The price of remoteness: Product availability and local cost of living in Ethiopia*

Julien Martin[†1,2], Florian Mayneris[‡1,2], and Ewane Theophile[§1]

[1]Université du Québec, Montréal
[2]Centre for Economic Policy Research

December 4, 2024

Abstract

We use the microdata underlying the Ethiopian CPI to examine the spatial dispersion in local prices and availability of 401 items across 106 cities. Remote cities face higher prices and have access to fewer products. Large cities also face higher individual prices but enjoy access to a wider set of products. To assess the welfare implications of these patterns, we aggregate the data and build spatial cost-of-living indexes that account for both the price of available products and product availability. The cost of living is higher in remote and large cities. Moving from the first to the ninth decile in terms of remoteness (holding population size constant) results in an 8.3% increase in the cost of living. A comparable move in terms of population size (holding remoteness constant) leads to a 3.7% increase in the cost of living.

*We thank Simon Alder, Nate Baum-Snow, Kristian Behrens, Remi Jedwab, François Libois, Marion Mercier, Daniel Pereira Arellano, Tsur Sommerville, Will Strange, Gonzague Vannoorenberghe, and participants at various seminars and conferences for helpful discussions. We also thank the editor and the referees for their insightful comments.

†martin.julien@uqam.ca.

‡Corresponding author: mayneris.florian@uqam.ca.

§william.ewane@gmail.com.

1 Introduction

How do prices and the availability of products vary across cities in developing countries? This question is important because spatial inequality is pervasive within poor countries, and accounting for cost-of-living differences across locations is key for a comprehensive view of these patterns.[1] The literature in economic geography emphasizes city size as a key determinant not only of the dispersion in prices but also of product availability across cities (Handbury and Weinstein, 2015; Feenstra et al., 2020). A key insight from these papers is that product availability increases with city size and greatly affects cost-of-living measures across regions. While they focus on the US and China, product availability is also likely to matter in developing countries (see, e.g., FAO et al., 2018; WFP and CSA, 2018; FAO et al., 2019). The literature in development economics often examines price differences between urban and rural areas (see, e.g., Ravallion and Van De Walle, 1991; Deaton and Tarozzi, 2000; Muller, 2002), and highlights the role of transport costs in determining the price of available products in remote destinations (see, e.g., Atkin and Donaldson, 2015; Rancourt et al., 2014). Our goal is to bridge these two strands of the literature and analyze how both city size and remoteness shape the spatial dispersion in the level of prices and the availability of a large number of products that are representative of households' consumption in the context of a large developing country - Ethiopia.

We leverage the microdata underlying the Ethiopian CPI to examine the spatial dispersion in local prices and product availability across more than 100 cities. Our findings indicate that individual prices tend to increase with both city size and remoteness. Furthermore, product availability, defined as the probability of finding a product in a given city, increases with city size but decreases with the level of remoteness. These results remain robust even after controlling for potential confounding factors such as income per capita, the prevalence of home production, and the proximity to international trade corridors. They also resist an IV strategy and various sample restrictions. The analysis at the micro-level shows that overall, remoteness has a clear impact on the local cost of living by driving up prices and reducing product availability, while the advantage of living in a larger city is ambiguous.

To quantify the impact of these micro-patterns for local consumers, we aggregate the data and calculate local cost-of-living indexes that capture variations in both price levels and product availability. The elasticity of substitution is the key parameter governing the utility cost of missing products in these spatial price indexes: the lower this elasticity, the

[1] See Ferré et al. (2012) and Young (2013) for nominal consumption and nominal income differences across urban and rural areas, and Gollin et al. (2021) for spatial differences in terms of non-monetary amenities.

more costly missing products are. We calibrate the elasticity using sectoral estimates by Broda et al. (2017) for 10 African countries.

The analysis reveals that remote locations are unambiguously more expensive than central cities. Comparing cities at the first and ninth deciles in terms of remoteness, the price of available products is 5.2% higher in remote cities. Furthermore, remote cities suffer less available products. When both margins are accounted for, the local cost-of-living is 8.3% higher in remote cities.

The effect of city size on the local price index is more ambiguous and depends on the calibrated elasticity of substitution. Available products are 11% more expensive in large cities relative to small ones (still based on the interdecile comparison). However, more products are available in large cities. Depending on the elasticity of substitution between these products, the greater availability of products in large cities may compensate for the higher individual prices. Based on our calibration, the cost-of-living that accounts for both margins remains 3.7% higher in large cities. Note that the greater product availability in large cities exactly compensates for the higher price of available products for an elasticity of 3.

All in all, this quantification shows that when both the price of available products and product availability are accounted for, remoteness imposes a higher toll on the local cost-of-living than city-size.

Theoretical mechanisms. From a theoretical viewpoint, there are multiple mechanisms that may govern the relationship between product prices and product availability on the one hand, and city size and remoteness on the other. Regarding the relationship between city remoteness and prices, several factors are at play. First, more remote locations, almost by definition, suffer higher trade costs (including transport costs, time costs and/or administrative costs), which should result into higher prices with a magnitude that depends on the pass-through of trade costs to final prices (Martin, 2012). Second, according to the well-known Alchian-Allen effect, when trade costs are (at least partly) additive, only higher-quality varieties will make it into remote locations (see, e.g., Hummels and Skiba, 2004; Martin and Mayneris, 2015). Third, competition in remote locations may be lower and, in case of imperfect competition, this can lead to higher markups and thus higher prices (see, e.g., Asturias et al., 2019). Hence, it makes almost no doubt that prices should be higher in remote locations. Moreover, due to the higher trade costs associated with remoteness (whether due to the distance, the quality of transport infrastructure, or in a developing country context the safety of the environment), we also unambiguously expect that more remote locations suffer more unavailable products/varieties.

The relationship between city size and prices is more ambiguou. On the one hand, competition tends to increase with city size which should induce, when markups are variable, lower prices in larger cities (Melitz and Ottaviano, 2008; Feenstra et al., 2020). On the other hand, production and distribution costs may be higher in larger cities due to higher wages and higher real estate prices (see, e.g., Combes et al., 2008, 2019), leading to higher prices for consumers. The net effect of both mechanisms is indeterminate.

Regarding the relationship between city size and product availability, models featuring consumers' love for variety and size differences across regions predict that the number of available varieties increases with market size (Krugman, 1991; Ottaviano et al., 2002), which may lead to self-reinforcing agglomeration mechanisms. We thus unambiguously expect a negative correlation between city size and the probability of a product or a variety being unavailable.

We provide exploratory analyses on some of these mechanisms but leave a complete assessment for future research.

Related literature. We contribute to three strands of the literature. First, we contribute to the literature on spatial differences in terms of the cost of living in developing countries (see, e.g., Ravallion and Van De Walle, 1991; Deaton and Tarozzi, 2000; Muller, 2002; Timmins, 2006; Ferré et al., 2012). Our work offers three important contributions: (1) Our measure of the cost of living takes into account both the availability of products and services, as well as their prices when available, which distinguishes it from existing papers that typically focus solely on prices. Our results suggest that the cost of living is higher in large cities despite the presence of a broader set of available varieties. This may partly explain the higher wage premia in large urban cities discussed in the literature (Gollin et al., 2021). (2) By leveraging the micro-data underlying the consumer price index in Ethiopia, our analysis provides broader spatial and industrial coverage compared to existing studies that often focus on a narrower range of products and specific villages or regions. (3) Instead of emphasizing the urban-rural divide, we analyze differences in the cost of living across cities and examine their relationship with economic geography factors, namely, population size and geographic remoteness of cities. Whereas we focus on differences in cost-of-living across space within a country, a related literature investigates how cost-of-living varies across income groups (see, e.g., Hottman and Monarch, 2020; Faber and Fally, 2022; Ma et al., 2024).

Second, a couple of recent papers examine prices and product availability in the context of Ethiopia (Gunning et al., 2024; Krishnan and Zhang, 2020). Gunning et al. (2024)

demonstrate that households in remote villages of Ethiopia have access to a lower variety of goods. Our work differs from these other papers along two important dimensions. First, the focus is different. These papers examine individual prices and product availability in remote villages. Our analysis instead covers Ethiopian main cities. Second, our data allow us to compute local price indexes and thus to directly compare the welfare across cities - and relate welfare differences to city size and remoteness.[2]

Third, our paper contributes to the literature on the measurement and determinants of price indexes in the presence of missing products. Seminal papers in this literature have developed methods to measure the costs and benefits of disappearing and appearing varieties over time (Feenstra, 1994; Hausman, 1996; Diewert and Feenstra, 2019). A few papers apply this method to a spatial context, and explore the link between the cost of living and city size (see, e.g., Handbury and Weinstein, 2015; Feenstra et al., 2020, using U.S. and Chinese data, respectively).[3] We complement this literature by applying the method to a developing country for the first time and showing that, in the case of Ethiopia, remoteness has a stronger impact on spatial variations in terms of cost of living than city size.

Three papers are particularly relevant to our analysis. Atkin and Donaldson (2015) use the same source of data as ours to estimate the size of intra-national trade costs and their pass-through in the price paid by final consumers. To implement their methodology, they have to focus on 15 products (out of 400) that are as precise as barcode categories and for which they can identify, based on interviews, the location of production. They find that internal trade costs are high in Ethiopia, and that falling trade barriers only modestly benefit final consumers as most of the gains are captured by intermediaries. Handbury and Weinstein (2015) use barcode data for food products in the U.S. and compute city-level measures of cost-of-living that account for both the availability and the price of varieties. They find that food products are more expensive in large cities because consumers there buy more expensive varieties in more expensive retail stores. On the other hand, more varieties are available in large cities. When both product availability and store heterogeneity are accounted for, large U.S. cities exhibit a lower price index for food products. Focusing on 19 grocery barcode products, Feenstra et al. (2020) show that in China, more varieties are available in large cities (as in the U.S.) and sold at a lower price (contrary to the U.S.). They attribute this latter pro-competitive effect to a more uneven distribution of manu-

[2]The literature also examines the impact of remoteness on outcomes other than prices or product availability. For instance, Dercon and Hoddinott (2005) show that better access to market towns allows rural households to buy their inputs at a lower price and to sell their outputs at a higher price in Ethiopia; in the same vein, Aggarwal et al. (2022) find poor market access implies a poor harvest output in rural Tanzania.

[3]See also Matsa (2011) on the link between competition and inventory shortfalls.

facturers' sales and retailers across space compared to the U.S. The cost-of-living is then unambiguously lower in large cities in China.

Our work differs from these three papers in two main aspects. Firstly, we examine the impact of city size and remoteness on the availability and price of products across space. To our knowledge, we are the first to consider all these dimensions simultaneously. Secondly, we rely on the price data used for the construction of the Ethiopian CPI rather than barcode data. The main advantage of such data is that it allows us to study a basket of goods and services representative of the consumption of households. The limit is that we compare the price of similar but not necessarily *identical* products. Barcode data instead allow for the comparison of identical products but for a limited set of food (and dry grocery) products. Barcode data are not available in the context of developing countries like Ethiopia. Put differently, our approach compares, across cities, the typical prices faced by a consumer that wants to buy a given product, while approaches relying on barcode data compare the prices faced by a consumer willing to buy the exact same variety within a given retail chain. These two approaches correspond to different thought experiments.

The rest of the paper is organized as follows. Section 2 presents the data, section 3 discusses the results at the level of individual products, and section 4 proposes an aggregation procedure to analyze spatial differences in terms of cost-of-living indexes. Finally, section 5 concludes.

2 Data

Prices and product availability. The price data we use are those collected by the Ethiopian Central Statistical Agency (CSA) to construct the national Consumer Price Index. A detailed description of the data extraction and treatment procedure is provided in the Online appendix. Prices are collected on a monthly basis by enumerators in 117 markets. In the original dataset, Addis-Ababa is divided into 12 markets. To consolidate the data, we merge these markets into a single entity by considering the median price across the 12 markets, resulting in a total of 106 markets. Within each market, enumerators survey a predetermined sample of outlets every day during the first two weeks of each month. These outlets encompass a representative selection of open markets, kiosks, groceries, butcheries, pharmacies, supermarkets, and other similar establishments. Enumerators are instructed to locate specific products and report them as missing if they are unable to find them. When a product is found, enumerators ascertain its typical price (after bargaining) by conducting interviews with both sellers and consumers (Atkin and Donaldson, 2015).

The survey encompasses 427 products and services, which can be categorized into 12 major groups and 55 categories. These items include food products like bread and cereals, as well as clothing and footwear, household equipment, and services such as haircuts and restaurants. The product descriptions range from specific branded items (e.g., "Coca-Cola bottle 300c") to more detailed products without a specific brand (e.g., "bed sheet (Patterned Kombolcha) 1.90m x 2.50m") and even generic product categories (e.g., "sorghum yellow, kg" or "rice imported, kg"). The index covers all type of expenditures including those related to housing such as stone and sand for construction or concrete blocks. However, there is no information on rents in our dataset. Another dataset provided by the Ethiopian Statistical Agency reports that rents account for only 3.3% of expenditures across Ethiopian regions (the maximum being 10% in Addis-Ababa). We are thus confident we cover the lion share of consumers' expenditures with the data we use.

To focus on location-based price differences rather than changes over time, and to address potential issues related to misreporting, we use the monthly data for the year 2015 and calculate the median price per product and location based on the months when the product is available.[4] In addition to price information, the survey allows us to identify unavailable products. We consider a product to be unavailable in a given month if the price is missing (Atkin and Donaldson, 2015, do the same for example). Then we require the product to be missing every month throughout 2015 to be classified as unavailable in that year. After excluding products reported in less than 10 cities, the final dataset contains information on 401 products across 106 cities, i.e. 42,506 product-market pairs, out of which approximately 34% are categorized as missing. Missing is more prevalent for some services related to education (45%), communication (49%), and for transport including airplane, bus etc. (62%). Two categories stand out with very few missing observations: health including drugs (15%) and restaurants, bars, and hotels (8%).

Because inflation is a significant political issue in Ethiopia, there may be concerns about potential manipulation of price-quote data for political reasons. However, it is worth noting that 2015 was not a year of hyperinflation, which limits the political motivations for price manipulation. To further evaluate the reliability of the price quotes, we test whether the data adhere to Benford's Law. Researchers have used deviations from this law to identify reporting issues in survey data (e.g., Judge and Schechter, 2009; Demir and Javorcik, 2020). The underlying concept of this test is that manipulating the data while still conforming to the Benford distribution of the first digit of numerical data is difficult. Figure OA.2 in Appendix presents the frequency distribution of the first digit of price quotes in our data

[4] The correlation between this measure of prices and the average or the mode is above 99%.

alongside the expected distribution according to Benford's Law. The observed frequencies align well with the frequencies predicted by the Benford law.[5] Hence, the distribution of the first digit of price quotes in the data is consistent with Benford's Law, indicating that price manipulation is not a major concern in this context.

Size and remoteness. Our objective is to assess the impact of city size and remoteness on the spatial dispersion of the cost of living. City size is measured by population, and the population data used are based on the 2007 population and housing census of Ethiopia, with population projection figures for 2015 provided by the Ethiopian Central Statistical Agency (CSA) (see CSA, 2013). However, for 25% of the urban centers, the CSA does not offer any projections. In such cases, we rely on the projection figures for 2015 provided by the Ethiopian Ministry of Water and Energy in 2011, as part of the urban water-supply universal access plan (see Ministry of Water and Energy, 2011). The city sizes in the dataset range from 764 inhabitants (Deri) to 3,273,000 (Addis-Ababa). Among the cities in the sample, 63% have a population of fewer than 30,000 inhabitants, and these cities are classified as "rural".

One might wonder how large or remote cities differ from other locations. First, the correlation between city size and remoteness is negative but not significantly different from zero in the data. Similarly, ethnic diversity is not correlated with either city size or remoteness. Both large and remote cities tend to have a dominant ethnicity that is significantly less connected to the ethnicity of the prime minister. Last, remote regions are far from Addis, while cities nearby international corridors tend to be bigger in terms of population (see Table OA.2 of the Online appendix).

The remoteness of a city is calculated as the average travel time to other Ethiopian cities, given by the formula:

$$\text{remote}_c = \frac{1}{105} \sum_{j \neq c} \text{travel time}_{cj},$$

where remote_c represents the remoteness index for city c, and travel time_{cj} denotes the travel time between cities c and j. The travel time between Ethiopian cities is computed using the Stata package Georoute (Weber and Péclat, 2016), which utilizes Google Maps data to compute travel times by road.[6] In robustness checks, we consider alternative measures of

[5]We also conducted the Kuiper test to assess deviations of the data from Benford's Law. The mean Kuiper statistic is 0.0264, which exceeds the threshold above which we can reject with a 99% confidence the null hypothesis that the observed distribution deviates from Benford's Law.

[6]The travel times to other cities provided by this package are missing for 10 cities. For these cities, we estimate travel time based on bilateral distance. Distance is computed using the Stata package Geodist. Travel

remoteness such as the population-weighted average travel time to other cities, travel time to the capital city (Addis-Ababa), and to the main international trade corridor (Kombolcha, which serves as a transit point for shipments to and from Djibouti).

Other data sources. The baseline specification links observed prices and product availability to city size and remoteness. We check the robustness of estimated relationships to possible confounding factors including income per capita, ethnic diversity, and home production. The construction of price indices further necessitates information on expenditure weights. The construction of each of these variables is described in the Online appendix.

3 Spatial dispersion of products' prices and availability

This section examines the link between the price and availability of *individual* items and the size and remoteness of cities.

3.1 Empirical approach

Our baseline specification is as follows :

$$y_{pc} = \alpha \text{pop}_c + \beta \text{remote}_c + \omega_p + \epsilon_{pc}, \quad (1)$$

where y_{pc} is either the log median price quote of product p in city c in 2015 or a dummy equal to 1 when p is missing every month of 2015 in city c. City-size is measured by the log population in c. The measure of remoteness is the log average time to reach all the other Ethiopian cities in the database from c by road. Variable ω_p represents product fixed effects that purge the left hand side variables from product characteristics. Product fixed effects are important as more than 95% of the dispersion in prices and 35% of the dispersion in availability is product-specific. Indeed, the per-unit cost varies a lot across products, and as already mentioned in the data section, the frequency of missing observations is also quite heterogeneous across products. Last, ϵ_{pc} is the error term. Since the level of prices exhibits significant spatial autocorrelation as measured by the Moran statistic[7], we allow for a possible correlation in the error terms within a 50km radius around cities using the

time is regressed on a polynomial of degree 7 of distance, along with fixed effects for origin-destination pairs of regions. For the 8,730 city pairs for which we have both forms of information, the R^2 is approximately 90%. Travel time is then predicted for all city pairs involving the 10 destinations for which information on travel time is missing. Travel times are those reported by GEOROUTE in 2018.

[7] But not product availability, results available upon request.

HAC method proposed by Conley (1999) and the Stata package developed by Colella et al. (2019).

The purpose of this regression analysis on individual data is mainly descriptive, as the exact value of the coefficients we obtain on the various covariates has no importance for the analysis of city-level price indices we conduct in the last part of the paper. We nonetheless check the robustness of our results to two econometric concerns: endogeneity of remoteness and selection bias. Endogeneity may arise in our specification if factors influencing travel time between cities also influence the level of prices and the availability of products. For example, local economic conditions, local conflicts, or connections between local politicians and the central government may all affect the quantity and the quality of road infrastructure on the one hand, and products' prices and availability on the other. To tackle this issue, we first include additional controls such as, among others, local ethnic diversity, distance to trade corridors, and home production. We then propose an IV strategy where remoteness is instrumented by the average bilateral geodesic distance to other cities.

Selection bias may arise in our price regression. Indeed, we do not observe the price when a product is missing, which may imply that the sample of cities for which prices are observed is not representative. For instance, if some observations are missing because a product becomes prohibitively expensive due to remoteness or city size, we may underestimate the impact of these variables if we don't account for selection bias. We develop two strategies to tackle this issue. First, we check the robustness of our results on a sample of products that are available across almost all location. Second, we implement a Heckman model of sample selection.

3.2 Baseline results

Columns (1) to (4) of Table 1 present the estimates of the relationship between prices and remoteness and city size. As can be seen from column (1), remoteness and city size are significantly related to the local prices of available products. Large and more remote cities are more expensive than the others. Individual prices are 11.6% higher in a city of 159,300 inhabitants (p90 in the sample) than in a city with 8,685 inhabitants (p10 in the sample). Individual prices are 5.9% higher in remote cities than in more central ones (again p90 *vs* p10 of the remoteness index in the sample). These results are robust to controlling for per capita income (column (2)), measuring remoteness using a population-weighted average of distance to other cities (column (3), p-value on population-weighted remoteness equal to 11%), and excluding Addis-Ababa from the sample (column (4)).

Columns (5) to (8) display the results on product availability. From column (5), we see that the probability that a product is missing is significantly higher in remote cities and significantly lower in large cities. Still using the first and ninth deciles of the distribution in terms of population size and remoteness, the probability that a product is unavailable is 21 percentage points higher in small cities than in big cities, and 5.2 percentage points higher in remote cities than in central ones. The results are overall robust to controlling for income per capita (column (6), p-value on remoteness equal to 12% in this case) and excluding Addis-Ababa (column (8)). The impact of remoteness remains positive but not statistically significant with an alternative population-weighted measure of distance to other cities (column (7)). This suggests that the distance to other markets, independent of their size, is the primary impediment to product availability in a city.

Table 1: ***Price and availability of products across Ethiopia***

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Price of individual product (log)				*Product not available (dummy)*			
Ln Remoteness	0.119[a]	0.112[a]		0.122[a]	0.105[c]	0.100		0.116[c]
	(0.032)	(0.036)		(0.034)	(0.061)	(0.065)		(0.061)
Ln Population	0.040[a]	0.044[a]	0.039[a]	0.038[a]	-0.072[a]	-0.068[a]	-0.072[a]	-0.075[a]
	(0.004)	(0.005)	(0.004)	(0.005)	(0.007)	(0.009)	(0.007)	(0.008)
Ln per cap. inc.		-0.025				-0.022		
		(0.022)				(0.032)		
Ln pop-weighted			0.041				0.058	
remoteness			(0.026)				(0.044)	
Addis	Yes	Yes	No	Yes	Yes	Yes	Yes	No
Observations	28,183	28,183	28,183	27,810	42,506	42,506	42,506	42,105
R-squared	0.03	0.03	0.02	0.03	0.06	0.06	0.06	0.06

Notes: The dependent variables are the log price (columns (1) to (5)) and a dummy equal to 1 if the product is unavailable (columns (6) to (10)). These dependent variables are defined in the $product \times city$ dimension. Standard errors account for spatial autocorrelation within a 50km radius around cities using the spatial HAC procedure. [a], [b] and [c] respectively, denote significance at the 1, 5, and 10% levels.

3.3 Sensitivity analysis

Sectoral and temporal heterogeneity. We run the benchmark regressions separately for the 37 product categories used later in the paper to build local price indexes. Our baseline results do not change much across product categories from a qualitative viewpoint, even though the value of the coefficients does vary. The coefficients are reported in Table OA.3 of the Online appendix.[8]

Note that we have also checked that our results are stable over years and across seasons (results are available upon request). We find that the effects of city size and remoteness on prices and product availability are stable if we consider prices for 2011 to 2015 (the years for which we manage to collect the data).[9] We also used our monthly data in 2015 to examine how our estimates vary across seasons. We considered three seasons: a dry season (Bega) from October to January; a short rainy season (Belg) from February to May; and the main rainy season (Kiremt) from June to September. We do not find a significant heterogeneity across seasons except in the case of agricultural products for which the effect of remoteness on prices shrinks during the harvest season (Belg). One possible interpretation is that the supply of locally produced goods in remote areas reduces reliance on external supplies, leading to a temporary convergence in prices between remote and more accessible regions.

Additional controls Table 2 examines the sensitivity of the results to the inclusion of several additional controls. Remoteness might be more related to the distance to the capital city, Addis-Ababa, which happens to be close to the geographic centre of the country too. Moreover, Ethiopia is a landlocked country and Atkin and Donaldson (2015) argue that 90% of trade flows enter Ethiopia via Djibouti through a commercial corridor, whose main Ethiopian city is Kombolcha. We thus control for distance to these two cities in columns (1) and (3). None of these variables is significantly related to the price and availability of products in our benchmark sample.

[8] A recent literature has examined the role of rationing in developing countries (Gadenne, 2020). Rationing involves setting a quota on consumed quantities while imposing a fixed price. This reduces the dispersion across cities in both the availability and prices of products. If anything, this should diminish the coefficients on remoteness and city size. In Ethiopia, less than 15% of cereal production is subject to rationing. When estimating the equations for cereals only, we find that the relationship between prices and remoteness is not significant, but the relationship with city size is. The relationship between product availability and remoteness and population size is similar to that observed for other products. The overall stability of the coefficients, combined with the fact that rationing affects only a fraction of a specific product, suggests that our results are not primarily driven by this policy.

[9] We do not exploit the panel dimension because our measures of remoteness and city size cannot be computed on a yearly basis.

Table 2: ***Sensitivity analysis***

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	price (log)		missing		price (log)		missing		price (log)	missing
Ln Remoteness		0.079^b		0.114^c	0.087^b	0.082^b	0.175^a	0.176^a	0.167^a	0.128^c
		(0.033)		(0.062)	(0.040)	(0.039)	(0.042)	(0.042)	(0.033)	(0.073)
Ln Population	0.039^a	0.035^a	-0.068^a	-0.071^a	0.046^a	0.038^a	-0.047^a	-0.044^a	0.040^a	-0.072^a
	(0.005)	(0.004)	(0.007)	(0.007)	(0.006)	(0.007)	(0.005)	(0.006)	(0.004)	(0.007)
Ln Travel time to Addis	0.009		0.005							
	(0.006)		(0.010)							
Ln Travel time to import corridor	-0.010		0.017							
	(0.008)		(0.011)							
Dominant ethnicity politically connected		-0.051^a		0.012						
		(0.013)		(0.021)						
Log Ethnic diversity index		0.006		0.004						
		(0.015)		(0.030)						
Home production						-0.089^a		0.023		
						(0.032)		(0.023)		
Observations	27,697	28,183	41,704	42,506	8,672	8,672	12,370	12,370	28,183	42,506
R-squared	0.03	0.03	0.03	0.06	0.04	0.04	0.04	0.04	0.03	0.06
Kleinbergen-Paap F-stat.									627.33	663.58

Notes: Dependent variables are defined in the $product \times city$ dimension. The variable "Dominant ethnicity politically connected" is a dummy equal to one when the dominant ethnic group in the city is the same as the Prime Minister's one. The index of ethnic diversity is the inverse of a Herfindahl index based on the share of the various ethnic groups in the population. Home production measures the share of auto-consumption in total households' consumption by region and type of area (rural or urban). Standard errors account for spatial autocorrelation within a 50km radius around cities using the spatial HAC procedure. ^{a}, ^{b} and ^{c} respectively, denote significance at the 1, 5, and 10% levels. Columns (9) and (10) are the outcomes of IV regressions where Ln Remoteness (which measures the average travel time by road to all the other cities) is instrumented by Ln Average geodesic distance to all the other cities.

Also, cities where the dominant ethnic group is the same as the Prime Minister's one could benefit from lower prices and greater product availability thanks to political connections and favoritism. Another issue with ethnicity is that if ethnic groups have very specific tastes, more ethnically diverse cities could be cities where products are more likely to be available (e.g., Schiff, 2015), and this diversity could also affect the price at which they are sold. We thus control for a dummy equal to one when the dominant ethnic group in the city is the same as the Prime Minister's one and for the inverse of a Herfindahl index based on the share of the various ethnic groups in the population in columns (2) and (4). Cities where the dominant ethic group is the same as the Prime Minister's one enjoy lower prices but not greater product availability. Ethnic diversity has no significant relationship with the price and availability of products. Both remoteness and city size remain significantly related to prices and product availability.

Home production. In the context of a developing country such as Ethiopia, another important issue that could undermine the benchmark results is home production. The phenomenon is particularly important in rural areas. If products are produced directly by those who consume them, they might be unavailable on the market without necessarily being unavailable for consumption. For a subset of food products in the database, we have information on the share of auto-consumption in total households' consumption by region and type of area (rural or urban). After reproducing in columns (5) and (7) the benchmark results for the subset of observations for which home-production is available, we directly introduce home-production in the regression. Column (6) shows that home-production pushes prices downward, which is consistent with the view that it increases competition. However, we do not find a significant relationship between home-production and product availability. Importantly, the sign, magnitude, and significance of the coefficients on remoteness and population size are not affected by the introduction of the home-production variable.

Endogeneity. There may be other confounders (e.g. conflicts, political connections) that affect both remoteness and the dependent variables, and that are not accounted for by the previous exercises. To push further the causal analysis of remoteness, we instrument it by the average geodesic distance to all the other cities, capturing the effect of remoteness related to the relative physical location of cities, and not to other geo-political or socio-economic factors. The results displayed in columns (9) and (10) of Table 2 show that if anything, the coefficient on remoteness is boosted (and the one on population remains unaffected).

Sample selection bias. We treat potential sample selection issues in Table 3. To deal with sample selection, we first replicate the benchmark regression on the subsamples of products for which the share of missing product $\times$ city cells is alternatively below 5%, 3% and 1%. We find that the coefficient on population is very stable across specifications. The coefficient on remoteness increases when we focus on samples of products which are less likely to be missing, which is coherent with the idea that products tend to be missing in remote locations when they would become prohibitively expensive if they were available.

We also implement a Heckman sample selection procedure. In the absence of any credible excluded variable that would affect the availability of products, but not their prices, the procedure relies on the functional form only (the first stage being non-linear). Again the results are qualitatively similar but the coefficients on remoteness and population are higher with the sample correction (especially for city size).

Table 3: ***Price and availability of products across Ethiopia***

	Benchmark	$\leq$ 5% missing	$\leq$ 3% missing	$\leq$ 1% missing	Heckman
	(1)	(2)	(3)	(4)	(5)
Ln Remoteness	0.119^{a}	0.117^{a}	0.136^{a}	0.191^{a}	0.127^{b}
	(0.035)	(0.032)	(0.059)	(0.038)	(0.041)
Ln Population	0.040^{a}	0.024^{a}	0.022^{a}	0.030^{a}	0.080^{a}
	(0.004)	(0.004)	(0.004)	(0.005)	(0.009)
Observations	28,183	8,529	6,802	3,801	28,183

Notes: The dependent variables are the log price defined in the $product \times city$ dimension. Standard errors account for spatial autocorrelation within a 50km radius around cities using the spatial HAC procedure. ^{a}, ^{b} and ^{c} respectively, denote significance at the 1, 5, and 10% levels. In columns (2)-(4) we reduce the sample to products that are missing in less than 5%, 3%, and 1% of cities.

3.4 Mechanisms behind the price effects

Price differences across locations may reflect differences in transport costs, in quality, in distribution costs, or in markups. Whereas a quantitative assessment of these mechanisms is beyond the scope of the paper, we have performed a series of exploratory analyses to shed light on them. The precise description of these analyses appears in Online appendix. Here is a summary of our findings.

Prices and remoteness. First, we investigate whether the price premium paid in remote location reflects that remote locations are farther away from places where goods are produced. To do so, we focus on imported products for which we can assess the travel time between the main Ethiopian point of entry and cities. The results show that cities that are farther away from the point of entry (the trade corridor) exhibit higher prices, and that while the coefficient on remoteness slightly decreases, it remains positively and highly significantly related to the price of available varieties. This suggests that remoteness captures factors beyond travel time for shipping goods. One possible explanation is that there is less competition in remote cities, which pushes prices up. For a given travel time, transportation costs might also be higher if risk is higher on these roads due to conflicts for instance, or if there is less competition among carriers.

Prices and city size. Then, we explore different channels that may explain the price premium in large cities. A first explanation is that price differences reflects difference in terms of the quality of goods sold. We performed our analysis on a subsample of barcode products available in our dataset. In this (tiny) sample, the price premium vanishes, which suggests that part of the price premium in large cities may be explained by composition effects.

To make further progress on this issue, we also examine whether the elasticity of prices to population size varies with the length of the quality ladder of products. Specifically, we define products with large cross-city price variations as more likely to exhibit quality differences. We find that the elasticity of prices to city size is higher for products potentially affected by vertical differentiation. Together, the results suggest that quality variations across cities may partially—but not entirely—explain the observed correlation between city size and prices.

An alternative explanation is that distribution costs are higher in large cities, which increases the price paid by consumers.[10] We leverage our dataset to compute proxies for

[10]Feenstra et al. (2020) show that population size has an unambiguously negative effect on producers'

local unskilled labor costs and local real-estate costs that may both affect distribution costs. Once we add these two proxies in our baseline specification, the price premium related to city size drops by 40%. One interpretation of this result is that higher distribution costs are passed through to consumers, partially explaining why prices are generally higher in larger cities.!!

There is no smoking gun, but the analysis suggests that part of the price premium in remote locations is due to higher transport costs, and part of the price premium in large cities is driven by higher quality and higher distribution costs. However, individual prices collected by enumerators reflect, for a given product, the typical price of available varieties within a market. We believe these cross-city differences in prices are informative of spatial difference in terms of cost-of-living, irrespective of the mechanisms underlying them. This is why we keep using observed prices, and not prices purged of quality differences and/or distribution costs, in the rest of the analysis.

4 Consumer price index across Ethiopian cities

So far, the results unambiguously suggest that remote cities are more expensive than central ones due to higher prices and lower product availability. On the other hand, prices are higher in large cities but consumers have access to a wider range of products. To our knowledge, we are the first to document this tension between product availability and individual prices in the context of a developing country. In this section, we put more structure and compute a spatial price index to assess which force dominates.

4.1 Spatial CES price index

We follow Handbury and Weinstein (2015) and compute the spatial version of the price index proposed by Feenstra (1994). The computation of this index rests on the assumption that welfare across cities can be represented by a representative agent with CES utility. Under this assumption, the price index in market c (EPI_c) can be written as:

markups and prices if sunk costs and marginal costs are similar across cities. However, if sunk costs depend on local costs or market size, then prices are less negatively related to market size and could even be higher in larger markets (anti-competitive effect of market size). Unfortunately, with the data at hand, there is nothing we can do to deal with markups. One can also think that the marginal costs inclusive of distribution costs are higher in large cities due to local wages and rents. This could lead to higher prices in large cities, even if producers' quality and markups remain constant across markets.

$$EPI_c = \Pi_{g \in G} \left(SPI_{gc} \times VA_{gc} \right)^{w_{gc}} \tag{2}$$

$$\tag{3}$$

where SPI_{gc} and VA_{gc} are sub-indices capturing respectively the prices of available products and product availability for the products of product-category g in city c, and w_{gc} is the log-ideal Sato-Vartia weight (built from the share of product-category g in consumers's total expenditures). These sub-indices are defined as follows :

$$SPI_{gc} = \Pi_{j \in J_{gc}} \left(\frac{p_{jc}}{p_{jE}} \right)^{w_{jc}} \tag{4}$$

$$VA_{gc} = \left(\frac{\sum_{j \in J_{gc}} x_j}{\sum_{j \in J_g} x_j} \right)^{\frac{1}{1-\sigma_g}}, \tag{5}$$

with p_{jc} the individual price of good j in city c, p_{jE} the median price of good j across Ethiopian cities, J_{gc} the set of products in category g that are available in city c (which size is also noted J_{gc}), and w_{gc} and w_{jc} the log-ideal Sato-Vartia weights.[11] In the composite index VA_{gc}, x_j is the total expenditures for product j of the nationally representative consumer, so that $\left(\frac{\sum_{j \in J_{gc}} x_j}{\sum_{j \in J_g} x_j} \right)$ represents the share of the products of product-category g that are available in c in the representative consumer's overall consumption of product category g. Finally, σ_g is the substitution elasticity between the products of the product-category g (the higher it is, the more substitutable the products, the lower consumers' love for variety).

To summarize, the price index EPI_c is a weighted average of g-category sub-indices that have two components: (i) an intensive component SPI_c that is a standard price index that tracks the price gap across products in location c compared to a location of reference; (ii) and an extensive part VA_c that measures the utility cost of unavailable products in location c compared to the same reference location. Here, we assume the reference location is a fictitious city where all the products of all the product-categories are available at a price equal to the median price observed in 2015 across Ethiopian cities in the sample.

We have information on expenditures across regions for 55 categories of products. Within these product categories, we assume that consumption is equally split across products. The formula of SPI_{gc} is valid if the set of available products from product-category g in

[11]The Sato-Vartia weights are given by

$$w_{gc} = \frac{(s_{gc} - s_g)/(logs_{gc} - logs_g)}{\sum_g (s_{gc} - s_g)/(logs_{gc} - logs_g)} \quad w_{jc} = \frac{(s_{jc} - s_j)/(logs_{jc} - logs_j)}{\sum_{jinJ_g} (s_{jc} - s_j)/(logs_{jc} - logs_j)}$$

where s_{gc} is the weight of product category g in total expenditure of city c and s_g is the weight of product category g in national expenditures, whereas s_{jc} is the weight of product j within product category g.

city c is not empty. However, many cities in the sample have empty sets for some categories of products. We thus group these categories into 37 g groups. Details on these groups are provided in the Online appendix. This aggregation allows us to compute the formula for 78 out of 106 cities in the sample.[12]

A key term in the formula is the elasticity of substitution σ_g, which affects the exact price index EPI_c through its extensive part VA_c. If products are poor substitutes, then missing products are more costly for consumers. With the data at hand, we cannot directly compute σ_g. We thus use the elasticities estimated by Broda et al. (2017) based on international trade data at the 3-digit level of the Harmonized System (HS) nomenclature for 10 African countries (Algeria, Central African Republic, Egypt, Gabon, Madagascar, Malawi, Mauritius, Morocco, Togo, Tunisia). We manually build a correspondence between the 37 product groups of the Ethiopian price data and the HS 3-digit nomenclature, and we propose two calibrations for σ_g. For the first one, we use the median of the sectoral estimates for the 10 African countries available in the dataset of Broda et al. (2017). Since the estimates are based on trade in goods data, we have no information on the value of the substitution elasticity in services sectors (e.g. transportation, restaurants and cafes, personal care). For these sectors, we calibrate the elasticity to 3.37 (the median elasticity across goods estimated for African countries in Broda et al., 2017). In the end, across the Ethiopian product categories used to build the local prices indexes, the median and the mean of the calibrated substitution elasticity are respectively equal to 3.37 and 5.08. In an alternative calibration, we do the same, but we restrict the sample of countries to Madagascar and Togo, which are the closest to Ethiopia in terms of income per capita and linguistic proximity. In this case, the substitution elasticity for services is set to 3.99 (i.e. the median across HS 3-digit sectors observed for the two countries).

4.2 Results

We first propose a visual inspection of the relationship between the cost of living as measured by the spatial price index and its various components on the one hand, and city size and remoteness on the other. The graphs in Figure 1 plot the level of the various components of the price index against the degree of remoteness of the city, taking the calibration

[12] Cities for which the index cannot be properly computed have the same median level of remoteness than the rest of the sample, but their population is four time smaller (in median). We thus cannot compute the index for small cities with many missing products, but these cities only account for 8% of the population covered by our sample. Moreover, for the intensive part of the price index (SPI) that can be computed for all cities, we have checked that the coefficients on remoteness and city size are almost the same in the full sample (106 cities) and the constrained sample (78 cities).

of substitution elasticities based on the 10 African countries in Broda et al. (2017); the size of the circle is proportional to the population of the city it represents. Panel (a) relates to the intensive component of the CES-price index SPI_c, that is how expensive available products are compared to the median price quote observed in Ethiopia. Two main messages emerge: once expenditure shares are accounted for, large cities are still more expensive (big circles are at the top of the graph), and remote cities also still tend to exhibit higher prices (the slope of the scatter plot is positive). A notable exception is Addis-Ababa whose price level is as high as the more remote cities. Regarding the extensive component of the price index VA_c in panel (b), when expenditure shares and the elasticity of substitution are taken into account, large and less remote cities appear with a lower value of the extensive component of the price index, which is coherent with the fact that the probability that a product is available is higher in large and less remote cities. The combination of the two components gives us the exact price index EPI_c, which appears on panel (c). The relative effects of city size on the level of prices and on product availability vary with the value of σ. With the current calibration, the graph shows that the correlation of the exact price index with city size is positive, which means that the access to a wider range of products in large cities does not fully offset the price premium. Remote cities have a higher cost of living than central cities, which is intuitive since they suffer from both higher prices and lower product variety.

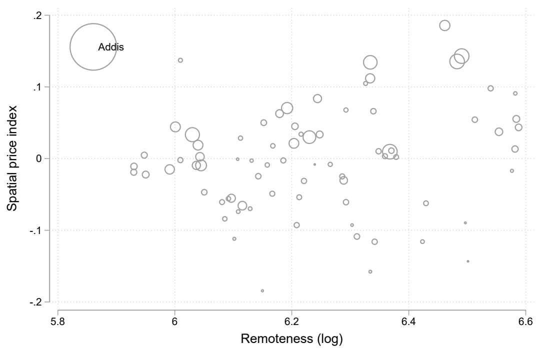
The graphical observations are largely confirmed by the econometric analysis reported in Table 4. The EPI is the product of the intensive and the extensive components, and thus log-linear in these two components.

In large cities, the intensive component (the weighted price index SPI_c, column (1) of Table 4) is higher, whereas the availability index (the weighted price index VA_c, columns (2) and (4) of Table 4) is lower, because more products are available. Which force dominates depends on the value of σ. Because consumers' valuation for variety is decreasing in σ, the intensive price channel dominates for higher values of σ. For the calibrated values of σ_g we use, we find that the price effect dominates the variety effect, which implies that, all else equal, the cost of living increases with city size. The positive relationship between city size and the cost of living is stronger when we use the calibrated elasticities based on the estimates for Togo and Madagascar. This reflects the fact that, as already mentioned, these calibrated elasticities are higher than those obtained when using the 10 African countries available in Broda et al. (2017). In unreported exercises, we find that the two effects exactly cancel out when σ_g is set to 3 for all product categories.

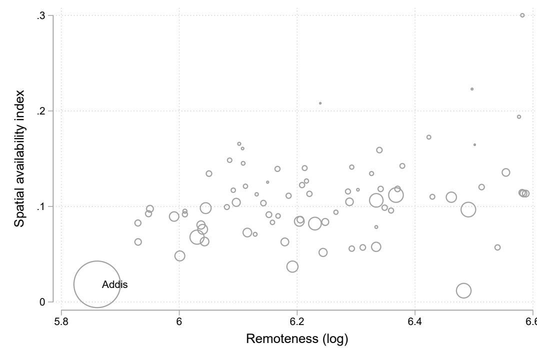
Note that if individual prices were not affected by city size (as we found for the barcode products), then the availability effect would unambiguously dominates, and the cost-of-living would always fall with city size.

On the other hand, remote locations exhibit both higher intensive and higher extensive components of the EPI, due to the detrimental effect of remoteness on both the price of available products and product availability. The cost-of-living is thus unambiguously higher in remote locations. The cost of remoteness decreases with σ though, since the higher the substitution elasticity, the lower the consumers' love for variety, and thus the less detrimental for consumers' welfare the unavailable products in remote locations.

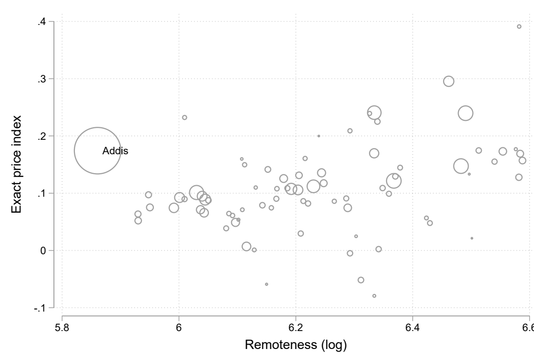
Figure 1: ***The geography of prices and product availability***



(a) ***SPI***



(b) ***VA***



(c) ***EPI***

Notes: Each dot is an Ethiopian city. The size of the circle is proportional to the city's population size. The log of the intensive part of the price index is in panel (a). The log of the extensive component is in panel (b). The exact price index in panel (c) is the sum of the indexes presented in panels (a) and (b). Indices presented in panels (b) and (c) are computed taking the calibration of substitution elasticities based on the 10 African countries in Broda et al. (2017).

Table 4: ***City-level regressions-local consumer price index***

	Ln SPI	Ln VA	Ln EPI	Ln VA	Ln EPI
		σ_g calibrated using the 10 African countries in Broda et al. (2017)		σ_g calibrated using Togo and Madagascar	
	(1)	(2)	(3)	(4)	(5)
Ln Remoteness	0.103^a	0.062^b	0.165^a	0.041^c	0.144^a
	(0.037)	(0.029)	(0.050)	(0.024)	(0.048)
Ln Population	0.042^a	-0.028^a	0.014^c	-0.020^a	0.022^a
	(0.006)	(0.004)	(0.009)	(0.003)	(0.008)
Observations	78	78	78	78	78
R-squared	0.39	0.48	0.16	0.42	0.18

Notes: The dependent variable is the log spatial price index (Ln SPI) in column (1), the log spatial availability index (Ln VA) in columns (2) and (4), and the log exact price index (Ln EPI=Ln SPI + Ln VA) in columns (3) and (5). Standard errors account for spatial autocorrelation within a 50km radius around cities using the spatial HAC procedure. ^{a}, ^{b}, and ^{c} respectively, denote significance at the 1%, 5%, and 10% levels.

Accounting for product availability quantitatively matters. We consider the first calibration of σ_g (10 African countries), and we compare two cities at the first and ninth deciles of the distribution in terms of population size and remoteness respectively. Based on column (1), which only accounts for the price of available products, the estimated coefficients imply that the cost-of-living is 11.0% higher in a large city, and 5.2% higher in a remote one. If we now take the coefficients in column (3) for the exact price index, which accounts for both the price of available products and for product availability, the cost-of-living is 3.7% higher in a large city, and 8.3% higher in a remote one. Hence, once product availability is accounted for, the toll imposed by remoteness on the local cost-of-living is higher than the one related to city size.

5 Conclusion

The strong dispersion in the cost of living across Ethiopian cities we document implies that nominal income should be deflated by a local price index to have a neat view of real income spatial heterogeneity. Importantly, such a local price index should account for product availability. In unreported investigations at the level of Ethiopian regions, we find that deflating nominal income can affect the ranking of regions, and that accounting or not for product availability in the deflator does matter too. Accounting for product availability could also matter for cross-country comparisons of cost-of-living. A more in-depth analysis of these issues is left for further research.

References

Aggarwal, Shilpa, Brian Giera, Dahyeon Jeong, Jonathan Robinson, and Alan Spearot, "Market Access, Trade Costs, and Technology Adoption: Evidence from Northern Tanzania," *The Review of Economics and Statistics*, 11 2022, pp. 1–45.

Asturias, Jose, Manuel García-Santana, and Roberto Ramos, "Competition and the Welfare Gains from Transportation Infrastructure: Evidence from the Golden Quadrilateral of India," *Journal of the European Economic Association*, 2019, *17* (6), 1881–1940.

Atkin, David and Dave Donaldson, "Who's Getting Globalized? The Size and Implications of Intra-national Trade Costs," Technical Report, National Bureau of Economic Research 2015.

Broda, Christian, Joshua Greenfield, and David E. Weinstein, "From groundnuts to globalization: A structural estimate of trade and growth," *Research in Economics*, 2017, *71* (4), 759 – 783. SPECIAL ISSUE IN HONOR OF 40 YEARS OF THE DIXIT-STIGLITZ MODEL.

Colella, Fabrizio, Rafael Lalive, Seyhun Orcan Sakalli, and Mathias Thoenig, "Inference with Arbitrary Clustering," IZA Discussion Papers 12584, Institute of Labor Economics (IZA) August 2019.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon, "Spatial wage disparities: Sorting matters!," *Journal of Urban Economics*, 2008, *63* (2), 723–742.

— , — , and — , "The Costs of Agglomeration: House and Land Prices in French Cities," *The Review of Economic Studies*, 2019, *86* (4), 1556–1589.

— , — , and — , "The Production Function for Housing: Evidence from France," *Journal of Political Economy*, 2021, *129* (10), 2766–2816.

Conley, T. G., "GMM estimation with cross sectional dependence," *Journal of Econometrics*, 1999, *92* (1), 1–45.

CSA, *The 2007 Population and Housing Census of Ethiopia*, Statistical Report at Country Level, 2008.

— , *Household consumption and expenditure (HCE) survey 2010/11*, Analytical report, October 2012.

— , *Population Projection of Ethiopia for All Regions At Wereda Level from 2014-2017,* August 2013.

— , *The 2015/16 Ethiopian Household Consumption-Expenditure(HCE) Survey, Results For Country Level*, Statistical Report, January 2018.

— , *Ethiopia - Socioeconomic Survey 2015-2016, Wave 3*, Report, April 2018.

Deaton, Angus and Alessandro Tarozzi, "Prices and poverty in India," *Princeton, July*, 2000.

Demir, Banu and Beata Javorcik, "Trade policy changes, tax evasion and Benford's law," *Journal of Development Economics*, 2020, *144* (C).

Dercon, Stefan and John Hoddinott, "Livelihoods, growth, and links to market towns in 15 Ethiopian villages," Technical Report 2005.

Diewert, W. Erwin and Robert C. Feenstra, "Estimating the Benefits of New Products: Some Approximations," in "Big Data for 21st Century Economic Statistics" NBER Chapters, National Bureau of Economic Research, Inc, March 2019.

Faber, Benjamin and Thibault Fally, "Firm Heterogeneity in Consumption Baskets: Evidence from Home and Store Scanner Data," *The Review of Economic Studies*, 2022, *89* (3), 1420–1459.

FAO, IFAD, WHO WFP, and UNICEF, "The State of Food Insecurity in the World 2018. Building climate resilience for food security and nutrition," 2018.

— , — , and — , "The State of Food Insecurity in the World 2019. Safeguarding against economic slowdowns and downturns," 2019.

Feenstra, Robert C, "New Product Varieties and the Measurement of International Prices," *American Economic Review*, March 1994, *84* (1), 157–177.

— , Mingzhi Xu, and Alexis Antoniades, "What is the Price of Tea in China? Goods Prices and Availability in Chinese Cities," *Economic Journal*, 2020, *130* (632), 2438–2467.

Ferré, Céline, Francisco H.G. Ferreira, and Peter Lanjouw, "Is There a Metropolitan Bias? The relationship between poverty and city size in a selection of developing countries," *World Bank Economic Review*, 2012, *26* (3), 351–382.

Gadenne, Lucie, "Can Rationing Increase Welfare? Theory and an Application to India's Ration Shop System," *American Economic Journal: Economic Policy*, November 2020, *12* (4), 144–177.

Gollin, Douglas, Martina Kirchberger, and David Lagakos, "Do urban wage premia reflect lower amenities? Evidence from Africa," *Journal of Urban Economics*, 2021, *121* (C).

Gunning, Jan Willem, Pramila Krishnan, and Andualem T. Mengistu, "Fading choice: transport costs and variety in consumer goods," *Economica*, 2024, *91* (363), 1100–1123.

Handbury, Jessie and David E Weinstein, "Goods prices and availability in cities," *The Review of Economic Studies*, 2015, *82* (1), 258–296.

Hausman, Jerry, "Valuation of New Goods under Perfect and Imperfect Competition," in "The Economics of New Goods," National Bureau of Economic Research, Inc, 1996, pp. 207–248.

Hottman, Colin J. and Ryan Monarch, "A matter of taste: Estimating import price inflation across U.S. income groups," *Journal of International Economics*, 2020, *127* (C).

Hummels, David and Alexandre Skiba, "Shipping the Good Apples Out? An Empirical Confirmation of the Alchian-Allen Conjecture," *Journal of Political Economy*, 2004, *112* (6), 1384–1402.

Judge, George and Laura Schechter, "Detecting Problems in Survey Data Using Benford's Law," *The Journal of Human Resources*, 2009, *44* (1), 1–24.

Krishnan, Pramila and Peng Zhang, "Restricting trade and reducing variety: Evidence from Ethiopia," *World Development*, 2020, *126*, 104695.

Krugman, Paul, "Increasing Returns and Economic Geography," *Journal of Political Economy*, June 1991, *99* (3), 483–499.

Ma, Hong, Luca Macedoni, Jingxin Ning, and Mingzhi Xu, "Tariffs Tax the Poor More: Evidence from Household Consumption During the US-China Trade War," Technical Report May 2024.

Martin, Julien, "Markups, quality, and transport costs," *European Economic Review*, 2012, *56* (4), 777–791.

_ and Florian Mayneris, "High-end variety exporters defying gravity: Micro facts and aggregate implications," *Journal of International Economics*, 2015, *96* (1), 55–71.

Matsa, David A, "Competition and product quality in the supermarket industry," *The Quarterly Journal of Economics*, 2011, *126* (3), 1539–1591.

Melitz, Marc J. and Gianmarco I. P. Ottaviano, "Market Size, Trade, and Productivity," *The Review of Economic Studies*, 2008, *75* (1), 295–316.

Ministry of Water and Energy, *Urban Water Supply Access Plan (Part III)*, The Federal Democratic Republic of Ethiopia, December 2011.

Muller, Christophe, "Prices and living standards: evidence for Rwanda," *Journal of Development Economics*, 2002, *68* (1), 187–203.

Ottaviano, Gianmarco, Takatoshi Tabuchi, and Jacques-FranÁois Thisse, "Agglomeration and Trade Revisited," *International Economic Review*, 2002, *43* (2), 409–436.

Rancourt, Marie-Ève, François Bellavance, and Jarrod Goentzel, "Market analysis and transportation procurement for food aid in Ethiopia," *Socio-Economic Planning Sciences*, 2014, *48* (3), 198–219.

Ravallion, Martin and Dominique Van De Walle, "Urban-rural cost-of-living differentials in a developing economy," *Journal of Urban Economics*, 1991, *29* (1), 113–127.

Schiff, Nathan, "Cities and product variety: evidence from restaurants," *Journal of Economic Geography*, 2015, *15* (6), 1085–1123.

Timmins, Christopher, "Estimating spatial differences in the Brazilian cost of living with household location choices," *Journal of Development Economics*, 2006, *80* (1), 59–83.

Weber, Sylvain and Martin Péclat, "GEOROUTE: Stata module to calculate travel distance and travel time between two addresses or two geographical points," Statistical Software Components, Boston College Department of Economics November 2016.

WFP and CSA, "Ethiopia-comprehensive food security and vulnerability analysis (CFSVA)," 2018.

Young, Alwyn, "Inequality, the Urban-Rural Gap, and Migration," *The Quarterly Journal of Economics*, 2013, *128* (4), 1727–1785.

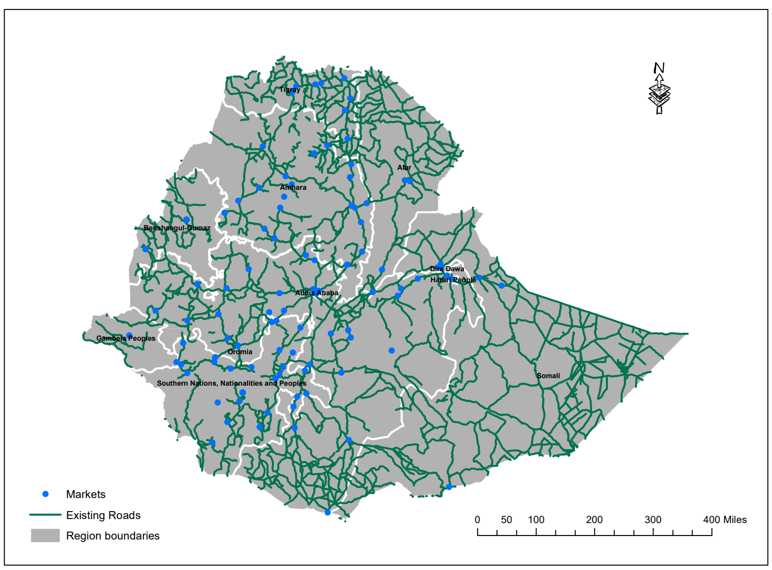
OA Online appendix

OA.1 Data and sample construction

In this appendix, we give additional details on the datasets and the construction of variables used in the empirical analysis.

Source. Individual price data are collected by the Ethiopian Central Statistical Agency (CSA) to build the national Consumer Price Index. We work with the data for the year 2015. The price quotes of more than 420 products are collected every month by enumerators in 117 markets. The markets are urban centers (cities or towns) in each woreda (a woreda being the smallest administrative division with local government in Ethiopia). Addis-Ababa, the Ethiopian capital city, is split into 12 markets. The different markets are mapped in Figure OA.1.

Figure OA.1: ***Markets covered by the Ethiopian CPI***



In each market, enumerators survey a pre-determined sample of outlets every day during the first two weeks of each month. Outlets comprise a representative sample of open markets, kiosks, groceries, butcheries, pharmacies, supermarkets, and so on. Enumerators are asked to find precisely defined products and to report the product as missing if they cannot find it. When they find the product, they determine its typical price (after bargaining) by interviewing both sellers and consumers (Atkin and Donaldson, 2015).

Extraction. Every month, the CSA releases a 200-page document reporting the price of every product in each of the 117 cities. We extract this information from pdf files for all the months of year 2015. We then manually check the obtained dataset. We change the label of some products or product categories that appear to be obviously wrong.

Description. The survey covers 427 products and services that can be grouped into 12 major groups and 55 categories (Tables OA.2 and OA.1). These products and services include food products such as bread and cereals, but also, among others, clothing and footwear products, household equipment products, or hair-cuts and restaurants. The product descriptions range from barcode-like data with brandname products ("Coca-Cola bottle 300c"), to very specific products without a brand (e.g., "bed sheet (Patterned Kombolcha) 1.90m x 2.50m"), to more generic product categories (e.g., "sorghum yellow, kg" or "rice imported, kg").

Outliers and yearly price. Prices are collected every month. In 2015, we thus have a maximum of 12 price quotes for a given product within a given location. We have detected a few outliers in the data. A visual check suggests two main types of outliers exist. First, the price might be abnormally high because, for instance, the marker for decimals was forgotten. Second, enumerators sometimes reported the monthly price of the product at the wrong line. For instance, in Akaki, the product "VCD-Player (Mayato Japan)" is missing every month of 2015 except the month of July. That month, the reported price is 5 Birrs. However, in the file, the product below "VCD-Player (Mayato Japan)" is "VCD Cassette rent". In Akaki, this product is available every month at the price of 5 Birrs, except for the month of July, where it is missing. Moreover, the median price of the "VCD-player (Mayato Japan)"' in markets where it is available is around 1,000 Birrs. Thus, an obvious mistake exists where the price of the "VCD Cassette rent" in Akaki was reported at the wrong line for the month of July.

To systematically detect and delete outliers, we drop monthly observations for which the price is 5 times higher or 5 times lower than the median price of the product across markets. We thereby exclude 469 observations out of more than 300,000. Note that with this procedure, we do not automatically create missing products, because, as explained below, our definition of "missing" is computed at the annual level.

Because we are interested in price differences across locations (and not over time), and to alleviate potential remaining issues related to misreporting, we take the median price per product and location computed over the months the product is not missing. In the original dataset, Addis-Ababa is split into 12 markets. We merge these markets into a single one by considering the median price across the 12 markets. After excluding products reported in less than 10 cities, the final dataset contains information on 401 products across 106 cities, i.e. 42,506 product-market pairs, out of which approximately 35% are categorized as missing.

Table OA.1: ***List of Product Categories***

Beverages - Alcoholic	Fuel And Power	Vegetables (Fresh)
Beverages - Non Alcoholic	Glass Ware	
Bread And Other Prepared Foods	Household Operation	
Bed Sheet (Non-Patterned Bahir Dar)	Jewellery	
Bed Sheet (Patterned Kombolcha)	Livestock	
Cereals Milled	Meat	
Cereals Unmilled	Medical Care	
Chairs, Tables, Etc.	Metal Ware	
Cigarettes	Miscellaneous Goods And Services	
Clothing	Oil Seeds	
Coffee, Tea, Chat And Buck-Thorn Leaves	Oils And Fats	
Communication	Other Food Items	
Construction Materials	Other Household Equipment	
Cost Of Milling	Other Medical Expenses	
Cost Of Tailoring	Personal Care	
Diary Products And Egg	Personal Effects	
Domestic Service	Plastic Ware (Local Made)	
Draught Animals	Potatoes, Other Tubers And Stems	
Earthen Ware	Prepared Cereal Products	
Education	Pulses Milled Or Split	
Equipment And Accessories	Pulses Unmilled	
Farm Equipment (Hand Made)	Ready-Made / For Adults	
Farm Equipment (Industrial Product)	Ready-Made / For Children	
Fish And Fish Products	Spices	
Food Taken Away From Home	Straw, Bamboo And Others	
Footwear (Men And Women)	Tobacco	
Fruits Fresh	Transport	

Income. To measure per-capita income, we use data from the 2015/2016 Ethiopian Household Consumption Expenditure survey (HCE) conducted by the CSA (see CSA, 2018a). Per-capita income is measured by region (Ethiopia has 11 regions in total), with two separate measures for urban and rural areas in each region. It is thus available at a more aggregate geographic level than prices. It measures people's total expenditures in Birr and it also includes consumption in kind. It is available by quintiles in each region and each type of area. The average income in the first quintile is 3,900 Birr, whereas it

Table OA.2: ***Product List***

'Katikalla'-Lt	Durrah-Kg	Internet Service-10hrs/month
'Tej' (Mead)-Lt	Hulled Barley-Kg	Mobile Apparatus(Nokia6200)
'Tella'-Lt	Maize (White)-Kg	Mobile call from Tel-Period
Araki (Local)-900cc	Oats -Kg	Telephone Charge (with town)-Period
Beer (Bedele)-330cc	Rice (Imported) -Kg	Telephone Charge to Addia Ababa -Period
Beer (Harar)-330cc	Sorghum Red-Kg	Telephone Line Installation Charge-Once
Beer (Meta Abo)-330cc	Sorghum White-Kg	Bricks (25cm x 12cm x 6cm)
Brandy (Local)-900cc	Sorghum Yellow-Kg	Cement/Bag/(Local)-50Kg
Cognac (Local)-900cc	Wheat Black (Red)-Kg	Chipwood (125cm x 250cm x 8mm)
Gin (Local)-900cc	Wheat Mixed-Kg	Coarse Aggregate Gravel-Meter cube
Saris Wine (Normal)-750cc	Wheat White-Kg	Corrugated Iron Sheet (.2mm)
Ambo Mineral Water-500cc	Book Shelves Wanza (3 Shelves) no Door No	Door made of iron
Coca Cola/Fanta-300cc	Chairs Wanza (Hand Made) Varnished	Floor Board 4m Length
Mineral Water-Lt	Chairs Wanza (Machine Made) Varnished. No	Gutter No 33 -Meter
Pepsi Cola/Mirinda-300cc	Chest of Drawer	Hollow Concrete Block(15x20x40 cm Cube)No
'Dabo' (Traditional Ambasha)-350gm	Cupboard Wanza (2 doors) Varnished	Iron Pipe 6mt. (1/2 inch Wide) Local
'Dabo' (Traditional Sheleto)-350gm	Double Bed Wanza (120cm) Varnished	Key (With hand)
'Enjera' ('Teff' Mixed)-325gm	Sofas (Complete)	Lime-Kg
Biscuits -150gm	Table Wanza (Hand made) Not Varnished. No	Lime/Jeso(Local) -Kg
Bread Wheat (Bakery) -350gm	Table Wanza (Machine made) Varnished	Mega Paints-4Kg
Bed sheet (Non-patterned Bahir Dar)-1.90m x 2.50m	Gissila-Packet	Nail (7cm - 12cm)-Kg
Bed sheet (Patterned Kombolcha)-1.90m x 2.50m	Marlboro-Packet	Nail With Cape-Kg
Bed Cover(Patterned Kombolcha)	Nyala -Packet	Nefas seleke Paints-4Kg
Blanket Woolen(Debre B.)-160cm x 220cm. No	Rothmans/England/ -Packet	Sand -Meter cube
Curtains-meter	Abujedid(Akaki/Bahir Dar)91cm. Meter	Stone for House Construction-Meter cube
Mattress-Sponge (A.A Foam) 120cm	Abujedid(Komb./Arba Min.)150cm Meter	Wall Paints-Super (Fluid) Normal -4Kg
Towel-Local (Kombolcha)	Cotton-Kg	Wall Paints-Super (Fluid) Plastic-4Kg
'Furno Duket' Locally Processed-Kg	Deriya-Meter	Water-Meter cube
'Teff' Black (Red)-Kg	Hisufi-Meter	Water Tanker,Roto (1 meter cube)
'Teff' Mixed-Kg	Jersi-Meter	Window Glass (50cm x 50cm x 3mm)
'Teff' White -Kg	Kashemire-Meter	Wood for House Construction('Atana')
Barley Mixed-Kg	Kefai-Meter	Yewellel Nitaf(Cement Made Tile)
Barley White-Kg	Khaki(Akaki) -Meter	Yewellel Nitaf(Plastic Made Tile)
Durrah-Kg	Khaki(S-10,000 Twil)150cm-Meter	Cereals-100Kg
Maize (White)-Kg	Nylon(Mojo)-Meter	Pepper Whole-100Kg
Oats-Kg	Polyster(Arba Minch/Awasa)-Meter	Pulses-100Kg
Sorghum-Kg	Poplin(Dire Dawa)105cm-Meter	Khaki/Teteron Suit (Boys)
Wheat Mixed-Kg	Poplin(Komb./Arba Minch)150cm-Meter. Meter	Khaki/Teteron Suit (Men)
Wheat White-Kg	Tetron(A.Minch)-Meter	Woolen Suit (Men)
'Teff' Black (Red)-Kg	Wool-England 100%-Meter	Camel Milk-Lt
'Teff' Mixed-Kg	'Chat'-Kg	Cheese Cottage-Kg
'Teff' White-Kg	Buck Thorn Leaves-Kg	Cow Milk (Unpasteurized)-Lt
African Millet-Kg	Coffee Beans-Kg	Cow Milk (pasteurized)-Lt
Barley Black-Kg	Coffee Leaves-Kg	Egg (Traditional)-Dozen
Barley Mixed-Kg	Coffee Whol-Kg	Goat Milk-Lt
Barley White-Kg	Malt-Barley-Kg	Powdered Milk (Me&My)-450gm
Barley for Beer-Kg	Malt-Wheat-Kg	Yoghurt (Traditional)-Lt

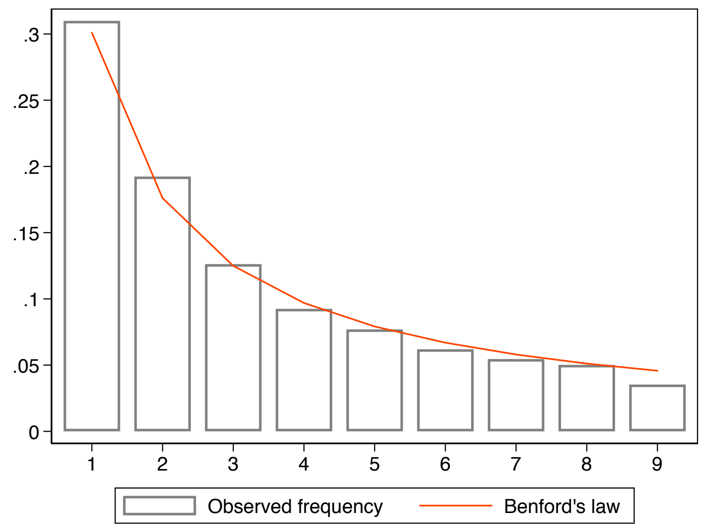
Table OA.2 (continued): **Product List**

Maid Servant-Month	Cup of Milk	Neck Laces (6gm 18 carat Local)
Salary for Guard-Month	Cup of Tea	Rings (4gm 18 carat Local)
Unskilled Service (Daily Laborer)-Day	Anbessa Leather Shoes Men(Local)Pair	Seiko-21 Jewels Automatic (Men)
Donkey	Boots for Men Plastic (Local)-Pair	Seiko-21 Jewels Automatic (Women)
Horse	Canvas Shoes (China) Men-Pair	Bull (2-4 Years)
Mule	Canvas Shoes (Local) Men-Pair	Cock (Indigenous)
'Jebena' Medium Size	Cost of Mending Shoes sole(Men)-Pair	Cow (4 Years and Above)
'Mitad' (Griddle of Clay)	Leather Shoes Men (Croft)-Pair	Goat (10-15Kg)
Cup for Coffee (China)	Plastic Shoes (Local) Women-Pair	Heifer (2-4 Years)
Local Stove '(Lakech)'	Plastic shoe(Childern)-Pair	Hen (Indigenous)
Plastic tile-Meter	Sendel Plastic Shoes (Imp.)-Pair	Ox (4 Years and Above)
Plate Clay (imported)	Shoe leather,Children(Imported)-Pair	Sheep (10-15Kg)
Water Pot	Shoe leather,Children(Local)-Pair	Beef-Kg
Ball Point-Bic England	Shoe leather,Women(China)-Pair	Camel Meat-Kg
Day School Fee-Private(Grade 9-10)-Month	Shoe leather,Women(Local)-Pair	Amoxacillin(500mg)-16 caps
Day School Fee-Public(Grade 7 & 8)-Month	Shoe sendel(Plastic) Childern-Pair	Ampicillin(250 mg) Local-56 caps
Exercise Book (50 Leaves) Local	Slippers Sponge Adult (China)-Pair	Asprin (300 mg) Local-20 pills
Night College Fee-Private-Credit/hr	Walking Shoes(Imp.) Non-Leather-Pair	Bactrim(480mg) Local -30 pills
Night School Fee-Government(Gr.9-10)-Month	Avocado-Kg	Chloramphenicol(250mg) Local-56 caps
Night School Fee-Private(Grade9-10)-Month	Banana-Kg	Cough Syrup (Efadykse) Local-125cc
Pencil (China)	Cactus-Kg	Fasider Table-1 pill
School Uniform Fee	Grapes-Kg	Insulin(Lente)-Buttle
Cassette Recorded Original(Local Music)No	Lemon -Kg	Magnesium Oxide -30 Pills
Expense for Photograph(Passport size)4Pho	Mango-Kg	Mezel(250mg) Local-30 caps
Newspaper (Addis Zemen)	Orange-Kg	Paracitamole(500mg) Local -20 Pills
Radio Set Philips 3 Band	Papaya-Kg	Penicillin injection(4 Mu. Local)-Buttle
T.V. Set Philips 21 inch(Colored)	Tangerine-Kg	Tetracycline (250 mg) Local-56 caps
Tape Recorder National (2 Speaker)	Buthane Gas (Shell)-12,5Kg	Vermox(100mg) Local-12 pills
Theater Enterance Fee-Once	Candles	Cooking Pan Medium (Local)
VCD Cassette rent	Charcoal-Kg	Electric 'Mitad' Aluminium
VCD-Player(Mayato Japan)	Diesel-Lt	Permuze(Japan)
'Digr'	Dung Cake-Kg	Refregrator
'Erfe '	Electric-Kwatt	Tray (Nickel) Medium N45 Local
'Kember'	Eveready Drycell	Charge for Money Transfer-Once
'Mofer'	Fire Wood-Meter cube	Coffin (Medium Quality)
Plough	Kerosine-Lt	Photocopy-Per page
Sickel	Matches-Box	Wedding Invitation Card-Per page
'Gejera'	Glass for Tea (Duralex)	Castor Beans-Kg
Pick Axe ('Doma')	Detergent (Omo)-50gm	Ground Nut Shelled-Kg
Sickel	Detergent(Zahira)-50gm	Linseed Red-Kg
Fish Fresh-Kg	Dry Cleaning (Suit Men)	Linseed White-Kg
Sardines (Imported)-125gm	Hard Soap (Imported)-200gm	Niger Seed-Kg
'Fasting Meal Without fish-One Meal	Hard Soap (Local)-200gm	Rape Seed-Kg
'Key Wot Yebeg/Yefyel'-One meal	Incense-Kg	Sesame Seed Red-Kg
'Key Wot Yebere'One meal	Sandal Wood	Sunflower-Kg
'Yebeg Kikil'-One meal	Toilet Paper (Mamko)-Roll	Butter Unrefined-Kg
'Yebeg Tibs'-One meal	Bracelet 20gm (18 carat Local)	Cooking Oil (Imported)-Lt
Cup of Coffee	Earrings (4gm 18 carat Local) -Pair	Cooking Oil (Local)-Lt

Table OA.2 (continued): **Product List**

Vegetable Butter (Sheno & Shady)-Kg	Lentils Split-Kg	Jeans trouser and Jacket
Canned Tomato (Local)-410gm	Mixed Pulses Milled-Kg	Kemise(for childeren)
Dry Yeast(Baking powder)-350gm	Peas Milled-Kg	Shirt Long Sleeved(Imported)Boys
Honey-Kg	Peas Split-Kg	Socks (Imported) Cotton-Pair
Salt-Kg	Peas Split(Roasted)-Kg	Sweater (England) for Girls
Sugar-Kg	Vetch Milled-Kg	Sweater (Local) for Boys
'Kuraz' Small Local Kerosine Lamp	Vetch Split(Roasted)-Kg	T-Shirt
Electric Bulb Philips(40/60 Watt)	Chick Peas-Kg	Basil Dry-Kg
Flash Light	Fenugreek(Green-Kg)	Black Cumin(Local)-Kg
Kerosine Lamp	Haricot Beans-Kg	Black Pepper(Local)-Kg
Bed Charge (Private-Per day	Horse Beans-Kg	Cardamon(Local)-Kg
Bed Charge (gov.)-Per day	Lentils-Kg	Chillies Whole-Kg
Doctor's Fee (Government)-Per visit	Lima Beans-Kg	Cinnamon(Imported)-Kg
Doctor's Fee (Private)-Per visit	Peas Green(dry)-Kg	Cloves(Imported)-Kg
Injection (Service Charge)-Once	Peas Mixed-Kg	Ginger Dry(Local)-Kg
X-Ray(For TB)-Once	Peas White-Kg	Ginger Wet(Local)-Kg
Barbery (Mens Hair Cut)	Soya Beans-Kg	Long Pepper(Local)-Kg
Blade-INDIA	Vetch-Kg	Pepper Whole-Kg
GIV Toilet Soap -90gm	'Gabi'	Tumeric Flour(Local)-Kg
Hair Dressing (Modern)	'Kemisna Netela'	White Cumin(Bishop's Weed) Local -Kg
Modes(Disposable napkins-Packet)	'Netela'	'Sefed'
Parafin Hair Oil -330cc	Geldem	Sack 100Kg Capacity
Perfume-100cc	Jeans Trouser	'Gaya'-Kg
Shaving Machine (medium)	Jogging Suit(sport tuta)	Air Plane (To Addis-Ababa) -Trip
Shoe Polish(Black/Brown)-Once	Khaki Jacket	Animal Transport fare-Trip
Zenith Hair Oil(Liquids Form)-330cc	Khaki Short	Benzene-Lt
Zenith Hair Oil(Non-Liquids Form)-330cc	Leather Jacket	Bus Fare (per km)
Belt (Local) Hand Made	Mekremia	Bus Fare (within Town)-Tarif
Belt (Local) Machine Made	Nylon Dress	Car Washing and Greasing -Trip
Hand Bag (Imported Synthetic)	Pants(for men)	Cart Fare-Trip
Umbrella-Men Medium (Local)	Polyester Suit	Motor Oil (Mobil)-Lt
Umbrella-Women Medium (Imported)	Polyester skirt	Taxi Fare-Trip
Bucket (20 Litres)	Shash (Imported)	Beet Root-Kg
Jerrycan (20 Litres)	Shirts Long Sleeved (Imported)	Cabbage-Kg
'Bula' -Kg	Shirts Long Sleeved (Local)	Carrot
'Kocho' (Unprocessed)-Kg	Shirts Short Sleeved (Imported)	Cauliflower-Kg
Potato-Kg	Singlets (Local) White	Ethiopian Kale-Kg
Sweet Potato-Kg	Socks (Cotton) Imported-Pair	Garlics-Kg
'Dube' Flour-Kg	Sweater (Local) Men	Green Peas-Kg
'Fafa' Flour-Kg	Sweater (Local) Women	Leaks-Kg
Macaroni (Local) Without Egg-Kg	Sweater-Men (Imported)	Lettuce-Kg
Pastini -Kg	Sweater-Women (Imported)	Onions-Kg
Spaghetti (Local) Without Egg-Kg	T-Shirts	Pepper Green-Kg
Chick Peas Milled-Kg	Tetron Trouser	Pumpkin-Kg
Chick Peas Split(Roasted)-Kg	Under Wear China	Spinach-Kg
Fenugreek Milled-Kg	Woolen Suit	Tomatoesv
Horse Beans Milled-Kg	Baby Cloths(Complete)	Tea Leaves(Local)-100gm
Horse Beans Split(Roasted)-Kg	Jeans Trouser	

Figure OA.2: ***Benford law for price quotes***



Each bar in the diagram represents the observed frequency of a number of the first digit of price quotes. The red curve represents the frequency of a price digit given by the Benford's law.

reaches 24,200 Birr for households in the fifth quintile. We use average income per capita as a measure of income in our analysis.

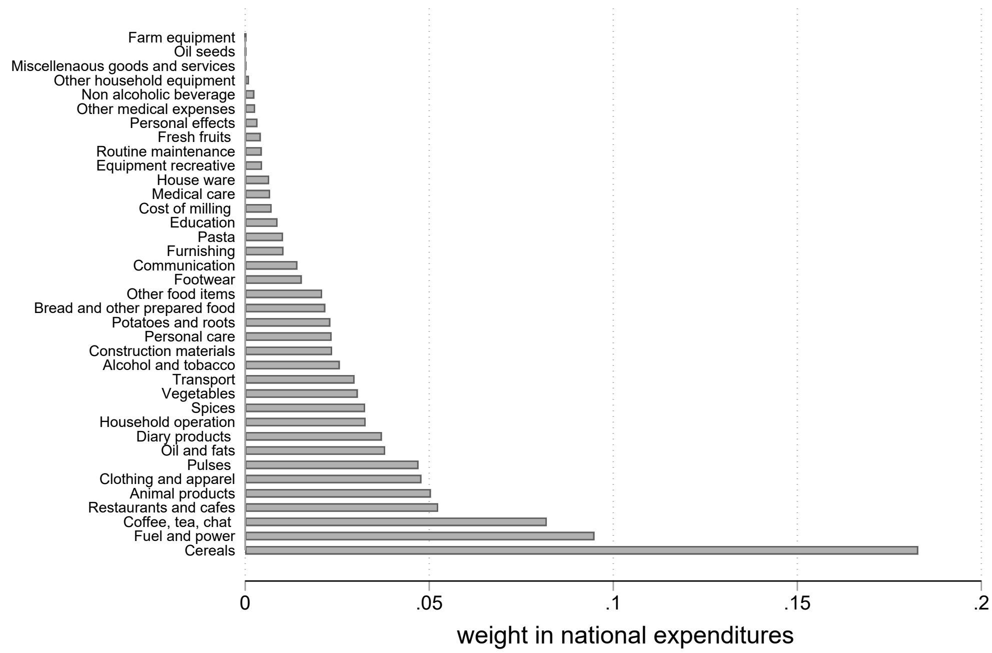
Ethnic diversity. Data on ethnic diversity comes from the 2007 Population and Housing Census of Ethiopia conducted by the CSA (see CSA, 2008). We compute the share of each ethnic group (93 in the country) in the total population of every woreda. Similar to Schiff (2015),we measure ethnic diversity as the inverse of a Herfindahl index computed with these shares. At the country-level, Oromos are the most represented ethnic group, with 34.6% of the total population.

Expenditure weights. Aggregate price indexes require information on expenditure weights across product categories. The price data do not include information on products weights. We thus rely on another dataset obtained from the Ethiopian Statistical Agency. From this dataset, we have information on the expenditure weights of 55 product categories covering the products and services in the Ethiopian CPI. The weights are specific to each Ethiopian region.[13] As explained in the core of the text, some of these categories are grouped together to be able to compute the ideal price index. We end up with 37 groups of products. Importantly, when several categories are aggregated within a group, we keep track of the weight of each of these category within the group. The groups and product

[13]An alternative is to build weights from the 2011 Ethiopian Household Consumption Expenditure survey (HCE) conducted by the CSA (see CSA, 2012). We have expenditure weights for 19 semi-aggregated categories. These categories are related to food and daily expenditures but there is no information on the expenditure weights of products that are purchased less frequently.

categories are presented in Table OA.3.

Figure OA.3: ***Weights in terms of expenditures***



Note: Each bar represents the share of an item in the national expenditures dedicated by a household for the selected products.

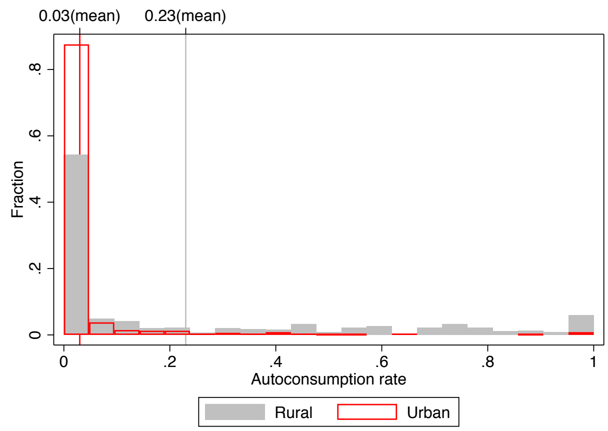
Home production. In some robustness checks, we account for auto-consumption, which could be correlated with both prices and product availability on the one hand, and city size and remoteness, which are our variables of interest, on the other. We compute data on auto-consumption using the third wave of the Ethiopian socioeconomic survey conducted for year 2016 by the CSA and the World Bank Living Standards Measurement Study-Integrated Surveys on Agriculture (LSMS-ISA) team (see CSA, 2018b). Obviously, auto-consumption is a concern for food products mainly, and information on auto-consumption is thus available only for these products. We define auto-consumption as the share of household consumption (in terms of quantity) that comes from its own production. These data are available for each product-region-area triplet, where area refers to urban versus rural areas. "Dairy products" (e.g., cow milk) and "Cereals unmilled" (e.g., sorghum) are the product categories with the highest share of auto-consumption in the country (average of 38% and 24%, respectively). Not surprisingly, the auto-consumption rate is higher in rural areas with an average of 23% compared to only 3% in urban areas. Note, however, that these averages hide

Table OA.3: ***Product categories with expenditure weights***

Group	Category	Group	Category
Alcohol and tobacco	beer	Farm equipment	Farm equipment
Alcohol and tobacco	cigarettes	Footwear	Footwear
Alcohol and tobacco	spirit	Fresh fruits	Fresh fruits
Alcohol and tobacco	wine	Fuel and power	Fuel and power
Animal products	Draught animal	Furnishing	Furnishing
Animal products	Fish	House ware	House ware
Animal products	Livestock	Household operation	Household operation
Animal products	Meat	Medical care	Medical care
Bread and other prepared food	Bread and other prepared food	Miscellenaous goods and services	Charge for Money Transfer
Cereals	Barley	Miscellenaous goods and services	Miscellenaous goods and services
Cereals	Flour	Non alcoholic beverage	Non alcoholic beverage
Cereals	Maize	Oil and fats	Oil and fats
Cereals	Millet	Oil seeds	Oil seeds
Cereals	Oats	Other food items	Other food items
Cereals	Rice	Other household equipment	Other household equipment
Cereals	Sorghum	Other medical expenses	Other medical expenses
Cereals	Teff	Pasta	Pasta
Cereals	Wheat	Personal care	Personal care
Clothing and apparel	Apparel	Personal effects	Jewelry
Clothing and apparel	Clothing	Personal effects	Personal effects
Clothing and apparel	Cost of tailoring	Potatoes and roots	Potatoes and roots
Coffee, tea, chat	Coffee, tea, chat	Pulses	Pulses
Communication	Communication	Restaurants and cafes	Restaurants and cafes
Construction materials	Construction materials	Routine maintenance	Routine maintenance
Cost of milling	Cost of milling	Spices	Spices
Diary products	Diary products	Transport	Transport
Education	Education	Vegetables	Vegetables
Equipment recreative	Equipment recreative		

large discrepancies at the household level, as shown in Figure OA.4. Auto-consumption is below 5% for almost all of the households living in urban districts. For households in rural districts, the distribution of auto-consumption shares is bi-model with a (large) peak around 0 and a (small) peak close to 1.

Figure OA.4: ***Home production***



Note: Auto-consumption rates for food products computed from the 2015 household survey (LSMS data). The plain gray bars represent the distribution of auto-consumption shares for households living in rural districts. The red empty bars are for households living in urban districts. The red and grey vertical lines indicate the average auto-consumption rate for urban and rural households, respectively.

Distance and travel time between cities. We first need to assess how far each city in the dataset is from the other cities. Distance is computed thanks to the Stata package Geodist. Travel time is computed using the Stata package Georoute (Weber and Péclat, 2016). The travel times to other cities provided by this package are missing for 10 cities. For these cities, we predict travel time based on bilateral distance.[14] We use a similar strategy to compute the distance and travel time to the capital city (Addis-Ababa) and to the main international trade corridor (Kombolcha, through which shipments from and to Djibouti transit).[15]

Remoteness. For each city, we aggregate the information on bilateral distances with the 105 other cities by building a remoteness index, which is the average travel time to the other

[14] Travel time is regressed on a polynomial of degree 7 of distance as well as fixed effects for origin-destination pairs of regions. For the 8,730 pairs of cities for which we have both forms of information, the R^2 is close to 90%. Travel time is then predicted for all the city-pairs involving the 10 destinations for which information on travel time is missing.

[15] But in this case, a linear fit with region fixed effects is enough.

Ethiopian cities:

$$\text{remote}_c = \frac{1}{105} \sum_{j \neq c} \text{travel time}_{cj},$$

where remote_c is the remoteness index for city c, and travel time_{cj} is the travel time between cities c and j. In robustness checks, we use a population-weighted version of this remoteness index.

OA.2 Additional results

A focus on imported products. A reason why remoteness is positively correlated with prices may be that remote locations are farther away from places where goods are produced. Unfortunately, we do not have information on where distributed goods are produced. However, we have already mentioned that 90% of imported products enter Ethiopia through Djibouti, and that Kombolcha is the main Ethiopian city on that commercial route. We identify a list of 30 imported products by keeping those featuring "imported," "England," "China," or "Japan," in their name, along with five additional products identified by Atkin and Donaldson (2015) as imported. We then run our baseline specification on this subsample, and add to the set of regressors the travel time to Kombolcha. The results show that cities that are farther away from the trade corridor exhibit higher prices, and that while the coefficient on remoteness slightly decreases, it remains positively and highly significantly related to the price of available varieties.[16] This suggests that remoteness captures factors beyond just travel time for shipping goods. One possible explanation is that there is less competition in remote cities pushing prices up. For a given travel time, transportation costs might also be higher if risk is higher on these roads due to conflicts for instance, or if there is less competition among carriers.

A focus on barcode-like products. Most of the daily transactions made by Ethiopian consumers do not involve barcode products. The rich data we have allow us to compare the typical price of a representative set of similar, but not identical, products and services across locations. This implies that price differences may be related to differences in varieties of a given product across locations. In a robustness exercise presented in Table OA.1, we reproduce the analysis on a subset of 27 brandname products with narrow details on the units. As noted by Atkin and Donaldson (2015), the description of these products is as detailed as barcode products.[17] These 27 products account for 7% of the consumption in our basket.

[16] Results available upon request.

[17] The products are: Ambo Mineral Water (500cc), Ball Point-Bic England, Beer Bedele (330cc), Beer Harar (330cc), Beer Meta Abo (330cc), Buthane Gas Shell (12.5 Kg), Coca Cola/Fanta (300cc), Detergent Omo (50gm),

For these products, the average missing rate is 31%, i,e, very close to the 34% measured in the sample of 401 products used in the benchmark analysis. Results in Table OA.1 show that the relationship between remoteness and the price and availability of products, as well as the relationship between market size and product availability, persist in this narrow sample (although sometimes less precisely estimated due to smaller sample size). Quantitatively, the relationship we estimate between prices and city remoteness for barcode-like products is very close to the one estimated for the whole sample: this suggests that the Alchian-Allen effect, where higher quality varieties are the only ones to be shipped to more remote markets, is probably not the main mechanism explaining the observed higher prices in remote locations.

On the opposite, the relationship between city size and prices turns from positive to almost zero in this narrow sample. This changing result suggests that part of the price premium in large cities may be explained by composition effects. Whether these composition effects reflect higher quality is unclear. Indeed, the sample used for this check is quite reduced compared to the initial one (27 out of 401), and we have shown in Table 1 of the paper that the results are robust to controlling for GDP per capita, which suggests that quality alone cannot explain these patterns.

To further investigate this issue, we also examine whether the elasticity of prices to population size varies with the length of the quality ladder of products. Specifically, we define products with large cross-city price variations as more likely to exhibit quality differences. To do so, we calculate the coefficient of variation of local prices for each product and create a dummy variable that takes the value of one for products with a coefficient of variation above the median (i.e. 0.06).[18] We then run our baseline price regression, adding an interaction term between city size and the dummy identifying products with longer quality ladders. Compared to the benchmark results, the coefficient on city size remains positive but drops from 0.04 to 0.025. Moreover, the interaction term between city size and the vertical differentiation dummy is positive and significant (coefficient of 0.027), suggesting that the elasticity of prices to city size is higher for products potentially affected by vertical differentiation.[19] This augmented specification suggests that quality variations across

Detergent Zahira (50gm), Electric Bulb Philips(40/60 Watt), Eveready Drycell, GIV Toilet Soap (90gm), Gissila Packet, Marlboro Packet, Liter of Motor Oil (Mobil), Newspaper (Addis Zemen), Nyala Packet, Pepsi Cola/Mirinda(300cc), Powdered Milk (Meamp) (450gm), Radio Set Philips 3 Band, Rothmans/England Packet, Saris Wine (Normal 750cc), T.V. Set Philips 21 inch (Colored), Toilet Paper (Mamko Roll), Water Tanker Roto (1 meter cube), Zenith Hair Oil (Liquids Form 330cc), Zenith Hair Oil (Non-Liquids Form, 330cc).

[18] Note that 75% of barcode products identified in the first check have a coefficient of variation below this threshold.

[19] Detailed results available upon request.

cities may partially—but not entirely—explain the observed correlation between city size and prices.

Distribution costs. Besides higher quality, higher prices in larger cities could also reflect higher producers' markups and/or higher distribution costs. Feenstra et al. (2020) show that if marginal and sunk costs are similar across cities, population size has an unambiguously negative effect on producers' markups and prices through higher entry, meaning that there is a pro-competitive effect of market size. However, if sunk costs depend on local costs, or if sunk costs become endogenous to market size (with higher advertising expenditures in the largest cities for example), then their model predicts that markups—and thus prices—will be less negatively related to market size and could even be higher in larger markets (anti-competitive effect of market size). It is also possible that the price paid by consumers depends on the combination of the wholesale price of the product, the retailer's marginal cost, and the retailer's markup. In such cases, the marginal cost inclusive of distribution costs would rise with local wages and rents. This could lead to higher prices in large cities, even if producers' quality and markups remain constant across markets.

We cannot estimate markups with our data, but we can leverage our dataset to compute proxies for local unskilled labor costs and local real-estate costs that both may affect distribution costs.[20] We then include these two proxies in our baseline specification, excluding the products used to compute the proxies. This reduces the direct effect of city size on prices by 40%.[21] Higher unskilled labor and real-estate costs in larger cities may be related to several different things, but one possible explanation to our results is that higher distribution costs are passed through to consumers, partially explaining why prices are generally higher in larger cities.

[20] Our data includes three measures of unskilled labor—Maid Servant, Guard Salary, and Unskilled Service (Daily Laborer). We run a regression of the price (salary) of these three types of workers on product and city fixed effects, and the estimated city fixed effects are used as proxies for local labor costs. As for real-estate costs, we make the assumption that construction costs can serve as a proxy for real-estate prices (for example Combes et al., 2021, incorporate proxies for construction costs in their investigation of the production function for housing). Focusing on products such as Sand, Stone for House Construction, Window Glass (50 cm x 50 cm x 3 mm), Wood for House Construction ('Atana'), and Cement Tile, we run an analysis similar to the one on unskilled labor to build a proxy for local real-estate costs.

[21] The coefficient on city size drops from 0.038 to 0.023, detailed results available upon request.

Table OA.4: ***Price and availability of barcode products across Ethiopia***

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Price of individual product (log)				*Product not available (dummy)*			
Ln Remoteness	0.094^a	0.088^a		0.099^a	0.086	0.073		0.097
	(0.029)	(0.028)		(0.028)	(0.062)	(0.065)		(0.062)
Ln Population	-0.004	-0.000	-0.003	-0.006	-0.060^a	-0.050^a	-0.060^a	-0.063^a
	(0.006)	(0.007)	(0.005)	(0.006)	(0.008)	(0.010)	(0.008)	(0.009)
Ln per cap. inc.		-0.021				-0.052		
		(0.020)				(0.035)		
Ln pop-weighted			0.098^a				0.048	
remoteness			(0.022)				(0.044)	
Addis	Yes	Yes	Yes	No	Yes	Yes	Yes	No
Observations	1,968	1,968	1,968	1,944	2,862	2,862	2,862	2,835
R-squared	0.02	0.02	0.03	0.02	0.06	0.06	0.06	0.05

Notes: The dependent variables are the log price (columns (1) to (4)) and a dummy equal to 1 if the product is unavailable (columns (5) to (8)). These dependent variables are defined in the $product \times city$ dimension. Standard errors account for spatial autocorrelation within a 50km radius around cities using the spatial HAC procedure. ^{a}, ^{b} and ^{c} respectively, denote significance at the 1, 5, and 10% levels.

Table OA.5: ***Correlation between remoteness, city size, and socio-demographic variables***

	Remoteness (log)	Pop. (log)
Remoteness (log)	1.000	
	(.)	
Pop. (log)	-0.099	1.000
	(0.313)	(.)
Dominant ethnicity politically connected	-0.307	-0.230
	(0.001)	(0.018)
Ethnic diversity index (log)	0.146	-0.050
	(0.134)	(0.610)
Time to Addis (log)	0.779	-0.131
	(0.000)	(0.185)
Time to corridor (log)	0.125	-0.223
	(0.205)	(0.022)

Notes: Correlation coefficients (with p-value in parentheses). The variable "Dominant ethnicity politically connected" is a dummy equal to one when the dominant ethnic group in the city is the same as the Prime Minister's one. The index of ethnic diversity is the inverse of a Herfindahl index based on the share of the various ethnic groups in the population.

Table OA.6: ***Coefficients estimated by sector***

	Prices				Product availability			
	Remoteness		Population		Remoteness		Population	
category	coef.	s.e.	coef.	s.e.	coef.	s.e.	coef.	s.e.
BEVERAGES - ALCOHOLIC	.101	.050	.015	.006	.221	.068	-.060	.009
BEVERAGES - NON ALCOHOLIC	.265	.047	-.011	.010	.033	.049	-.015	.008
BREAD AND O. PREPARED FOODS	.137	.064	-.003	.010	.249	.107	-.038	.017
CEREALS MILLED	.061	.055	.035	.007	.258	.148	-.081	.021
CHAIRS, TABLES, ETC.	.102	.068	.068	.010	.161	.074	-.035	.009
COFFEE, TEA, CHAT ...	.468	.126	.074	.019	.099	.093	-.097	.010
COMMUNICATION	.028	.098	-.012	.014	.162	.066	-.035	.012
CONSTRUCTION MATERIALS	.355	.046	-.010	.007	-.160	.209	-.114	.021
COST OF MILLING	.516	.106	.026	.019	.008	.086	-.121	.010
DIARY PRODUCTS AND EGG	.043	.083	.064	.013	.287	.147	-.081	.017
DOMESTIC SERVICE	.407	.157	.099	.021	.181	.062	-.087	.010
EDUCATION	.017	.100	.023	.010	-.013	.115	-.063	.019
EQUIPMENT AND ACCESSORIES	.265	.065	.002	.011	-.128	.088	-.015	.010
FARM EQUIPMENT (HAND MADE)	.252	.106	.068	.014	.148	.091	-.066	.012
FOOD TAKEN AWAY FROM HOME	-.012	.066	.125	.009	.062	.128	-.096	.015
FOOTWEAR (MEN AND WOMEN)	.080	.056	.056	.008	-.028	.088	-.092	.016
FRUITS FRESH	.149	.228	.188	.031	.137	.143	-.046	.017
FUEL AND POWER	.037	.072	.026	.008	.029	.050	-.034	.012
HOUSEHOLD OPERATION	-.025	.059	.010	.007	.048	.117	-.063	.014
LIVESTOCK	.010	.059	.073	.009	.012	.066	-.053	.011
MEDICAL CARE	.069	.092	-.030	.016	.054	.063	-.083	.011
MISCELLANEOUS	.293	.111	-.062	.019	.082	.099	-.079	.016
OIL SEEDS	.006	.092	.053	.012	.011	.140	-.095	.019
OILS AND FATS	-.130	.066	.014	.010	.061	.081	-.045	.014
OTHER FOOD ITEMS	.132	.057	.008	.009	.080	.095	-.096	.013
OTHER HOUSEHOLD EQUIP.T	-.136	.101	.016	.018	.045	.112	-.067	.012
OTHER MEDICAL EXPENSES	.127	.169	.191	.025	.371	.102	-.061	.011
PERSONAL CARE	.113	.051	.007	.007	.236	.115	-.067	.013
PERSONAL EFFECTS	.087	.059	.048	.012	-.014	.116	-.065	.014
PLASTIC WARE (LOCAL MADE)	.029	.081	.046	.008	.111	.095	-.133	.016
POTATOES, OTHER TUBERS	.408	.198	.134	.019	-.044	.060	-.043	.008
PREPARED CEREAL PRODUCTS	.121	.023	-.004	.004	.556	.132	-.028	.018
PULSES UNMILLED	.069	.044	.029	.006	.210	.148	-.049	.017
READY-MADE / FOR ADULTS	.026	.061	.052	.011	.288	.077	-.057	.010
SPICES	.198	.109	.033	.017	.142	.107	-.082	.014
TRANSPORT	.186	.077	-.017	.015	.017	.107	-.169	.015
VEGETABLES (FRESH)	.291	.116	.033	.015	.066	.043	-.055	.006

Notes: Each line is a broad product category. Each column corresponds to the coefficients and the standard errors estimated by sector using the main specification (Columns (1) and (5) of Table 1). The first four columns correspond to the regression of prices on remoteness and city size. The last four columns correspond to the regression of product availability. Standard errors account for spatial autocorrelation within a 50km radius around cities using the spatial HAC procedure.