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

Shilpa Aggarwal Brian Giera Dahyeon Jeong
Jonathan Robinson Alan Spearot

October 17, 2022

Abstract

We collect data on prices, travel costs and farmer decisions to quantify market access for chemical fertilizer and its impact on agricultural productivity in 1,180 villages in Northern Tanzania. Villages at the bottom of the travel cost-adjusted input price distribution face 40-55% less favorable prices than those at the top. A standard deviation increase in village-level remoteness is associated with 20-25% lower input adoption. A spatial model of input adoption conservatively estimates that total trade costs are 4X pecuniary travel costs. Counterfactuals suggest that halving travel costs would more than double adoption and reduce the adoption-remoteness gradient by 63%.

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

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

*Aggarwal: Indian School of Business and J-PAL (shilpa_aggarwal@isb.edu); Giera: Amazon Research (bgiera14@gmail.com); Jeong: World Bank (dahyeonjeong@worldbank.org); Robinson: UCSC, BREAD, CEGA, J-PAL and NBER (jmrtwo@ucsc.edu); Spearot: UCSC (aspearot@ucsc.edu). 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 3 anonymous referees, Lauren Falcao Bergquist, Shoumitro Chatterjee, Denise Horner, Alain de Janvry, Betty Sadoulet and many seminar participants 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). The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

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 (World Bank 2007; World Bank 2017).[1] However, while remoteness no doubt limits input market access and, by extension, input adoption and agricultural productivity, there is little research to quantify its effect.

In this paper, we rigorously document input market access for farmers in 1,180 villages, essentially the universe of villages in two regions – Kilimanjaro and Manyara – of Northern Tanzania, which together comprise 6 percent of the land area and population of the country.[2] We map the entire supply chain for fertilizer in these regions using (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) surveys with transportation operators which measure road quality, travel times, and travel costs; and (4) driving times and distances from Google Maps API. The data we collect enables us to provide new evidence on variation in the availability of inputs at the micro level, and the input choices of farmers relative to locations of retailers in spatial equilibrium.

We make three main contributions. First, we precisely document spatial dispersion in input prices, inclusive of travel 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 input prices, where the price difference between the 90th and the 10th percentile of delivered

[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 (Foster and Briceño-Garmendia, 2010). For comparison, the US has 679 kilometers per 1000 square kilometers, with nearly two thirds paved.

[2] These regions in fact have 1,183 villages in all, but 3 villages are not included in our analysis due to errors in GPS information.

input prices is equivalent to about 50% of the mean. Second, we conduct a reduced-form investigation of the correlation between input usage and remoteness, 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. These effect sizes are meaningful: input usage in the most remote villages is only a third of that in the least remote villages.

Third, we develop a quantitative spatial model of fertilizer adoption and pricing, in which the decision to adopt fertilizer is based on local output conditions, 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 the pricing of inputs as determined by the spatial distribution of demand and competition. The structure of the model (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. The solution to the retailer choice problem is facilitated by the novel data collected at the farmer-level on precisely where a farmer sourced inputs, and critically, the choice set of available agrovets at which she could have purchased inputs. Further, the data allows us to calibrate local factors that may affect adoption (eg. output conditions), as well as the implied trade costs incurred while sourcing from each agrovet location. Using this new retailer choice data, we derive a structural multinomial logit specification that estimates the implied iceberg trade costs to each location as a function of distance and road-type. The results suggest that farmer-level trade costs are extremely large: we estimate local iceberg costs that are approximately 72% when buying from the closest agrovet, which is approximately 5 times our accounting estimate of pecuniary costs.[3]

[3]This is higher than Allen (2014), where 50% of trade costs are non-pecuniary.

Using these estimates, we build a market-clearing condition for fertilizer for each agrovet, which is a function of expected 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.[4]

We use the estimated parameters from the model to simulate market access counterfactuals. Exact log changes in adoption are characterized by the log change in market access, and an interaction between the percent change in market access and baseline adoption levels. 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 more than doubles adoption relative to baseline, and also reduces the remoteness gradient by 63%. We also leverage our extensive transport surveys to assess counterfactuals in which transport improvements are targeted toward main roads and rural roads separately. While both increase adoption, it is only the improvement of main roads which reduces the remoteness gradient through market access; surprisingly, improving only rural roads has a smaller effect on both adoption and the gradient, and does not disproportionately increase market access in rural areas. The reason for this is that there are not many retailers located in remote areas, and so farmers in these areas typically must travel on main roads to expand their choice set. We also study hypothetical entry counterfactuals, and find that entry conditions are worse in more remote markets, but improve disproportionately after the 50% reduction in trade costs. Finally, reducing distribution transport costs from wholesalers has a modest impact on adoption, but disproportionately improves adoption in rural areas, and exclusively via improved 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

[4] This is similar in spirit to Redding and Venables (2004); Redding and Sturm (2008); Head and Mayer (2011); and Donaldson and Hornbeck (2016).

behind the rest of the developing world in agricultural technology adoption. Many studies find evidence of large *yield* increases due to using improved inputs, 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; Udry and Anagol 2006). Our results document how access to inputs will erode this profitability, especially in remote areas. This result 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. We also differentiate results by focusing on heterogeneity in market access and input costs, rather than on heterogeneity in returns.[5]

Our paper is also related to a large literature about the effect of transportation infrastructure improvements on development outcomes and spatial economic activity,[6] which includes an expanded set of outcomes such as consumption, investments, migration, and occupational choice. In our paper, we focus narrowly on the specific effect of trade costs on input market access (i.e. trade costs and the presence of intermediaries and the prices they charge) in isolation, without changing other margins.[7]

We contribute to a voluminous trade literature which attributes spatial price differentials to three primary components: marginal trade costs,[8] spatially varying mark-ups,[9] and the organization of intermediaries.[10] Our work is also related to the literature that quantifies the extent of trade costs in Africa, though most existing work focuses on trade between larger cities and markets, typically using trucks or trains (Atkin and Donaldson 2015; Porteous 2020; Porteous 2019) whereas our paper focuses on last-mile costs to farmers (which are

[5]Related work in Minten, Koru and Stifel (2013) also focuses on remoteness and profitability, documenting significant farmer-to-retailer transaction costs in northern Ethiopia.

[6]A partial listing of papers includes Aggarwal (2018), Aggarwal (2021), Brooks and Donovan (2020), and Asher and Novosad (2020). See Donaldson (2015) for a review.

[7]Gebresilasse (2021) evaluates the joint role of road improvements and extension services in increasing productivity in Ethiopia.

[8]See Donaldson (2018); Eaton and Kortum (2002); and Sotelo (2020)

[9]See (Atkin and Donaldson, 2015) and (Asturias, García-Santana and Ramos, 2019)

[10]See, for example, Allen and Atkin (2016), Dhingra, Tenreyro et al. (2021), Bergquist and Dinerstein (2020), Casaburi and Reed (2019) and Chatterjee (2020)).

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 (2020), while our estimated ad-valorem travel costs to the typical nearest retailer is 72% over just 13.5 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.

2 Study region and context

This study took place in the Kilimanjaro and Manyara regions 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). The quality of roads in Kilimanjaro and Manyara is objectively poor, although neither is an outlier for Tanzania: 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.[11] In our study, we differentiate between larger paved roads, which we refer to as "trunk roads," and smaller, mostly unpaved feeder or rural roads.[12]

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, and specific to maize farming.

As in much of Sub-Saharan Africa, production of fertilizer is virtually nil and almost all

[11] Information compiled from various resources.

[12] The Tanzania Roads Act of 2007 defines a trunk road as (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.

of what is used is imported at Dar es Salaam, and then transported throughout the country over surface roads. In these respects, the study area is similar to other countries in East Africa that predominantly import fertilizer, such as Kenya, and perhaps a bit better than landlocked countries such as Uganda, that can receive fertilizer only after it has traversed the distance between a neighboring coastal nation's port and their shared border, and then traveled further inland to reach farmers.[13] Fertilizer is purchased at retail shops known as "agrovets," which sell various farming inputs and, in some cases, veterinary supplies. Urea is the most widely used 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.[14] 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.

3 Data and summary statistics

We have four main sources of data: an agrovet census and detailed surveys, farmer surveys, a village census, and transport surveys. We also collected detailed information on maize markets and prices, and results related to output market acccess and sales are presented and discussed in Online Appendix D. All were collected from January 2016 to December 2017 in Kilimanjaro and February to May 2018 in Manyara.

[13]Online Appendix C compares our study sample with other countries in Africa

[14]While fertilizer choice should ideally be made based on a soil test, these tests remain out of reach for most farmers in Tanzania. In the absence of this information at the plot-level, urea is the most optimal fertilizer choice for farmers as some estimates suggest that Tanzanian land is almost universally deficient in Nitrogen. Harou et al. (2022) corroborate this using data from the Dodoma district, where they found 31% of the plots to be deficient in phosphorous and 100% to be deficient in Nitrogen. Country-level maps of soil nutrient availability can be found here: http://vitalsigns.org/soil-nutrient-maps.

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, and characteristics about the business and its owner. These surveys identify the complete choice set of retailers available to farmers for purchasing chemical fertilizer, which is universally unavailable in other datasets that cover this region.

3.2 Farmer surveys

We conducted farmer surveys in 246 randomly selected villages in three waves.[15] The surveys included questions on input usage and prices, transport costs, and importantly, agrovet choice and the number of trips taken to agrovets. While some 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,[16] while in rounds 2-3 households were pre-identified from a listing exercise conducted with village leaders. In Wave 1, we sampled only

[15]The first wave included 115 villages in Kilimanjaro in early 2016, the second wave included 97 villages in Kilimanjaro in 2017, and the third wave included 50 villages in Manyara in 2018. This survey also served as the sampling frame for Jeong (2020).

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

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.[17]

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,[18] from which we calculated driving times and distances using the Google API. Second, we conducted surveys of transportation operators in every village in the two regions, 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.[19]

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. We also use this data to validate the Google API measures used for calculating travel costs.

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

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

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

3.4 Summary statistics on villages

A map of Kilimanjaro and Manyara is shown in Web Appendix Figure A1. Summary statistics on villages are provided in Table 1. In Panel A, 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 nearest 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.[20] 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 Empirical Results

In this section, we examine dispersion in input prices, and document the relationship between remoteness and prices, adoption, yields, and other related outcomes. The goals of this section are three-fold. First, we use our agrovet and village censuses to document the dispersion in fertilizer access and transport cost adjusted prices. Second, we evaluate how the remoteness of a village - defined by proximity to major hubs - correlates with the access to retailers at the village-level, and the best prices available to a farmer, and adoption of inputs.

4.1 Travel cost-adjusted price dispersion

From their village, we assume that farmers are free to travel anywhere to buy inputs, 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

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

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 on local and trunk roads using our transport operator surveys, and information collected by enumerators during their own travel. We present these results in Web Appendix Table A2. 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 \$0.02 per km, or \$1.26 per hour of travel. The columns 4-6 present figures for rural roads. As expected, we find higher costs for rural travel: \$0.10 per km, or \$3.54 per hour of travel.

We use these estimates to calibrate village to retailer costs, depending on whether the route includes only local roads or involves travel along trunk roads as well.[21] We then calculate a travel cost-adjusted price of fertilizer for every village in two ways. First, we define the minimum travel cost-adjusted price that is available to villagers from i as follows:

$$r_i^{min} = \min_j \{r_j + c_{ij}\} \tag{1}$$

where r_j is the price at agrovet j and c_{ij} is the cost of traveling to agrovet j, and returning to village i with a bag of fertilizer.[22] This price is relevant if 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.

[21] The reason for this is because we did not measure village-to-village transport costs using only local roads; we only measured village-to-market costs (and markets tend to be more central locations, near trunk roads). Therefore, we have to impute costs of village-to-village travel. For trips which involve travel on trunk roads, we have a direct measure of the cost from origin and destination village to their local market, and we calibrate the cost of trunk road travel.

[22] 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 (Web Appendix Table A3). We do this for a 50 kg bag of fertilizer, the modal amount purchased by farmers. We find that transporting a 50 kg bag 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).

In contrast to (1), farmers may make decisions using a simpler decision rule by simply choosing the nearest retailer (so that prices at all retailers other than the nearest are irrelevant). Under this decision rule, the travel cost adjusted price that farmers face is:

$$r_i^{nearest} = r_{nearest} + c_{nearest,i} \tag{2}$$

We calculate these prices for every village-agrovet and village-market pair. Figure 1 plots CDFs of village-level best prices of inputs, adjusting for travel costs, and shows tremendous heterogeneity in prices across villages. In Panel A, the 90-10 difference for the best travel cost-adjusted price for fertilizer is about 45% of the mean, and the standard deviation is about 21% of the mean. Later we compare the distribution of these delivered prices with those using model-based estimates of both pecuniary and non-pecuniary trade costs.[23]

4.2 Reduced form regression specification

With a non-trivial degree of price variation established, we now evaluate in which villages prices are less favorable, and in particular, how remote villages compare to the average village. When evaluating the relationship between any market conditions and remoteness, the primary specification we utilize is,

$$m_{it} = \beta_r \cdot R_i + \epsilon_{it} \tag{3}$$

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

For the measures of village-level market conditions estimated in (3), we include no controls. However, for farmer outcomes such as input adoption, we naturally include other

[23]Web Appendix Figure A2 shows analogous figures for prices at the nearest location, and figures look similar. Web Appendix D1 Does the same for maize prices, and the ratio of maize prices to input prices to measure profitability. See Appendix D for details.

farmer-specific characteristics. Therefore, these are estimated as:

$$m_{fit} = \beta_r \cdot R_i + \beta_X X_{fit} + \epsilon_{fit} \tag{4}$$

where subscript f refers to farmer and X_{fit} 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.

4.3 Defining remoteness

To measure the remoteness of each village i, we focus on its proximity to selected "hubs" that are proximate to the study regions: Arusha, Babati, Dodoma, Moshi, and Tanga. These locations are chosen because distributors for both maize and fertilizer are commonly located here.[24] 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:

$$remoteness_i = \sum_h d_{hi} pop_h \tag{5}$$

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_{hi} is distance from village i to hub h. 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 (2016), we use distance to measure proximity to hubs, rather than calculate the ad-valorem costs of travel and estimate

[24] Web Appendix Table A4 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 Web Appendix Figure 1.

a distance elasticity (because it simplifies the construction of the measure substantially).[25]

The second measure engages on Donaldson and Hornbeck (2016) more directly, and calculates the market access of each village using the following formulation:

$$MA_i = \sum_h \tau_{hi}^{-\theta} pop_h \tag{6}$$

MA_i 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, $\tau_{hi}^{-\theta}$. The cost term τ_{hi} is calculated as

$$\tau_{hi} = 1 + \frac{2.69 * cost_{hi}}{avgprice} \tag{7}$$

where $cost_{hi}$ 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. For $-\theta$, we appeal to the substitution elasticity across agrovets which is estimated later in the paper to be -7.39, though the results are robust to other values.

We standardize both measures to have mean 0 and standard deviation 1 (and put a negative sign in front of MA_i, so that it measures remoteness rather than market access). The distributions of these variables are illustrated in Web Appendix Figure A3. Both measures feature two modes, and the least remote areas (near major hubs) are approximately -2 to -3 standard deviations from the mean, while the most remote are +1 to +3 standard deviations away. The difference between these (4-5 standard deviations) is useful for benchmarking differences in outcomes between the most and least remote areas (similar to the approach

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

taken in Atkin and Donaldson (2015)).[26]

4.4 Summary statistics and correlations with remoteness

Table 2 presents summary statistics, and shows how these variables vary with remoteness. From Panel A, we see that farmers in more remote areas are less educated, own fewer assets, have less access to finance, and earn less income from sources outside of farming. These farmers also tend to have larger families and larger farms. Hence, in regressions using our farmer surveys, we will present results with and without these additional controls.

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 13% 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 about two times 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. We find clear evidence that output per acre is lower in more remote areas (which is consistent with reduced access to inputs). However, we find that the *total* value of harvest (which is simply the total harvest multiplied by the same average price of output for all farmers) is higher in remote areas. The explanation for this is that farms in more remote areas are larger (as discussed above). Also of note in this table is that the average value of harvest output is less than $200.

In conclusion, Table 2 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,

[26]We show relationship between the two variables in Web Appendix Table A5, in which we regress one measure on the other. The coefficient is 0.81 with an R-squared of 0.66.

this motivates the use of farmer-level controls in our empirical work, and comprehensive economic model to conduct policy counterfactuals.

4.5 Access to input markets

Table 3 shows how the two remoteness measures correlate with access to input markets.[27] We first tabulate access to retailers within 10 km, which is a distance that is reasonably traveled by farmers.[28] 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 3 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 $1.94-2.34, equivalent to just shy of 10% of the mean. This implies a difference in prices of approximately 40-50% 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 per-unit 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.[29]

[27] Web Appendix Table D1 and D2 perform a similar analysis for output markets. We find similarly strong evidence of reduced market access in more remote areas, but do not discuss results in detail in this paper.

[28] Appendix Figure A4 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.

[29] In Appendix Table A8, Panels A and B, we present two 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 A, 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

Panel B2 of Table 3 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 23% of the optimal purchase; for the nearest retailer (Panel B2), they are about 16%. Critically, these results suggest that the optimal (unconstrained) behavior of farmers involves traveling further to find a lower price.[30] Indeed, farmers are estimated to pay 20% more in transport to reach the optimal agovet. In our trips module, which we utilize extensively in the quantitative model, about 1/3 of farmers report *not* choosing the closest agrovet, mostly due to price and quality of the retailer. Further, in 60% of villages, farmers do not uniformly choose the same agrovet, suggesting that there are other factors at play in agrovet choice.

a bounding exercise in Panel B of Appendix Table A8, 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 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.

[30]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 Web Appendix Table A6. 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.

4.6 Farmer decisions

The results so far show clear evidence of reduced market access in more remote areas for input and of higher prices for inputs. These results lead us to expect lower input usage in more remote areas. We investigate this in Panel C of Table 3, where we present results with and without a full set of farmer controls. In almost all specifications, these relationships are strong (significant at 1%) and large. We find that a one standard deviation increase in remoteness reduces fertilizer use by 7-20 percentage points. Since the distance between the least and most remote regions is about 5 standard deviations, the regressions predict at least 35 percentage point lower usage of fertilizer in the most remote villages relative to the least remote villages.

5 Model

In this section, we quantify the impact of access to input markets by developing a rigorous spatial model of fertilizer adoption and retailer pricing. Farmers determine whether to buy fertilizer and where to buy it. A fixed set of agrovets compete in prices, and the industry is be pinned down by spatial supply and demand. We focus on the market for chemical fertilizer, and abstract from endogenous changes to the broader economy (eg. food prices, wages). Within this context, we run counterfactuals that study the role of last-mile trade costs and other frictions in the adoption decision. All derivations are available in Web Appendix A.

5.1 Model Preliminaries

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

$$Y_{fi0} = \widetilde{\theta}_{fi0} K_{fi}^{\alpha_0} L_{fi0}^{\beta_0} \tag{8}$$

Here, $\widetilde{\theta}_{fi0}$ is a baseline productivity measure when not using fertilizer, K_{fi} is land held and L_{fi0} is labor hired/used by farmer f in village i. The coefficients $\alpha_0 > 0$ and $\beta_0 > 0$ represent the intensity of land and labor in production of maize. If the wage rate in village i is w_i and the selling price of maize is p_i, holding land fixed, profits can be derived as

$$\mathit{\Pi}_{fi0} \quad = \quad \theta_{fi0} \pi_{i0} K_{fi}^{\frac{\alpha_0}{1-\beta_0}} \tag{9}$$

where $\theta_{fi0} = (1-\beta)\,(\beta_0)^{\frac{\beta_0}{1-\beta}} \widetilde{\theta}_{fi0}^{\frac{1}{1-\beta_0}}$ and $\pi_{i0} = p_i^{\frac{1}{1-\beta_0}} w_i^{-\frac{\beta_0}{1-\beta_0}}$.[31] The term θ_{fi0} will be represented by a random variable with a village-specific mean, and will be absorbed in estimation.

The production function *with* fertilizer requires both labor and fertilizer as variable inputs, and farmers not only have a choice of how much fertilizer to buy, but also which retailer to choose. Supposing that farmer f from village i buys fertilizer from agrovet j in location v, and maintaining the basic Cobb-Douglas assumption (for tractability), the production function is written as:

$$Y_{fijv} = \widetilde{\theta}_{fijv} \left(\theta_i K_{fi}\right)^{\alpha} L_{fijv}^{\beta\gamma} M_{fijv}^{\beta(1-\gamma)} \tag{10}$$

where M_{fijv} is the quantity of fertilizer purchased. Note we are allowing the exponents on land and labor to be different from the technology without fertilizer, which as we will show below, allows for land to affect adoption decisions.[32] 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 farming. This is meant to capture the local suitability of using fertilizer in village i. The second is a productivity shock for farmer f, $\widetilde{\theta}_{fijv}$, that potentially varies by the agrovet j

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

[32] This maintains the analytical simplicity of a basic Cobb-Douglas technology. The Cobb-Douglas framework to model agricultural production shares similarities with recent work in papers including Chatterjee (2020) and Gollin and Udry (2020), though unlike the latter, we do not model allocations across different plots within the household.

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 village i from agrovet j in location v as r_{ijv}, solving for the optimal labor and fertilizer inputs, profits are written as

$$\Pi_{fijv} = \theta_{fijv} \pi_i r_{ijv}^{-\sigma} K_{fi}^{\frac{\alpha}{1-\beta}} \tag{11}$$

where $\sigma \equiv \frac{\beta}{1-\beta}(1-\gamma)$, $\pi_i = \theta_i p_i^{\frac{1}{1-\beta}} w_i^{-\gamma \frac{\beta}{1-\beta}}$, and $\theta_{fijv} = \kappa_2 \widetilde{\theta}_{fijv}^{\kappa_1}$.[33] Here, the profitability of fertilizer is a function of the productivity shock, θ_{fijv}, the (delivered) price of fertilizer itself, r_{ijv}, land holdings, and deterministic profits based on local factors and technology, π_{fi}.

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 trade 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 $j \in \mathcal{J}_v$ is r_{jv}. The per-unit cost to a farmer from i buying from j in v is $r_{ijv} = r_{jv}\tau_{iv}$, where τ_{iv} is an iceberg trade cost for farmer i buying from v. The iceberg assumption facilitates a decomposition of the model that aids estimation and calibration.

We assume that θ_{fijv} is a random variable that measures the benefit of farmer f who lives in i purchasing at agrovet j in location v. The central tendency of this random variable will depend on benefits of purchasing at agrovet j in location v, and could represent other inputs or information acquired at location v, or unobserved quality at a retail location. For analytical convenience we assume that θ_{fijv} is distributed according to a Fréchet distribution with location parameter T_{jv} and dispersion parameter ε: $\Pr(\theta_{fijv} < \theta) = \exp(-T_{jv}\theta^{-\varepsilon})$. That is, while each farmer from i may get a random draw from this distribution, its central moments are specific to the retail location itself. The unconditional distribution of profits

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

for farmer f from village i buying from agrovet j in location v is written as:

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

We also assume that the outside option of not buying fertilizer is random, with θ_{fi0} distributed Fréchet with location parameter T_{i0} and the same dispersion parameter ε. Thus, the distribution of profits without fertilizer is written as:

$$\Pr\left(\mathit{\Pi}_{fi0} < \pi\right) = \exp\left(-T_{i0}\pi_{i0}^{\varepsilon}\pi^{-\varepsilon}\right)$$

Here, average productivity when not buying fertilizer varies 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.).

Farmers choose among locations $v \in \mathcal{V}$ and agrovets $j \in \mathcal{J}_v$ at each location to find the most profitable option. Solving a standard discrete choice problem, the probability that farmer f in village i buys from agrovet j at location v is written as:

$$\lambda_{fijv} = \frac{T_{jv} r_{ijv}^{-\varepsilon_a} K_{fi}^{\varepsilon_k}}{T_{i0}\left(\frac{\pi_{i0}}{\pi_i}\right)^{\varepsilon} + \sum_{v'\in\mathcal{V}}\sum_{l\in\mathcal{J}_{v'}} T_{lv'} r_{ilv'}^{-\varepsilon_a} K_{fi}^{\varepsilon_k}} \tag{12}$$

Here, $\varepsilon_a = \varepsilon\sigma$ and $\varepsilon_k = \varepsilon\left(\frac{\alpha}{1-\beta} - \frac{\alpha_0}{1-\beta_0}\right)$, with ε_a being the critical substitution elasticity across agrovets. The probability of adoption by farmer f in village i is written as:

$$\mu_i = \frac{K_{fi}^{\varepsilon_k}\sum_{v'\in\mathcal{V}}\sum_{l\in\mathcal{J}_{v'}} T_{lv'} r_{ilv'}^{-\varepsilon_a}}{T_{i0}\left(\frac{\pi_{i0}}{\pi_i}\right)^{\varepsilon} + \sum_{v'\in\mathcal{V}}\sum_{l\in\mathcal{J}_{v'}} T_{lv'} r_{ilv'}^{-\varepsilon_a} K_{fi}^{\varepsilon_k}} \equiv \frac{\mathit{\Phi}_i K_{fi}^{\varepsilon_k}}{\mathit{\Phi}_{i0} + \mathit{\Phi}_i K_{fi}^{\varepsilon_k}} \tag{13}$$

Note that the elasticity ε_k is crucial for determining whether there is within-village variation in adoption. If $\varepsilon_k = 0$, which is true if technologies exhibit HD1 production, then the probability of adopting fertilizer will be identical for every farmer in village i. When $\varepsilon_k \neq 0$ farm size will lead to variation across farmers, within villages.

In (13), we also define the two village-level terms that characterize the adoption decision for each farmer f from i along with the individual land-holdings raised to ε_k. First, we define village i's *market access* to inputs as $\Phi_i \equiv \sum_{v \in \mathcal{V}} \tau_{iv}^{-\varepsilon_a} \phi_v$, where $\phi_v = \sum_{l \in \mathcal{J}_v} T_{lv} r_{lv}^{-\varepsilon_a}$. Market access is a function of the elasticity-adjusted iceberg to a given village, $\tau_{iv}^{-\varepsilon_a}$, and a local retailer index, ϕ_v, which is the sum of elasticity-adjusted local prices weighted by the local input-quality. Second, we define $\Phi_{i0} = T_{i0} \left(\frac{\pi_{i0}}{\pi_i}\right)^{\varepsilon}$, which is the relative profitability of using fertilizer, adjusted for local productivity factors that are unrelated to market access or land holdings. This term will be absorbed in the calibration strategy defined below.

To build intuition that will help in understanding the counterfactuals, we now discuss the role of changes in market access, and how it leads to changes in adoption by a given farmer. Defining the value of a variable after a shock as X', we derive that the exact log change in individual adoption probabilities in response to a shock is the following:

$$\log\left(\frac{\mu'_{fi}}{\mu_{fi}}\right) = \log\left(\frac{\Phi'_i}{\Phi_i}\right) - \log\left(1 + \mu_{fi}\left(\frac{\Phi'_i}{\Phi_i} - 1\right)\right) \tag{14}$$

Here, the log change in adoption will be an additive function of the log change in market access, and a non-linear function in which the percent change in market access is interacted with an individuals baseline adoption probability. It is easy to show through the second term that market access has a larger impact on those farmers with a lower adoption probability. Further, the non-linear nature of the second term shows why analysis is required at the farmer level for an accurate counterfactual, rather than by village level averages. When presenting the results below, we will not only focus on the exact change in log adoption, but also the log change in market access, and also the interaction term, to better interpret in what way changes to market access are driving the results.

To conclude the farmers problem, we derive expected farmer-level expenditures,

$$\mathbf{E}[F_{fi}] = \kappa \Phi_i^{\frac{1}{\varepsilon}} \pi_i K_{fi}^{\frac{\alpha}{1-\beta}} \mu_{fi}^{\frac{\varepsilon-1}{\varepsilon}} \tag{15}$$

where κ is a function of model parameters.

5.2 Spatial Equilibrium and Pricing

The first equilibrium condition relevant for the analysis is the market clearing condition for fertilizer at each agrovet. Defining expected agrovet sales at j in v as $\mathbb{E}[v_{jv}]$, expenditures at this agrovet as aggregated from expected spatial farmer-level demand is written as:

$$\mathbb{E}[v_{jv}] = \sum_i N_i \sum_f \mu_{fi} \lambda_{ijv|adopt} \mathbb{E}[F_{fi}|adopt\ at\ jv]$$

where $\mathbb{E}[F_{fi}|adopt\ at\ jv]$ is expected fertilizer expenditures by farmer f in i, conditional on adopting at jv, and N_i is the village population. Using the properties of the Fréchet distribution, it is straightforward to show that $\mathbb{E}[F_{fi}|adopt\ at\ jv] = \mathbb{E}[F_{fi}|adopt]$.[34] Combining with $\mathbb{E}[F_{fi}] = \mu_{fi}\mathbb{E}[F_{fi}|adopt]$, and defining $\mathbb{E}[F_i] = \sum_f \mathbb{E}[F_{fi}]$ as expected expenditures at the village level, we get:

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

This equation will be used for the calibration of agrovet quality, and will also be satisfied in any counterfactual equilibrium.

The second equilibrium condition is agrovet pricing where within the study region agrovets compete in prices, conditional on the quality of their own product, the spatial location of demand, and the pricing/quality of their competitors. Entry is fixed, though we discuss this limitation further within the counterfactuals The first order condition for an agrovet j located in village v is a mark-up over marginal cost:

$$r_{jv} = \frac{\varepsilon^v_{jv} - 1}{\varepsilon^v_{jv}} c_{jv} \tag{16}$$

[34]See technical appendix for a proof.

where c_{jv} is the marginal cost for agrovet j in location v, and ε^v_{jv} as the price elasticity of revenue at that location. Defining $s_{ijv} = \frac{N_i \lambda_{ijv|adopt} \mathbb{E}\left[F_{fi}\right]}{\sum_{i'} N_{i'} \lambda_{i'jv|adopt} \mathbb{E}\left[F_{fi'}\right]}$ as the expenditure share of farmers from i within jv, and $\omega_{fi} = \frac{\mathbb{E}\left[F_{fi}\right]}{\sum_{f'} \mathbb{E}\left[F_{f'i}\right]}$ as the expenditure share of farmer f within i, in the technical appendix ε_{jv} is derived as:

$$\varepsilon_{jv} = -\varepsilon_a + \varepsilon_a \frac{\varepsilon - 1}{\varepsilon} \sum_i s_{ijv} \lambda_{ijv|adopt} \sum_f \omega_{fi} \mu_{fi} \tag{17}$$

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, $\lambda_{ijv|adopt} \approx 0$ for all i and the mark-up is pinned down by the substitution across agrovets through the elasticity, ε_a.[35]

5.3 Model Calibration

The calibration of the model will proceed in four parts. At the farmer level, the two equations to calibrate are individual adoption probabilities, and expected fertilizer expenditures conditional on adoption. Second, elasticity-adjusted trade costs will be estimated using the unique "trips" data on retailer choice from the farmer surveys within a multinomial logit framework. Third, on the aggregate level, we will calibrate the model using the market clearing condition as described in (16), and back-out agrovet-specific quality terms from this calibration. Finally, we will calculate agrovet-level demand elasticities to back-out marginal costs from the pricing equation. After these steps, we will conduct counterfactuals on "last-mile" access to agrovets, and measure the impacts on adoption, expenditures, and the remoteness gradient.

[35] If production is constant returns, $\mu_{fi} = \mu_i$, and since $\sum_f \omega_{fi} = 1$, we can write:

$$\varepsilon_{jv} = -\varepsilon_a + \varepsilon_a \frac{\varepsilon - 1}{\varepsilon} \sum_i s_{ijv} \lambda_{ijv} \tag{18}$$

Thus, more flexible production leads to a different mark-up equation

5.3.1 Adoption and Expenditures

As described in (13), adoption follows a standard logit form with village level fixed-effects, and a coefficient on the log of land held by each farmer. The coefficient on this term, ε_k, measures the effect of land holdings on the relative incentive to adopt fertilizer. This coefficient is non-zero within the context of the model if technologies with and without fertilizer have different intensities of land vs. variable factors. In practice, any vector of fixed attributes can be treated in the same way as land in this equation, and indeed, we include a vector fixed individual-level controls in the preferred specification used for calibration. The estimates of this logit model with fixed effects is presented in Table B1. Here, we present results with and without additional farmer controls (treated in the same way as capital) as used in Table 4. The estimate $\varepsilon_k = 0.167$ indicates a significant and positive effect of the log(land) on use of fertilizer, as well as other attributes. The predicted probabilities of adoption, $\widehat{\mu}_{fi}$, and their changes under counterfactuals, will be the focus on the empirical exercise. Moving forward, we use the predicted probabilities from the regression with controls.[36]

Expenditure measures are required for the market clearing conditions and for agrovet pricing. However, since all farmers do not report adopting fertilizer, we must impute expected expenditures using the structure of the model and underlying selection into adoption. For farmers who adopt, we simply use the observed level of expenditures as $\mathbb{E}\left[F_{fi}|adopt\right]$ as an unbiased estimate of this term, and then use $\mathbb{E}\left[F_{fi}\right] = \mathbb{E}\left[F_{fi}|adopt\right] \cdot \widehat{\mu}_{fi}$ to calculate unconditional expenditures. To calibrate unconditional expected expenditures for those farmers that do not adopt, we run a Poison model using (15) and reported expenditures in the sample, including those with zero expenditures. The predictions of this model are assigned to farmers

[36] There are a number of villages for which predicted adoption probabilities cannot be estimated using the above model because there is uniform adoption (or non adoption) within each village (and hence, the village fixed effect cannot be estimated). To compute probabilities for these villages, we use the estimates from the logit in Table B1, with controls, and then choose village fixed effect values such that the average adoption in non-adoption villages is 0.0275, and adoption in full adoption villages is 0.975. Results using different trimming values are available in Web Appendix B7.

that do not adopt for their value of $\mathbb{E}[F_{fi}]$. Estimation details for expected expenditures are presented in Appendix B, and results presented in Table B1.

5.3.2 Elasticity adjusted trade costs through retailer choice

A novel component of the analysis is estimating the revealed trades costs faced by farmers. To do this, we exploit the discrete choice structure of the model, and estimate trade costs using multinomial logit. To begin, we note that (12) can be broken up into the probability of adoption by farmer f from village i, μ_{fi}; the probability that a farmer from 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, $\lambda_{j|adopt\,at\,v}$:

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

The farmer surveys collect data to calculate all three probabilities, though since some retailers don't all have known names and signage, the first two are used in this paper. For estimating trade costs, $\lambda_{iv|adopt}$ contains the all relevant information. Exponentiating the village share equation, and re-writing $\log(\phi_v)$ into a location v fixed effect, d_v, we can write:[37]

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

To estimate equation (19), we need a specification for the ad-valorem trade cost, τ_{iv}. Mimicking the reduced form, we will measure the kilometers traveled on main roads and rural roads separately between a given village i and agrovet location v. However, since we are modeling consumer behavior rather than accounting for pecuniary costs, we adopt a flexible

[37]Note that we have suppressed f's since this equation is conditional on adoption, which conditions out any farmer-level effects. We have experimented with interactions between trade costs and non-farming income, and these results are presented in Online Appendix B5.

non-linear relationship between road type distance and agrovet choice. To do so in a parsimonious fashion, we assign distance bins for each road type, where $Main_{iv}^{b}$ will be equal to one if the distance on main roads between i and v is in bin b, and $Rural_{iv}^{b}$ will be defined analogously for rural roads. The specification taken to estimation can be written as:

$$\lambda_{iv|adopt} = \frac{\exp\left(d_v + \sum_{b \in B_m} \beta_b^m Main_{iv}^b + \sum_{b \in B_r} \beta_b^r Rural_{iv}^b\right)}{\sum_{v' \in \mathcal{V}} \exp\left(d_{v'} + \sum_{b \in B_m} \beta_b^m Main_{iv'}^b + \sum_{b \in B_r} \beta_b^r Rural_{iv'}^b\right)} \tag{20}$$

Equation (20) can be estimated by McFadden's alternative-specific conditional logit. The results from doing so are presented in Table 4. The estimates for main road bins are in the left-hand column, and those for rural roads are on the right. To the right of each column is an estimate of the ad-valorem equivalent trade cost per kilometer, when evaluated at the farthest distance that defines each bin, and with trade costs compounded each kilometer.[38] For this interpretation, we use a value of $\varepsilon_a = 7.39$, which is estimated using an IV regression in the next section using data obtained from market clearing conditions.

To interpret the estimates, let us first begin with the bins for main roads. When interpreting main road coefficients on their own, we are presenting the marginal effects of main roads, conditional on whatever travel occurred on rural roads. Traveling a short distance on a main road does not appear to effect adoption choices. However, when traveling more the 5KM on a main road, the likelihood of choosing an agrovet between 5 and 10km away compared to a local agrovet, conditional on the price and quality of those retailers is 0.165.[39]. When looking at rural roads, there is a large reduction in the likelihood of choosing an agrovet with any travel on rural roads compared with no travel on rural road, with only a 0.1 probability of choosing an agrovet outside your village for idiosyncratic reasons.

[38]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.

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

$$\frac{\lambda_{0-5km}}{\lambda_{0km}} = \frac{\exp\left(d_v - 1.8\right)}{\exp\left(d_v - 0\right)} = 0.165$$

Interestingly, travel on rural roads appears to be more discrete in nature, where needing to travel any distance (under 20km) on the margin leads to a similar reduction in the probability of choosing a different agrovet. This is in contrast with main roads, where distance has a minimal impact at first, but an increasingly negative impact at longer distances.

Since not every village has its own agrovet (according to Table 3, only 62% have one within 10 km), a better interpretation is the average ad-valorem trade cost to the closet agrovet. Using our MNL estimates, and again using $\varepsilon_a = 7.39$, the mean and median ad-valorem equivalent trade costs to the lowest trade cost agrovet is 72% and 37%, respectively. These are extremely large trade costs, and approximately 4 times average pecuniary costs of travel as summarized in Panel B2 of Table 3.

5.3.3 Market Clearing

For the second step of the calibration, we solve for the vector of quality adjusted fertilizer prices $\eta_{jv} \equiv T_{jv} r_{jv}^{-\varepsilon_a}$ that exactly equates supply and expected demand for fertilizer at each agrovet.[40] Using (16) and imposing the specification for transportation costs, we get:

$$\mathbb{E}\left[v_{jv}\right] = \sum_i N_i \left(\frac{\exp\left(\sum_{b \in B_m} \widehat{\beta}_b^m Main_{iv}^b + \sum_{b \in B_r} \widehat{\beta}_b^r Rural_{iv}^b\right) \eta_{jv}}{\sum_{v' \in \mathcal{V}} \sum_{l \in \mathcal{J}_{v'}} \exp\left(\widehat{\beta}_b^m Main_{iv'}^b + \sum_{b \in B_r} \widehat{\beta}_b^r Rural_{iv'}^b\right) \eta_{lv'}} \right) \mathbb{E}\left[F_i\right] \quad (21)$$

where $\mathbb{E}\left[F_i\right]$ is construted from the calibrated farmer-level expenditure levels. Since agrovet fertilizer revenues and farmer expenditures are from different surveys, 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 (21), under the normalizing

[40] The sampling of villages complicates the calibration of the market equilibrium, where 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 bias the estimates of η_{jv} by assigning a large value for agrovet locations without any surveyed farmers to make-up for the incorrectly assigned remoteness. To correct for this, we assume that unsampled villages have an identical sample of farmers to the closest sampled village.

assumption that $\sum_v \sum_j \eta_{jv} = 1$.[41]

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

$$\hat{\Phi}_i = \sum_{v \in \mathcal{V}} \exp\left(\sum_{b \in B_m} \widehat{\beta}_b^m Main_{iv}^b + \sum_{b \in B_r} \widehat{\beta}_b^r Rural_{iv}^b\right) \sum_{l \in \mathcal{J}_v} \hat{\eta}_{lv}$$

Changes to this measure of market access will be crucial in characterizing the exact impact of counterfactual changes in trade costs.

5.3.4 Elasticities and markups

We now estimate the key elasticity in the paper that guides any change in trade costs, ε_a, using the calibrated values of $\hat{\eta}_{jv}$. Taking logs of η_{jv}, we get:

$$\log(\hat{\eta}_{jv}) = -\varepsilon_a \log(r_{jv}) + \log(T_{jv})$$

If we estimate this equation literally, proxying for $\log(T_{jv})$ using the log of an agrovet owner's experience, and district fixed effects, we get an estimate of 4.6. However, as $\hat{\eta}_{jv}$ is calibrated using revenue-expenditure market-clearing conditions, and $\log(T_{jv})$ represents unobserved quality in these conditions, there is an endogeneity problem in estimating ε_a. To address this endogeneity, we can either instrument for the mark-ups that are embedded in prices, or the marginal cost of fertilizer at that location. As a simple instrument, we use the wholesale price at which the agrovet was sold fertilizer but lagged from the year of analysis to avoid any mechanical issues with how η_{jv} was estimated. The full IV regression, which is presented in Appendix Table B2, includes log experience of the agrovet owner and district fixed effects. The first stage is strong, though the question remains whether the instrument satisfies the

[41] This is required as the probabilities within the sum in equation (21) are homogeneous degree zero in η_{jv}. Eckert et al. (2019) 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.

exclusion restriction. Since the wholesale price is from a prior year, and set by another party (who are often distributors for extremely large firms outside the study area), there is no mechanical relationship with η_{jv}. One possible threat to the validity of this instrument is if higher quality agrovets are more likely to find lower prices through experience in search. As we are also including experience as a control, we believe the instrument to be appropriate in this regard. Estimating the IV regression using this instrument, we get a value of $\varepsilon_a = 7.39$ which we use for the counterfactuals. We also present a full set of results for different values of ε_a in Appendix Table B4.[42] In terms of the reasonableness of this estimate, we note that it is within "typical" bounds for substitution elasticities for Urea fertilizer.[43] Our estimate will also help rationalize observed mark-ups, as we discuss shortly.

As illustrated in the mark-up equation, we need a value of ε to characterize any changes in pricing due to counterfactual shocks. Recalling that $\varepsilon_a = \varepsilon \left(1 - \gamma\right) \frac{\beta}{1-\beta}$, to back out ε, we need estimates for γ, the relative weight on labor within variable factors, and β, the weight on variable factors in the production function, both when using fertilizer. There are a variety of approaches to estimate these parameters; in the appendix, we implement one using basic production function estimation. However, the results are ultimately not particularly sensitive to these parameters, something we document in Appendix Table B4. So to move forward with the counterfactuals, we assume that $\varepsilon \approx 2\varepsilon_a$ (which is a central estimate from the production function approach) and refer the reader to the appendix for details of different parameter estimates, and a table describing the sensitivity of our results.

The last step prior to running counterfactuals is to calibrate marginal costs using the

[42]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.

[43]Usings WITS data, this ranges between 4.88 to 11.78. Results available upon request.

mark-up equation. Mark-ups can be calculated using (18), and then using (16), marginal cost are calculated by dividing the observed price by the mark-up over marginal cost.

5.3.5 Model Performance

The calibrated model matches quite well the observed behavior that we see in the raw data. The first place to observe this is in the estimation of mark-ups. The predicted markups have a mean and median of roughly 14.05%, which is almost exactly what we find in the reduced form (13.4% - see Web Appendix A6). This is notable because estimated mark-ups do not use any marginal cost information measured for retailers.

Second, we focus on the observed behavior of farmers, and the predicted share of villagers who purchase at the closest agrovet. In the raw data, 67% of farmers report purchasing at their closest agrovet, with the rest traveling a further distance. In the model, we estimate that for 57.3% of villagers the agrovet with the highest probability of being chosen is the closest agrovet. So these moments are comparable to each other, though we are undershooting a bit. To further evaluate this, we calculate the relationship between the share of purchases of farmers in each village at an agrovet location and the predicted probability that farmer chooses that agrovet location. Running a Poisson model (to account for zeros) of actual trip shares on the log of the predicted probabilities, we get an estimate of 0.69, which is also highly significant. Thus, after calibration of quality terms in the model, and using our census of prices, our model predicts actual travel behaviour quite nicely. Both pieces of evidence here suggests that the model does an adequate job at modeling actual farmer behaviour.

6 Counterfactuals

In this section, we use the calibrated model to evaluate counterfactuals on input market access. To implement the counterfactuals, we solve for new fertilizer prices that satisfy the first order conditions for pricing in (16), the market clearing conditions in (21) while taking into account equilibrium changes in the farmer's problem in response to new agrovet prices and/or trade costs. Aggregate terms such as output prices, wages, and entry are assumed to be fixed, as are individual characteristics used in the calibration of the model. Thus, these counterfactuals can be considered the response of the existing industry equilibrium for retail fertilizer to last-mile trade cost shocks. Sensitivity analysis to a larger support of parameter values is presented in Online Appendix B4.

To describe the results, we focus on four key metrics, how they change under the counterfactuals, and how the counterfactual changes the relationship of remoteness to these metrics. The four key metrics are log adoption, log expected expenditures, log market access, and the second term in equation (14) that summarizes the effects of market access interacted with baseline adoption. We also provide an estimate for the overall change in surplus in the fertilizer market, and how division of this surplus affects farmers.[44]

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. The first row of Panel B in Table 5 reports the results from this counterfactual on the four metrics described above. The results indicate a large effect of the 50% reduction in trade costs on adoption, and a reduction of the remoteness gradient. Precisely, log adoption increases by 1.38, and increases an additional 0.41 with a one standard deviation of remoteness. Compared to the baseline relationship between log adoption and remoteness this corresponds to a 63% reduction in the remoteness gradient. Similar effects

[44] Agrovet surplus is directly calculated in the counterfactuals after calibrating mark-ups based on equilibrium market shares. Farmer surplus is a straightforward implication of proportionality in the production side of the model. Precisely, $\mathbf{E}[F_{fi}] = \sigma \mu_{fi} \mathbf{E}[\Pi_{fi}]$, where $\sigma = \varepsilon_a/\varepsilon$. Since we have estimates for σ, $E[F_{fi}]$, and μ_{fi}, we can calculate $\mathbf{E}[\Pi_{fi}]$ easily, and calculate the change in total surplus, as well as the change in the farmer share of surplus.

are seen with log of expenditures. In column three, we see that log market access increases by 2.23, with a 7.8% additional increase per standard deviation of remoteness. The interaction between market access and baseline adoption is also particularly important, and appears to drive much (though not all) of the change in the remoteness gradient under the 50% reduction in trade costs. Overall, the results show a clear effect reduced trade costs on adoption, expenditures, and market access, a larger effect in more remote areas, and a more pronounced effect on those farmers with low baseline adoption. In terms of welfare, in the last column of Panel B, total surplus within the fertilizer market increase by 23% after this reduction in transaction costs. However, this shifts the division of surplus toward agrovets, where the share of the market surplus falls from 0.8 by 0.075.

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 though of poor quality. Thus, both types of road improvements may be necessary to quantify. Further, 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.

By using the parameterization of trade costs in section 4.2, counterfactuals for cutting main road costs by 50%, and rural road costs by 50%, are presented in rows two and three of Panel B of Table 5. Both counterfactuals increase log adoption significantly, and more so in remote areas, but interestingly, it appears that main road improvements have a much larger effect on adoption and the remoteness gradient. This is most apparent in the effects of the different road improvements on log market access, where rural road improvements do not disproportionately improve market access in remote areas, but main roads do. One possible explanation for this result is that in remote areas, farmers tend to be farther from the nearest village with an agrovet, and that travel is necessary on main roads to reach additional agrovets. Thus, if one is to increase the probability of adoption for the most remote farmers disproportionately, better main-road access is necessary, not necessarily just rural roads. This may be different in other contexts, but in the present area of study it appears that further trunk road improvements would be required to yield a larger impact in access.

Distributor-Agrovet Costs

We also evaluate how the input sourcing costs for retailers affect the adoption decision. We document in our reduced-form analysis that sourcing costs rise significantly with remoteness. To evaluate the effects of wholesale distribution, we halve distributor-agrovet transportation costs. The results are presented in row four of Panel B in Table 5. Log adoption rises by about 0.142, and this increases by 0.055 with a standard deviation in remoteness. Interestingly, in this case, these impacts are driven almost exclusively by the market access term, and not the interaction between baseline adoption and market access.[45]

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. One likely reason is profitability, and the costs of opening a business in areas relative to lower demand in those areas in the first place. This represents a limitation of the analysis, though a full entry model is all 1180 villages is beyond the scope of the current manuscript. However we can use counterfactuals to study the impact of potential entry, and any possible bias in our counterfactual estimates.

[45]We have also experimented with generic price subsidies, as used in many developing contexts. Results for these policy experiments are available upon request

To study the baseline profitability of potential entry, we run a simple counterfactual to examine how profitability may vary with the remoteness of villages. Specifically, we force an average quality agrovet to enter every village in the sample (one at a time, not simultaneously), and then measure the profitability of the entrant after entry. We also compute the profitability of the entrant after the counterfactual 50% reduction in trade costs. We then regress the profitability of entry, before and after the reduction in trade costs, on the remoteness of the village in which the entry occurred. These results are presented in Panel C of Table 5. Prior to the reduction in trade costs, there was a strong, negative relationship between the profitability of the hypothetical entry and remoteness. Specifically, a one standard deviation increase in remoteness reduces the log(profitability) of entry by -1.141. After reducing last-mile costs by 50%, the relationship between remoteness and profitability of entry falls to -0.831 log points per standard deviation of remoteness, or 27% . Importantly, this suggests that the main counterfactual disproportionately increases the profitability of entry in more remote areas. In this sense, allowing for endogenous entry would amplify the reduction in the remoteness gradient that we document without entry.[46]

7 Conclusion

We collect detailed data on transportation costs, input prices, and input usage along the supply chain for fertilizer in 1,180 villages in the Kilimanjaro and Manyara regions of Tanzania. These villages comprise 99.7% of the villages in the region and thus represent essentially the universe of locations. We find that there is meaningful price dispersion, especially when accounting for travel costs. Access to retailers for inputs is much lower in remote regions, and consequently farmers in remote villages are much less likely to use fertilizer. Counterfactuals suggest that lowering transportation costs would substantially reduce the gradient

[46]We have experimented with an endogenous entry process that is overlayed on the existing model, and simulations confirm our conjecture that an endogeneous entry would serve to amplify the improvements to the remoteness gradient. Results available upon request.

between input usage and remoteness.

An important question is whether our results generalize to other settings. We provide two sets of evidence in the appendix. First, in Web Appendix D, we replicate much of our analysis for output sales - although our surveys are designed and focused on fertilizer purchases, we also find conditions such that remote regions are likely to face worse selling prices, and accordingly, less likely to sell output. Second, in Web Appendix C we use secondary datasets and a dataset of prices we collected in our study area to document that the patterns in Northern Tanzania are similar to other African countries.

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.

References

Aggarwal, Shilpa. 2018. "Do rural roads create pathways out of poverty? Evidence from India." *Journal of Development Economics*, 133: 375–395.

Aggarwal, Shilpa. 2021. "The long road to health: Healthcare utilization impacts of a road pavement policy in rural India." *Journal of Development Economics*, 102667.

Allen, Treb. 2014. "Information frictions in trade." *Econometrica*, 82(6): 2041–2083.

Allen, Treb, and David Atkin. 2016. "Volatility and the Gains from Trade." National Bureau of Economic Research.

Asher, Sam, and Paul Novosad. 2020. "Rural roads and local economic development." *American Economic Review*, 110(3): 797–823.

Asturias, Jose, Manuel García-Santana, and Roberto Ramos. 2019. "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.

Atkin, David, and Dave Donaldson. 2015. "Who's getting globalized? The size and implications of intra-national trade costs." National Bureau of Economic Research.

Beaman, Lori, Dean Karlan, Bram Thuysbaert, and Christopher Udry. 2013. "Profitability of fertilizer: Experimental evidence from female rice farmers in Mali." *American Economic Review (Papers & Proceedings Issue)*, 103(3): 381–86.

Bergquist, Lauren Falcao, and Michael Dinerstein. 2020. "Competition and entry in agricultural markets: Experimental evidence from Kenya." *American Economic Review*, 110(12): 3705–47.

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

Casaburi, Lorenzo, and Tristan Reed. 2019. "Interlinked Transactions and Competition: Experimental Evidence from Cocoa Markets."

Casaburi, Lorenzo, Rachel Glennerster, and Tavneet Suri. 2013. "Rural Roads and Intermediated Trade: Regression Discontinuity Evidence from Sierra Leone."

Chatterjee, Shoumitro. 2020. "Market Power and Spatial Competition in Rural India." Unpublished.

Dhingra, Swati, Silvana Tenreyro, et al. 2021. "The Rise of Agribusinesses and its Distributional Consequences." CEPR Discussion Papers.

Donaldson, Dave. 2015. "The gains from market integration." *Annual Review of Economics*, 7(1): 619–647.

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

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

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 (Papers & Proceedings Issue)*, 98(2): 482–88.

Eaton, Jonathan, and Samuel Kortum. 2002. "Technology, geography, and trade." *Econometrica*, 70(5): 1741–1779.

Eckert, Fabian, et al. 2019. "Growing apart: Tradable services and the fragmentation of the us economy." *Working Paper, Yale University.*

Foster, Vivien, and Cecilia Briceño-Garmendia. 2010. *Africa's infrastructure: a time for transformation.* World Bank.

Gebresilasse, Mesay. 2021. "Rural roads, agricultural extension, and productivity." *Job Market Paper.*

Harou, Aurélie P., Malgosia Madajewicz, Hope Michelson, Cheryl A. Palm, Nyambilila Amuri, Christopher Magomba, Johnson M. Semoka, Kevin Tschirhart, and Ray Weil. 2022. "The joint effects of information and financing constraints on technology adoption: Evidence from a field experiment in rural Tanzania." *Journal of Development Economics*, 155: 102707.

Head, Keith, and Thierry Mayer. 2011. "Gravity, market potential and economic development." *Journal of Economic Geography*, 11(2): 281–294.

Jeong, Dahyeon. 2020. "Creating (Digital) Labor Markets in Rural Tanzania."

Minten, Bart, Bethlehem Koru, and David Stifel. 2013. "The last mile (s) in modern input distribution: Pricing, profitability, and adoption." *Agricultural economics*, 44(6): 629–646.

National Bureau of Statistics, Tanzania. 2013. "Population and housing census 2013: General Report." National Bureau of Statistics: Dar es Salaam.

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): 327–66.

Porteous, Obie. 2020. "Trade and agricultural technology adoption: Evidence from Africa." *Journal of Development Economics*, 144: 102440.

Redding, Stephen, and Anthony J Venables. 2004. "Economic geography and international inequality." *Journal of international Economics*, 62(1): 53–82.

Redding, Stephen J, and Daniel M Sturm. 2008. "The costs of remoteness: Evidence from German division and reunification." *American Economic Review*, 98(5): 1766–97.

Sotelo, Sebastian. 2020. "Domestic trade frictions and agriculture." *Journal of Political Economy*, 128(7): 2690–2738.

Suri, Tavneet. 2011. "Selection and comparative advantage in technology adoption." *Econometrica*, 79(1): 159–209.

Udry, Christopher, and Santosh Anagol. 2006. "The return to capital in Ghana." *American Economic Review*, 96(2): 388–393.

World Bank. 2007. *World development report 2008: Agriculture for development.* The World Bank.

World Bank. 2017. "Enabling the business of agriculture 2017."

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. Remoteness and farmer characteristics

	(1)	(2)	(3)
	Mean	(Standardized) coefficient on remoteness measure based on (population-weighted):	
		Distance to hubs	Elasticity-adjusted travel costs to hubs
Panel A. Demographic and background characteristics			
Age	49.68	-0.92*	-1.44***
	(15.22)	(0.50)	(0.47)
Female	0.45	-0.02	-0.02
		(0.02)	(0.02)
Married	0.76	0.00	0.02*
		(0.01)	(0.01)
Household size	4.95	0.27**	0.32***
	(2.75)	(0.11)	(0.10)
Years of education	6.55	-0.29**	-0.42***
	(3.55)	(0.11)	(0.11)
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	5.41	1.33**	2.50***
	(13.49)	(0.56)	(0.65)
Has market business	0.27	-0.05***	-0.06***
		(0.01)	(0.01)
Annual total income from non-farming (USD)	400.2	-66.70**	-82.78***
	(761.00)	(29.57)	(28.38)
Panel B. Production Capacity (in kg/acre)[1]			
FAO-GAEZ production capacity for low input level	768.1	38.31	28.99
	(291.30)	(24.42)	(20.91)
FAO-GAEZ production capacity for high input level	3206	-280.98***	-279.52***
	(870.40)	(60.15)	(57.70)
FAO-GAEZ production difference between high and low	2438	-319.29***	-308.51***
	(734.40)	(49.35)	(47.79)
Panel C. Harvest Output			
Harvest output per acre	458.1	-76.45***	-80.85***
	(435.80)	(18.42)	(16.83)
Value of harvest output at average post-harvest price in the sample	197.5	-0.65	25.41**
	(288.60)	(10.86)	(11.27)

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 3. Access to input markets, fertilizer retail price heterogeneity, and adoption

	(1)	(2)	(3)	(4)	(5)
		(Standardized) coefficient on remoteness measure based on (population-weighted):			
	Mean	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 which sells fertilizer	0.62	-0.08***		-0.11***	
		(0.01)		(0.01)	
Number of agrovets within 10 km of village which sells fertilizer	4.14	-1.45***		-2.17***	
	(5.58)	(0.12)		(0.14)	
Distance to nearest agrovet which sells fertilizer	13.32	3.26***		3.96***	
	(27.33)	(0.96)		(0.85)	
Distance to the second nearest village with an agrovet which sells fertilizer	30.72	1.43		4.86***	
	(44.80)	(1.44)		(1.37)	
Panel B1. Travel-cost adjusted prices faced by farmers					
Minimum travel-cost adjusted price for 50 kg of Urea (USD)[1]	24.38	1.94***		2.34***	
	(5.41)	(0.17)		(0.14)	
Decomposition of price between retail price and cost of transportation					
Retail price at the location with the lowest travel-cost adjusted price (USD)	19.74	0.90***		1.10***	
	(2.55)	(0.07)		(0.06)	
Cost of travel to obtain minimum travel-cost adjusted price (USD)	4.638	1.04***		1.24***	
	(5.00)	(0.16)		(0.14)	
Panel B2. Travel-cost adjusted prices at the nearest agro-input shop					
Travel-cost adjusted price at the nearest input seller for 50 kg of Urea (USD)[1]	27.25	1.76***		2.12***	
	(7.23)	(0.22)		(0.19)	
Decomposition of price between retail price and cost of transportation					
Retail price at the nearest input seller (USD)	23.49	0.80***		1.02***	
	(3.34)	(0.09)		(0.08)	
Cost of travel to the nearest input seller (USD)	3.77	0.96***		1.10***	
	(6.02)	(0.19)		(0.16)	
Panel C: Input usage					
Used chemical fertilizer in previous long rains	0.39	-0.16***	-0.07**	-0.20***	-0.13***
		(0.03)	(0.03)	(0.03)	(0.03)
Quantity of chemical fertilizer used (kg)	19.55	-12.46***	-5.77***	-14.30***	-9.27***
	(31.39)	(2.04)	(1.58)	(1.83)	(1.82)
Controls for farmer and soil characteristics?		N	Y	N	Y

Notes: For Panels A and B, the unit of observation is the village. Data is from the near universe of villages in Kilimanjaro and Manyara regions (N = 1,180). Travel costs imputed from transport surveys and Google maps. For Panel C, the unit of observations is farmer (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 Columns 2-5, standard errors in parentheses; standard errors are clustered at the village level in Panel C.

*, **, 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 4. Multinomial logit of agrovet choice

	Main Roads			Rural Roads	
	Coef (se)	AVE/KM (elas=7.39)		Coef (se)	AVE/KM (elas=7.39)
between (0,5] km	0.322	0.9%	between (0,5] km	-2.322***	6.5%
	(0.274)			(0.402)	
between (5,10] km	-1.803***	2.5%	between (5,10] km	-1.041**	1.4%
	(0.342)			(0.424)	
between (10,15] km	-2.609***	2.4%	between (10,15] km	-2.139***	1.9%
	(0.329)			(0.485)	
between (15,20] km	-3.132***	2.1%	between (15,20] km	-2.890***	2.0%
	(0.332)			(0.648)	
between (20,30] km	-3.864***	1.8%	over 20 km	-3.949***	
	(0.310)			(0.926)	
between (30,40] km	-6.478***	2.2%			
	(0.575)				
between (40,50] km	-6.347***	1.7%			
	(0.529)				
between (50,75] km	-6.421***	0.9%			
	(0.394)				
over 75 km	-8.393***				
	(0.468)				

Notes: N = 515 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. Columns 1 and 3 are the estimates for trade costs bins for main and rural roads, respectively. Columns 2 and 4 report ad-valorem equivalent estimates of trade costs using use the preferred estimate of 7.9 for the substitution elasticity. Standard errors in parentheses. *, **, and *** indicate significance at 10%, 5%, and 1%.

Table 5. Counterfactuals

Panel A: Baseline Adoption, Expenditures, and Remoteness

	log(μ)		Δlog(E[F])	
	B_0	R	B_0	R
	-2.42***	-0.66***	-15.79***	-3.7378***
	(0.042)	(0.038)	(0.417)	(0.392)

Panel B: Effects of 50% lower trade costs

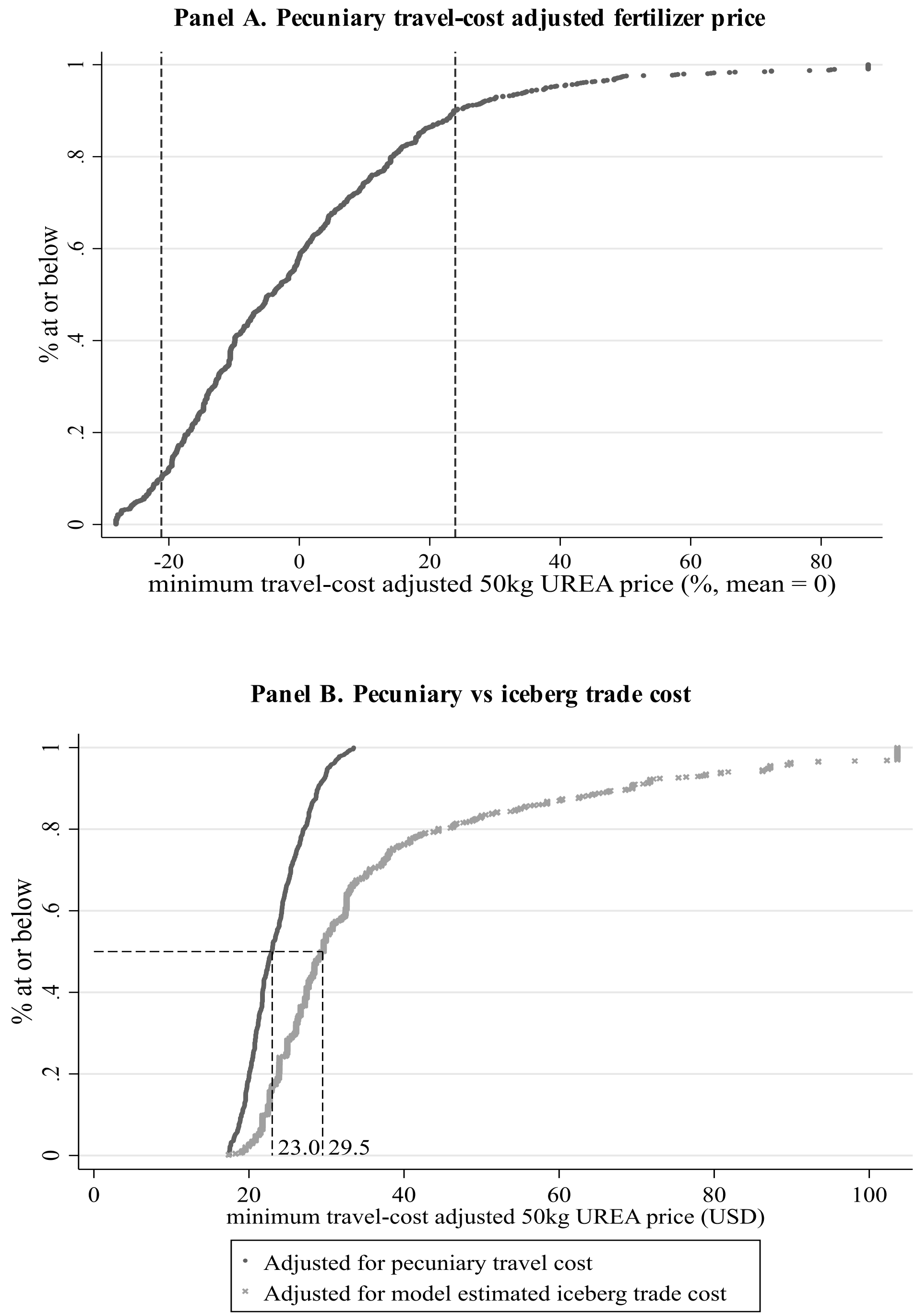
	Δlog(μ)		Δlog(E[F])		Δlog(Φ)		Δ-log(1+μ(ΔΦ-1))		%ΔW	ΔFS
	B_0	R	B_0	R	B_0	R	B_0	R		
All Roads	1.384***	0.414***	1.443***	0.392***	2.231***	0.078***	-0.863***	0.325***	23%	-0.075
	(0.029)	(0.029)	(0.028)	(0.028)	(0.028)	(0.028)	(0.024)	(0.024)		
Rural Roads	0.651***	0.121***	0.669***	0.113***	0.908***	0.01	-0.263***	0.118***	16%	-0.053
	(0.013)	(0.013)	(0.013)	(0.013)	(0.011)	(0.011)	(0.010)	(0.010)		
Main Roads	1.071***	0.352***	1.111***	0.334***	1.64***	0.079***	-0.586***	0.257***	10%	-0.042
	(0.028)	(0.028)	(0.027)	(0.028)	(0.030)	(0.030)	(0.020)	(0.020)		
From Distributor	0.142***	0.055***	0.144***	0.055***	0.163***	0.055***	-0.026***	0.005***	1%	-0.005
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.001)	(0.001)		

Panel C: Entry Profit

	log(Profitability)	
	B0	R
Baseline	-6.058***	-1.141***
	(0.054)	(0.054)
50% lower trade costs	-5.39***	-0.831***
	(0.042)	(0.037)

Panel A: Robust standard errors, clustered by village, in parentheses.
Panel B: Robust standard errors, clustered by village, in parentheses. %ΔW is the percent change in total surplus in the chemical fertilizer market, and ΔFS is the change in the share of surplus that accrues to farmers (0.8 at baseline).
Panel C: Robust Standard Errors, in parentheses.
*, **, and *** indicate significance at 10%, 5%, and 1%

Figure 1. CDF of travel-cost adjusted prices across villages



Notes: Each observation represents a village. Travel-cost adjusted prices are calculated through observed prices from a survey with fertilizer retailers and transport cost information collected from interviews with transport operators. In Panel A, the vertical dotted lines represent the 10th and 90th percentile. In Panel B, the vertical lines represent the median.