end{aligned}$$

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

$$\Pr\left(\Pi_i^{max} > \pi\right) = 1 - \exp\left(-\tilde{\Phi}_i \pi^{-\varepsilon}\right) \quad (40)$$

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

$$G_i^{max}\left(\pi\right) = \Pr\left(\Pi_i^{max} < \pi\right) = \exp\left(-\tilde{\Phi}_i \pi^{-\varepsilon}\right) \quad (41)$$

with pdf:

$$g_i^{max}\left(\pi\right) = \varepsilon \tilde{\Phi}_i \pi^{-\varepsilon-1} \exp\left(-\tilde{\Phi}_i \pi^{-\varepsilon}\right) \quad (42)$$

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

$$\begin{aligned}
\Pr\left(\Pi_i > \pi\right) &= \Pr\left(\Pi_{ijv} > \pi \text{ for any } jv \cup \Pi_{i0} > \pi\right) && (43)\\
&= 1 - \Pr\left(\Pi_{ijv} < \pi \ \forall \ jv \ \cap \ \Pi_{i0} < \pi\right) && (44)
\end{aligned}$$

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

$$\begin{aligned}
\Pr\left(\Pi_i > \pi\right) &= 1 - \Pr\left(\Pi_{ijv} < \pi \ \forall \ jv \ \cap \ \Pi_{i0} < \pi\right) && (45)\\
&= 1 - \Pr\left(\Pi_{i0} < \pi\right) \prod_{v' \in \mathcal{V}} \prod_{j \in \mathcal{J}_v} \Pr\left(\Pi_{ijv} < \pi\right) && (46)\\
&= 1 - \exp\left(-T_{i0}\gamma_{i0}\pi^{-\varepsilon}\right) \prod_{v' \in \mathcal{V}} \prod_{j \in \mathcal{J}_v} \exp\left(-T_{jv}\gamma_{ijv}\pi^{-\varepsilon}\right) && (47)
\end{aligned}$$

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

$$\Pr\left(\Pi_i > \pi\right) = 1 - \exp\left(-\left(\tilde{\Phi}_{i0} + \tilde{\Phi}_i\right)\pi^{-\varepsilon}\right) \quad (48)$$

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

$$G_i\left(\pi\right) = \exp\left(-\left(\tilde{\Phi}_{i0} + \tilde{\Phi}_i\right)\pi^{-\varepsilon}\right) \quad (49)$$

with pdf:

$$g_i(\pi) = \varepsilon\left(\tilde{\Phi}_{i0}+\tilde{\Phi}_i\right)\pi^{-\varepsilon-1}\exp\left(-\left(\tilde{\Phi}_{i0}+\tilde{\Phi}_i\right)\pi^{-\varepsilon}\right) \tag{50}$$

Profits conditional on adoption

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

$$\Pr\left(\Pi_i^{max}<\pi\middle|adopt\right) = \frac{\Pr\left(\Pi_i^{max}<\pi\cap\Pi_i^{max}>\Pi_{i0}\right)}{\Pr\left(\Pi_i^{max}>\Pi_{i0}\right)} \tag{51}$$

This can be re-written as:

$$\begin{aligned}\Pr\left(\Pi_i^{max}<\pi\middle|adopt\right) &= \frac{1}{\Pr\left(\Pi_i^{max}>\Pi_{i0}\right)}\int_0^{\pi}\Pr\left(s>\Pi_{i0}\right)g_i^{max}(s)ds\\ &= \frac{1}{\Pr\left(\Pi_i^{max}>\Pi_{i0}\right)}\int_0^{\pi}\exp\left(-\tilde{\Phi}_{i0}s^{-\varepsilon}\right)\varepsilon\tilde{\Phi}_i s^{-\varepsilon-1}\exp\left(-\tilde{\Phi}_i s^{-\varepsilon}\right)ds\\ &= \frac{1}{\Pr\left(\Pi_i^{max}>\Pi_{i0}\right)}\int_0^{\pi}\varepsilon\tilde{\Phi}_i s^{-\varepsilon-1}\exp\left(-\left(\tilde{\Phi}_{i0}+\tilde{\Phi}_i\right)s^{-\varepsilon}\right)ds \end{aligned}\tag{52}$$

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

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

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

$$\Pr\left(\Pi_i^{max}<\pi\middle|adopt\right) = \int_0^{\pi}\varepsilon\left(\tilde{\Phi}_{i0}+\tilde{\Phi}_i\right)s^{-\varepsilon-1}\exp\left(-\left(\tilde{\Phi}_{i0}+\tilde{\Phi}_i\right)s^{-\varepsilon}\right)ds \tag{53}$$

$$= \Pr\left(\Pi_i<\pi\right) \tag{54}$$

Profits conditional on adoption from j

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

$$\Pr\left(\Pi_{ijv}<\pi\middle|adopt\ from\ j\ in\ v\right) = \frac{\Pr\left(\Pi_{ijl}<\pi\cap\Pi_{ijv}>\Pi_{ij'l}\forall(j',l)\cap\Pi_{ijv}>\Pi_{i0}\right)}{\Pr\left(\Pi_{ijv}>\Pi_{ij'l}\forall(j',l)\cap\Pi_{ijv}>\Pi_{i0}\right)} \tag{55}$$

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

$$\Pr\left(\Pi_{ijv}<\pi\middle|adopt\ from\ j\ in\ v\right) = \frac{1}{\lambda_{ijv}}\int_0^{\pi}Pr\left(s>\Pi_{ij'l}\forall(j',l)\cap s>\Pi_{i0}\right)g_{ijv}(s)ds \tag{56}$$

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

$$\begin{aligned} \Pr\left(s > \Pi_{ij'l} \forall (j', l) \cap s > \Pi_{i0}\right) &= \exp\left(-\tilde{\Phi}_{i0} s^{-\varepsilon}\right) \exp\left(-\widetilde{\tilde{\Phi}}_{ijv} s^{-\varepsilon}\right) && (57) \\ &= \exp\left(-\left(\tilde{\Phi}_{i0} + \widetilde{\tilde{\Phi}}_{ijv}\right) s^{-\varepsilon}\right) && (58) \end{aligned}$$

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

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

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

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

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

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

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

Appendix D: Production Function Estimation with and without Fertilizer

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

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

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

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

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

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

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

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

Appendix E: Mark-ups

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

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

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

$$\mathbb{E}\left[v_{jv}\right] = (1-\beta)\frac{1-\alpha}{\alpha}\sum_i \lambda_{ijv}\mathbb{E}\left[\Pi_i\right] \tag{59}$$

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

$$\frac{d\mathbb{E}\left[v_{jv}\right]}{dr_{jv}}\frac{r_{jv}}{\mathbb{E}\left[v_{jv}\right]} = \sum s_{ijv}\left(\frac{d\lambda_{ijv}}{dr_{jv}}\frac{r_{jv}}{\lambda_{ijv}} + \frac{d\mathbb{E}\left[\Pi_i\right]}{dr_{jv}}\frac{r_{jv}}{\mathbb{E}\left[\Pi_i\right]}\right) \tag{60}$$

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 κ 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:

$$\frac{d\mathbb{E}\left[v_{jv}\right]}{dr_{jv}}\frac{r_{jv}}{\mathbb{E}\left[v_{jv}\right]} = -\varepsilon_a\sum_i s_{ijv}\left((1-\lambda_{ijv}) + \frac{1}{\varepsilon}\lambda_{ijv}\right) \tag{61}$$

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

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

Appendix F: External Validity

F1. Price Dispersion

To compare price dispersion in our study region to other parts of Africa, we bring in evidence from five secondary datasets of prices in 1,512 markets in 56 African countries[35] We compare this to

[35]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

a small dataset we assembled between March and April 2016 with 251 retailers of various sorts (shops, agrovets, and maize traders) in 82 markets in the Kilimanjaro region.[36] To quantify price dispersion, we decompose variation in spatial prices by running the following regression:

$$\log(p_{mcjt}) = \gamma_c + \gamma_j + \gamma_t + \epsilon_{mjt} \tag{63}$$

where p_{mcjt} (log) prices in market m for product j at time t in country c, and the γ terms are country, product, and time fixed effects. We report the standard deviation of the resulting residual in Appendix Table A9. In the secondary datasets, the standard deviation is 0.45 for all products, 0.34 for maize, and 0.12 for fertilizer; in our Tanzania data, the figures are 0.22, 0.14, and 0.09. While price dispersion is lower in our data (perhaps because of reduced measurement error, or that prices vary less within the geographic concentrated area of Kilimanjaro), this simple exercise sugggests substantial price dispersion in Northern Tanzania.

We also follow the literature,[37] to run dyadic regressions to look at price gaps, as follows:

$$\log(\mid p_{mjt} - p_{m'jt} \mid) = \theta \log(c_{mm'}) + \gamma_m + \gamma_{m'} + \gamma_j + \epsilon_{mm'jt} \tag{64}$$

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.[38]

Results are presented in Appendix Table A10. For each dyad, we regress the absolute difference in log prices on two measures of distance: (1) kilometers between locations in Columns 1, 4, and 7, and (2) driving time between locations in Columns 2, 5, and 8 (both calculated via Google Maps API), and 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. We can also use this data to provide some descriptive evidence on road upgrading. We conjecture that price gaps should respond to the time it takes to travel from point to point, and not the geographic distance (since the time and other costs of traveling to sell items should be what is important). To examine this, we regress price gaps on both distance and duration in Columns 3, 6, and 9. Consistent with priors, we find that duration is significant, whereas distance is not – which suggests that improving road

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.

[36] 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.

[37] 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).

[38] These regressions are motivated by an assumption of free entry where an arbitrageur will enter if $\mid (p_m - p_{m'}) \mid \geq c_{mm'}$.

quality would reduce these gaps.

F2. Fertilizer adoption

Finally, we use data assembled data from World Bank LSMS-ISA household panel surveys to study how remoteness affects fertilizer adoption in other African countries.[39] Using both measures of remoteness available in the dataset (distance to the main market, and distance to a population center), we find a negative association between remoteness and technology adoption.

[39]The countries included here are Ethiopia, Niger, Nigeria, Malawi, Tanzania, and Uganda.

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	0.04*		0.05***	
	(0.02)		(0.01)	
Google maps: hours to destination		1.28*		1.83***
		(0.68)		(0.26)
Number of villages	73	73	119	119
Number of observations	341	341	988	988

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

Web Appendix Table A3. Locations of Input and Output Distributors

Panel A. Locations of Agro-Input Distributors

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

Panel B. Locations of Output Distributors

B1. 2017 Maize Store Census

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

B2. 2016 Maize Store Census

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

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

Web Appendix Table A4. Summary Statistics of Market Access Proxies

	(1) Remoteness measured by distance
Remoteness measured by elasticity-adjusted travel costs to hubs	0.84*** (0.02)
Dependent variable mean before standardization	304.19
Dependent variable sd before standardization	31.56
Independent variable mean before standardization	-0.11
Independent variable sd before standardization	0.04
Observations	1,135

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

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

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

Notes: In Column 1, standard deviations are in parentheses. Columns 2 and 3 show regression coefficients from separate regressions of the dependent variable on a measure of remoteness (equations 5 and 6 in the paper). See text for further discussion of these measures. Regressions in Panel B includes type and brand fixed effects.

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

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

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

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

Notes: Data is from the universe of villages in Kilimanjaro and Manyara region (N = 1,183). The unit of observation is the village. Travel costs imputed from transport surveys and Google maps. In Column 1, standard deviations are in parentheses. Columns 2 and 3 show regression coefficients from separate regressions of the dependent variable on a measure of remoteness (equations 5 and 6 in the paper). See text for further discussion of these measures.

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

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

Web Appendix Table A7

	(1)	(2)	(3)	(4)
	Dependent variable: log(Eta) (from Calibration)			
log(Price)	-5.566***	-7.516**	-4.731***	-8.121**
	(0.990)	(3.729)	(0.960)	(3.330)
log(Experience)			0.938***	0.870***
			(0.154)	(0.171)
	OLS	IV	OLS	IV
R-squared	0.37	0.35	0.43	0.4
Cragg-Donald Wald F statistic		33.93		32.34
Wu-Hausman		0.35		1.22
Observations	374	242	374	242

Notes: District fixed effects used in all regressions. Instrumental variable is 2011 population in the market catchment area from the Tanzanian census. ***, **, and * indicate significance at 1%, 5%, and 10%.

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

	(1)	(2)	(3)
	Dependent variable: log(Harvest/Acres)		
log(Labor/Acres)	0.42***	0.43***	0.43***
	(0.04)	(0.04)	(0.04)
log(Labor/Acres) x Used Fertilizer?	0.12*	0.12*	0.15*
	(0.08)	(0.07)	(0.08)
Used Fertilizer?	(0.33)	(0.33)	
	(0.30)	(0.30)	
District-Year fixed effects		X	
District-Year-Fertilizer Use fixed effects			X
Plot fixed effects	X	X	X
Observations	3,395	3,395	3,395
Plots	2,554	2,554	2,554

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

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

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

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

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

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

Web Appendix Table A10. Dyadic price dispersion

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Dependent variable: Absolute log price difference								
Panel A. Secondary									
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									
Log (distance)	0.01***		-0.030	0.03***		-0.10**	0.003*		0.007
	(0.003)		(0.020)	(0.011)		(0.050)	(0.002)		(0.017)
Log (travel time)		0.01***	0.04*		0.04***	0.16**		0.004	-0.004
		(0.004)	(0.025)		(0.016)	(0.069)		(0.002)	(0.019)
Products	All	All	All	Maize	Maize	Maize	Fertilizer	Fertilizer	Fertilizer
Dependent variable mean	0.16	0.16	0.16	0.21	0.21	0.21	0.13	0.13	0.13
Dependent variable sd	0.14	0.14	0.14	0.18	0.18	0.18	0.10	0.10	0.10
Observations	22,386	22,376	22,376	6,873	6,873	6,873	15,064	15,056	15,056
Number of locations	82	82	82	65	65	65	60	60	60

Notes: Regressions include product, month and year fixed effects. All regressions are within country. Travel time and distances calculated from Google maps. See Web Appendix Table A3 and text for discussion of datasets.

Two-way clustered standard errors in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10%.

Web Appendix Table A11. Adoption in LSMS-ISA surveys

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

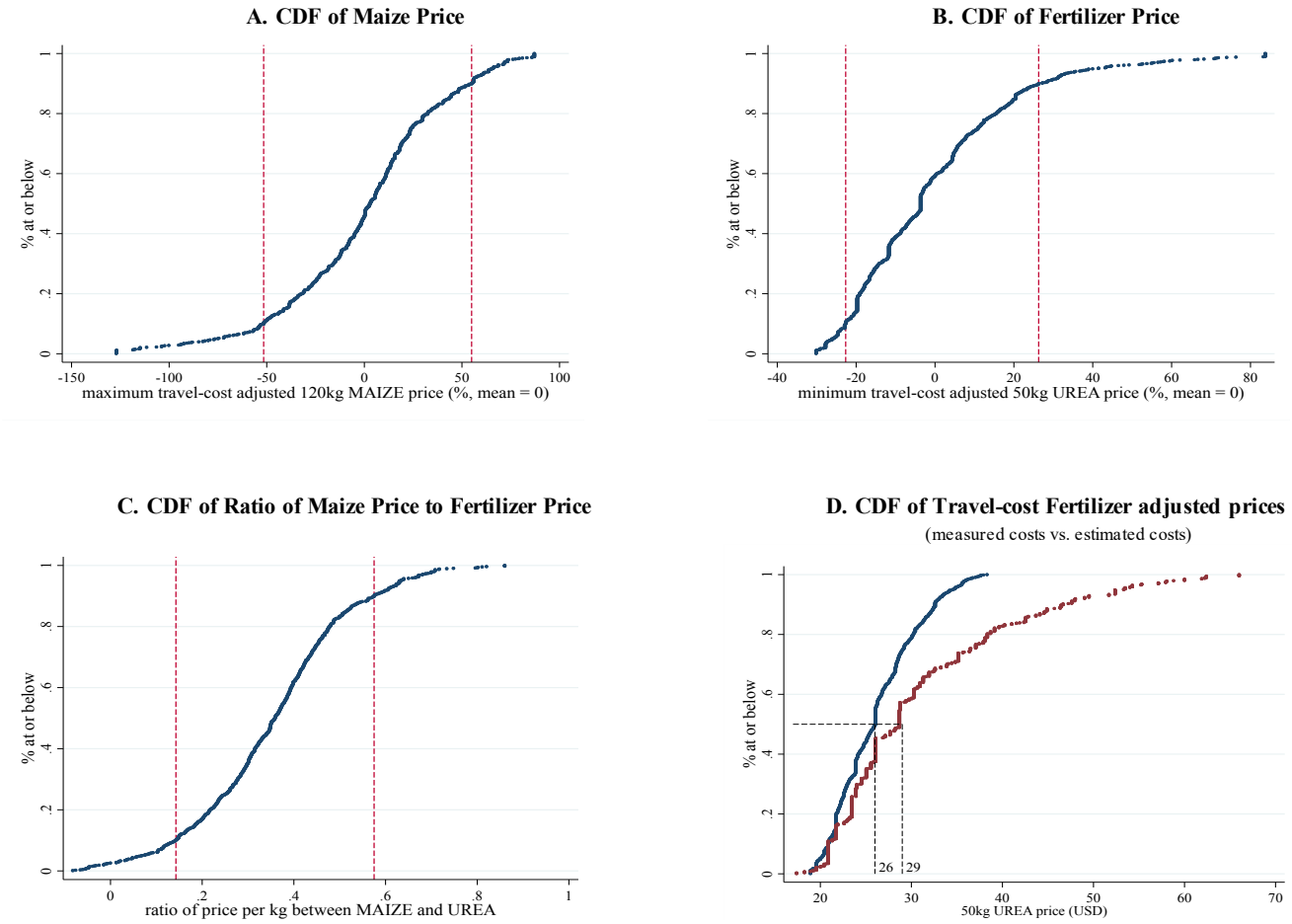
Notes: Regressions include World Bank LSMS-ISA household panel surveys in Ethiopia, Niger, Nigeria, Malawi, Tanzania, and Uganda. Standard errors clustered at the enumeration area level are in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10%.

Web Appendix Table A12. Road density in East Africa

	(1) Road Density (km per '00 sq km area)	(2) Percentage Roads Paved
Burundi	44.28	12.17%
Democratic Republic of Congo	6.54	1.82%
Djibouti	13.21	40.00%
Eritrea	3.41	21.80%
Ethiopia	10.00	13.00%
Kenya	27.82	8.93%
Madagascar	11.18	11.60%
Malawi	13.04	26.37%
Mozambique	3.88	23.70%
Rwanda	17.84	25.68%
Somalia	3.47	11.80%
South Sudan	1.13	2.74%
Tanzania	9.13	8.20%
Uganda	8.52	20.72%
Zambia	12.15	22.00%
Zimbabwe	24.90	19.00%
Sub-Saharan Africa Average	13.70	22.63%

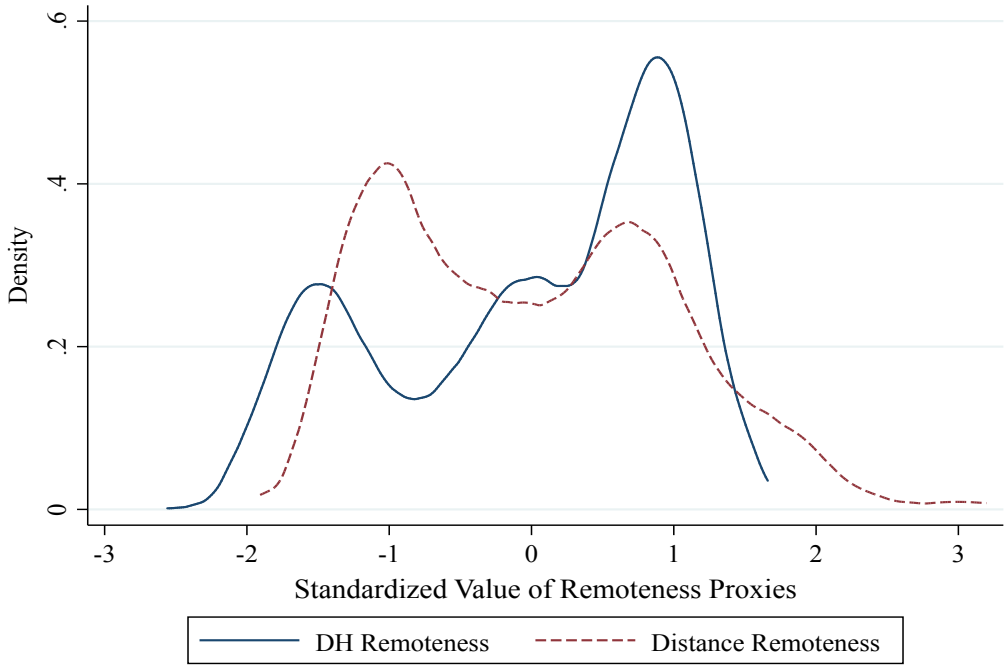
Notes: Data compiled from various World Bank and AfDB reports. Statistics correspond to years ranging between 2010 and 2016; DRC statistics are from 2001.
We include all countries classified as Eastern African as per the United Nations Statistics Division scheme of geographic regions, except the island nations of Comoros, Mauritius and Seychelles, and the French Overseas Territories of Réunion and Mayotte. We also exclude Sudan because there is no data available for after it split from South Sudan.

Web Appendix Figure A1. CDF of travel-cost adjusted prices at the nearest locations



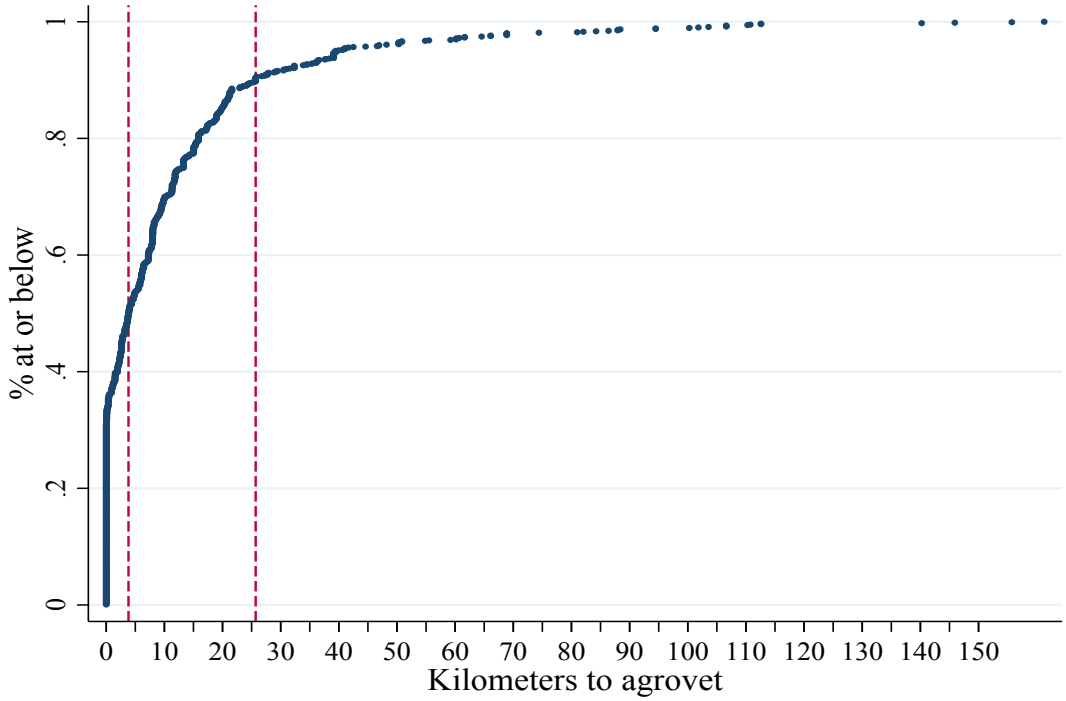
Notes: Each observation represents a village. Travel-cost adjusted prices are calculated through observed prices from an agrovet survey, a maize price survey at markets and transport cost information collected from interviews with transport operators. Vertical dotted lines represent a 10 percentile and a 90 percentile of the distribution.

Web Appendix Figure A2. The Distribution of Remoteness Proxies



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

Web Appendix Figure A3. CDF of Distance Farmers Travel to Purchase Inputs



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