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

February 2, 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 4.7% 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.

†Corresponding author: martin.julien@uqam.ca.

‡mayneris.florian@uqam.ca.

§william_theophile.ewane@courrier.uqam.ca.

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] Some papers emphasize city size as a key determinant of the dispersion in prices and 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). Studies focusing on developing countries often examine price differences between urban and rural areas (see, e.g., Ravallion and Van De Walle, 1991; Deaton and Tarozzi, 2000; Muller, 2002), and highlight 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 analyze how both city size and remoteness shape the spatial dispersion in the level of prices and the availability of products 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. The analysis at the micro-level shows that overall, the 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 more costly missing products are.

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

[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.

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. This quantification is based on an elasticity of substitution equal to 4, most estimates in the literature ranging from 1.5 to 6.

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. With an elasticity of substitution of 4, the cost-of-living that accounts for both margins remains 4.7% higher in large cities. Note that the greater product availability in large cities 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.

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.[2]

[2]Ethiopia serves as an ideal laboratory for understanding the impact of intra-national isolation on price variations across space within a developing economy context. With a population of over 100 million, it is the second most populous country in Africa and the largest landlocked country in the world. The population is unevenly distributed across the country, with around 3% to 5% residing in the metropolitan area of Addis Ababa, while approximately 20 cities throughout the country have populations in the five-digit range. Similar

Second, a couple of recent papers examine prices and product availability in the context of Ethiopia (Gunning et al., 2018; Krishnan and Zhang, 2020). Gunning et al. (2018) 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.[3]

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).[4] We complement this literature by showing that, in a country such as 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

to other countries in the region, Ethiopia has experienced significant GDP growth since the early 2000s, along with notable improvements in several development indicators.

[3]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.

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

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 manufacturers' 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.

In the paper, we run a check where we reduce the sample to 27 products that are as precise as barcode categories. Neither the impact of city remoteness and city size on product availability, nor the impact of city remoteness on prices are significantly affected by this; but the positive relationship between city size and prices vanishes. Still, we argue that the positive impact of city size on prices found in the full sample provides valuable insights into the analysis of spatial differences in the cost of living, even if it partly reflects the fact that consumers purchase heterogeneous varieties across locations. Our approach indeed compares, across cities, the typical prices faced by a consumer that wants to buy a given product, while approaches relying on barcode (scanner) data compare the prices faced by a consumer willing to buy the exact same variety within a given retail chain. These two approaches simply 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).[5] 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. 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 35% are categorized as missing.

[5]Note that unlike our data, there is no information on the level of rents (and prices in general) at the city level in this other dataset.

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. Benford's Law states that the distribution of the first digit of numerical data remains stable across different samples, with the frequency of number 1 being around 30%, number 2 around 17%, and number 9 appearing as the first digit in only 5% of occurrences. 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 is difficult. Figure OA.2 in Appendix presents the frequency distribution of the first digit of price quotes in our data alongside the expected distribution according to Benford's Law. The observed frequencies align well with the frequencies predicted by the Benford law.[6] 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".

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

[6] 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.

travel time between cities c and j. The travel time between Ethiopian cities is computed using the Stata package GEOROUTE (Weber and Péclat, 2016).[7] In robustness checks, we consider alternative measures of 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.

Empirical approach. Our baseline specification is as follows :

$$\mathrm{y}_{pc} = \alpha \mathrm{pop}_c + \beta \mathrm{remote}_c + \omega_p + \epsilon_{pc}, \tag{1}$$

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 (unweighted) average time to reach all the other Ethiopian cities in the database from c. 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. Last, ϵ_{pc} is the error term. We allow for a possible correlation in the errors across products within a city and thus cluster standard errors at the city-level.

[7] 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 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.

Baseline results. Columns (1) to (4) of Table 1 present the estimates of the effect of remoteness and city-size on prices. As can be seen from column (1), remoteness and city-size are significant determinants of 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)), 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 robust to controlling for income per capita and excluding Addis-Ababa. The impact of remoteness remains positive but not statistically significant with an alternative population-weighted measure of distance to other cities. This suggests that the distance to other markets, independent of their size, is the primary impediment to product availability in a city.

Sensitivity analysis. 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), or to the most important commercial routes, namely, Kombolcha, which is on the corridor to Djibouti, than to distance to all other cities. We thus control for distance to these two cities in columns (1) and (3). None of these variables significantly impacts the price and availability of products.

Also, 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 the inverse of a Herfindahl index based on the share of the various ethnic groups in the population in columns (2) and (4). Ethnic diversity has no significant impact on the price and availability of products.

In the context of a developing country such as Ethiopia, another important issue that

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^b	0.100^c		0.116^b
	(0.030)	(0.033)		(0.030)	(0.052)	(0.056)		(0.052)
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)	(0.009)	(0.009)	(0.009)	(0.010)
Ln per cap. inc.		-0.025				-0.022		
		(0.023)				(0.029)		
Ln pop-weighted			0.041^c				0.058	
remoteness			(0.025)				(0.038)	
Addis	Yes	Yes	Yes	No	Yes	Yes	Yes	No
Observations	28,183	28,183	28,183	27,810	42,506	42,506	42,506	42,105
R-squared	0.97	0.97	0.97	0.97	0.39	0.39	0.39	0.39

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 clustered at the city-level in parentheses, ^{a}, ^{b} and ^{c} respectively, denote significance at the 1, 5, and 10% levels.

Table 2: ***Sensitivity analysis***

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	price (log)		missing		price (log)		missing	
Ln Population	0.039^a	0.040^a	-0.068^a	-0.072^a	0.046^a	0.038^a	-0.047^a	-0.044^a
	(0.005)	(0.004)	(0.010)	(0.009)	(0.006)	(0.007)	(0.006)	(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					
	(0.006)		(0.010)					
Ln avg. dist. to other cities		0.121^a		0.103^c	0.087^b	0.082^b	0.175^a	0.176^a
		(0.030)		(0.053)	(0.040)	(0.040)	(0.039)	(0.039)
Log Ethnic diversity index		-0.007		0.007				
		(0.013)		(0.026)				
Home production						-0.089^a		0.023
						(0.028)		(0.022)
Observations	27,697	28,183	41,704	42,506	8,672	8,672	12,370	12,370
R-squared	0.97	0.97	0.39	0.39	0.93	0.93	0.40	0.40

Notes: Dependent variables are defined in the *product* × *city* dimension. Standard errors clustered at the city-level in parentheses, ^{a}, ^{b} and ^{c} respectively, denote significance at the 1, 5, and 10% levels.

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 impact of home-production on 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.

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 3, 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.[8] These 27 products account for 7% of the consumption in our basket. 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 3 show that the impact of remoteness on prices and product availability, as well as the impact of market size on product availability persist in this narrow sample (although sometimes less precisely estimated due to smaller sample size).

However, the impact of city size on 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

[8]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), 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).

quality is unclear. Indeed, we have shown in the previous section that the results are robust to controlling for GDP per capita, which suggests that quality alone cannot explain these patterns.

In the end, 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 even if they capture differences in the set of available varieties.

Table 3: ***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^c
	(0.028)	(0.026)		(0.030)	(0.027)	(0.059)		(0.057)
Ln Population	-0.004	-0.000	-0.003	-0.006	-0.060^a	-0.050^a	-0.060^a	-0.063^a
	(0.005)	(0.007)	(0.005)	(0.005)	(0.006)	(0.011)	(0.010)	(0.010)
Ln per cap. inc.		-0.021				-0.052		
		(0.020)				(0.034)		
Ln pop-weighted			0.098^a				0.048	
remoteness			(0.021)				(0.041)	
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	0.99	0.99	0.99	0.99	0.53	0.53	0.53	0.53

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 clustered at the city-level in parentheses, ^{a}, ^{b} and ^{c} respectively, denote significance at the 1, 5, and 10% levels.

Summary. So far, the results unambiguously suggest that remote cities are more expensive than central ones. A fewer number of products are available in remote cities and when they are available, products are sold at a higher price (which is consistent with the results in Atkin and Donaldson (2015) on intra-national trade costs). The impact of city size on the cost of living is more ambiguous. All else equal, prices are higher in large cities (as shown by Handbury and Weinstein (2015) for the U.S.) but consumers have access to a wider range of products (as shown by Handbury and Weinstein (2015) for the U.S. and Feenstra et al. (2020) for China). 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 section 4, we put more structure and compute a spatial price index to assess which force dominates.

4 Consumer price index across Ethiopian cities

We now use the product-market level information we have to build local price indexes that account for both the level of prices and the availability of products and services.

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:

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

$$\tag{3}$$

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_j 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.[9] In the composite index

[9]The Sato-Vartia weights are given by

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

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 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.[10]

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 consider different values of σ (1.5; 4; 6) that correspond to the range of values estimated in the literature using different methods (see, among others, Imbs and Mejean, 2015; Aghion et al., 2019; Boehm et al., 2023).

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

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.

[10] 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. We have compared the impact of size and remoteness on the intensive part of the price index (SPI) that can be computed for all cities. The coefficients on remoteness and city size are almost the same in the full sample (106 cities) and the constrained sample (78 cities).

components of the price index against the degree of remoteness of the city, taking 4 as a value for the substitution elasticity; 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. When the latter is equal to 4, 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.

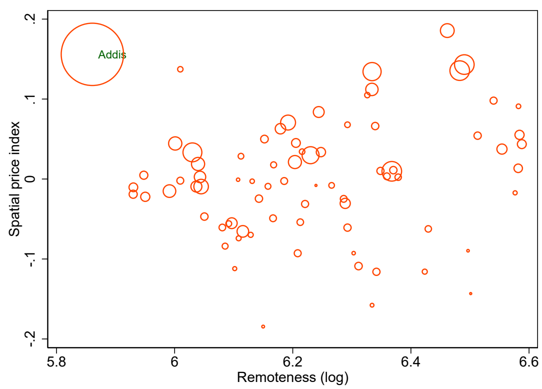
The graphical observations are largely confirmed by the econometric analysis reported in Table 4. The EPI is the product of the intensive and the extensive components, and thus log-linear in these two components.

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), (4) and (6) 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 value of σ. For values of σ of 4 or 6 (columns (5) and (7)), we find that the price effect dominates the variety effect, which implies that, all else equal, the cost of living increases with city size. If instead $\sigma = 1.5$, the variety term dominates (column (3)), which means that the access to a broader set of variety in large cities offsets the higher prices.[11] Note that if the elasticity of prices to distance was close to zero (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.

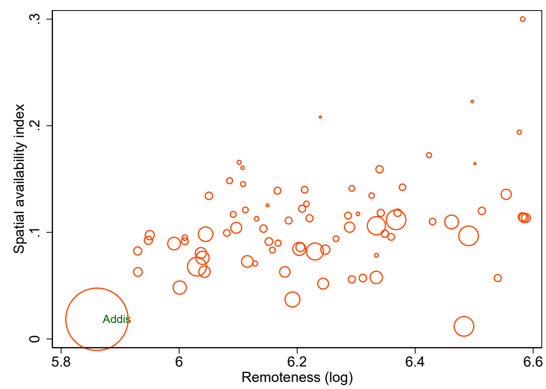
On the other hand, remote locations exhibit both higher intensive and higher extensive

[11] In unreported exercises, we find that the two effects exactly cancel out when σ is equal to 3.

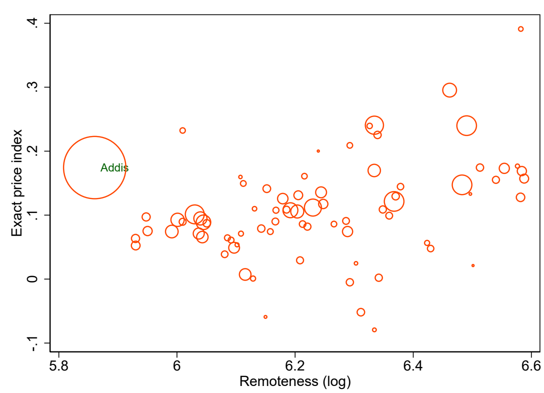
Figure 1: ***The geography of prices and product availability***



(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 using $\sigma = 4$. The relationship between size, remoteness and prices for alternative values of σ are presented in Table 4.

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.

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

	Ln SPI	Ln VA	Ln EPI	Ln VA	Ln EPI	Ln VA	Ln EPI
		$\sigma = 1.5$		$\sigma = 4$		$\sigma = 6$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Ln Population	0.042^a	-0.143^a	-0.101^a	-0.024^a	0.018^b	-0.014^a	0.028^a
	(0.006)	(0.021)	(0.022)	(0.003)	(0.007)	(0.002)	(0.007)
Ln Remoteness	0.103^a	0.373^a	0.476^a	0.062^a	0.165^a	0.037^a	0.140^a
	(0.037)	(0.125)	(0.133)	(0.021)	(0.044)	(0.012)	(0.040)
Observations	78	78	78	78	78	78	78
R-squared	0.39	0.47	0.35	0.47	0.17	0.47	0.20

Notes: The dependent variable is the log spatial price index (Ln SPI) in columns 1, the log spatial availability index (Ln VA) in columns 2-4-6, and the log exact price index (Ln EPI=Ln SPI + Ln VA) in column 3-5-7. In columns 2-7, different values of σ are used to compute the availability index and the exact price index ($\{1.5; 4; 6\}$). Robust standard errors are in parentheses. ^{a}, ^{b}, and ^{c} respectively, denote significance at the 1%, 5%, and 10% levels.

Accounting for product availability quantitatively matters. We consider a σ of 4, which is our preferred value based on our reading of the literature, 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 (5) for the exact price index, which accounts for both the price of available products and for product availability, the cost-of-living is 4.7% 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

This paper documents spatial dispersion in the price and availability of products purchased by consumers across Ethiopian cities. An analysis at the product level shows that consumers in remote locations suffer higher prices and lower product availability than in more central locations. In large cities, the results are more ambiguous: consumers in large cities pay higher prices but have access to a broader set of products and services.

To quantify the relative importance of these forces for consumers' welfare, we compute city-level price indices that vary with the level of prices and the local availability of products. The intensity of the impact of product availability on consumers' welfare depends on the substitution elasticity between products. The less substitutable the products, the higher the consumers' taste for variety, and the more consumers suffers from missing products in small and remote cities. Considering the estimates of the substitution elasticity available in the literature, the results show that in Ethiopia, the cost of living is higher in large cities, even when taking into account the access to a wider range of products. Living in a remote city is particularly costly as consumers have access to a narrower range of products at a higher price.

The strong dispersion in the cost of living across Ethiopian cities 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. 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.

Aghion, Philippe, Antonin Bergeaud, Timo Boppart, Peter J. Klenow, and Huiyu Li, "Missing Growth from Creative Destruction," *American Economic Review*, August 2019, *109* (8), 2795–2822.

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.

Boehm, Christoph E., Andrei A. Levchenko, and Nitya Pandalai-Nayar, "The Long and Short (Run) of Trade Elasticities," *American Economic Review*, April 2023, *113* (4), 861–905.

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.

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.

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, Andualem T Mengistu et al., "Fading Choice: Transport Costs and Variety in Consumer Goods," Technical Report, Centre for the Study of African Economies, University of Oxford 2018.

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.

Imbs, Jean and Isabelle Mejean, "Elasticity Optimism," *American Economic Journal: Macroeconomics*, July 2015, *7* (3), 43–83.

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.

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

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.

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.

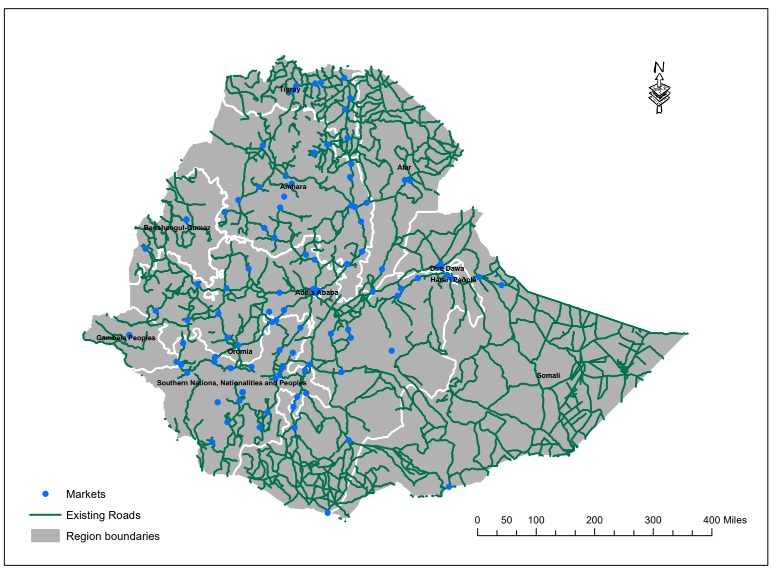
Young, Alwyn, "Inequality, the Urban-Rural Gap, and Migration," *The Quarterly Journal of Economics*, 2013, *128* (4), 1727–1785.

OA Online appendix

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 reaches 24,200 Birr for households in the fifth quintile. We use average income per capita

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

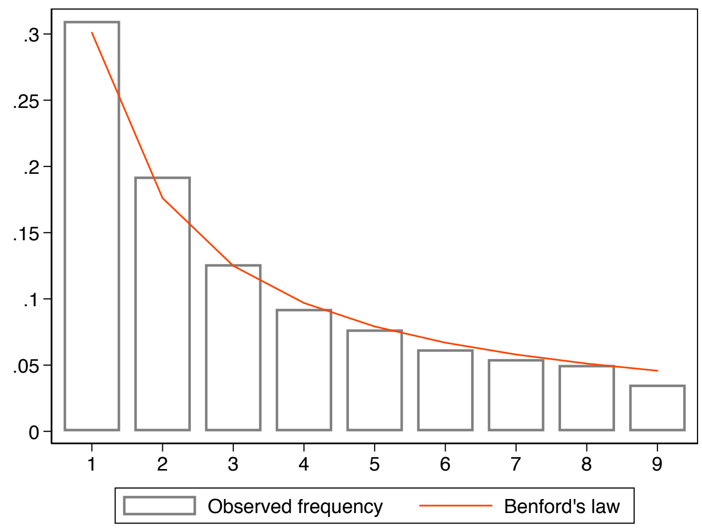
Table A.2 (continued).

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 A.2 (continued).

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.

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.

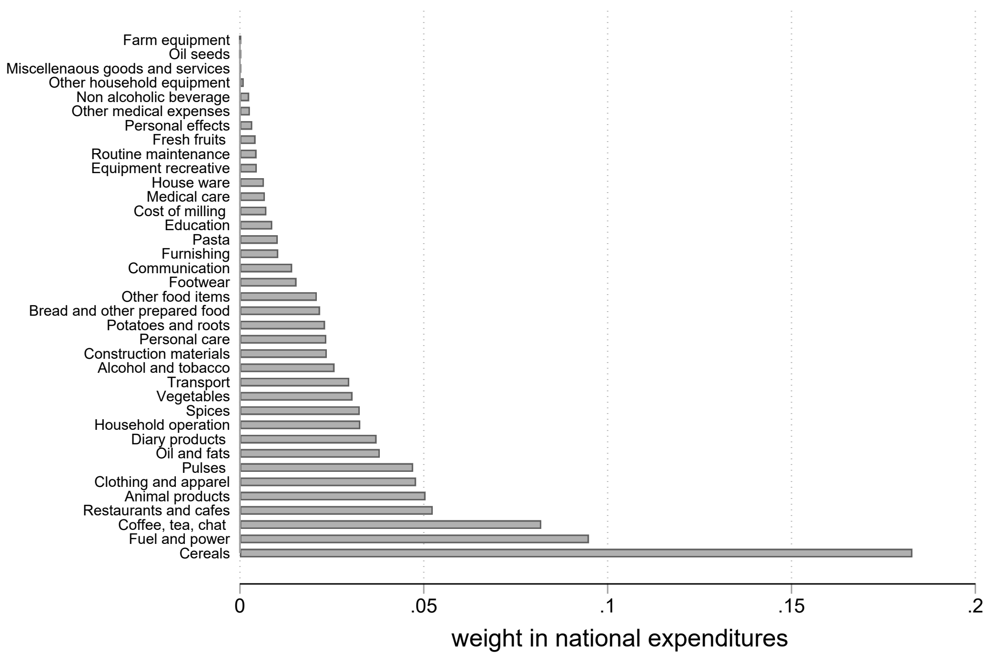
Expenditure weights. Aggregate price indexes require information on expenditure weights across product categories. The price data do not include information on products weights. We thus rely on another dataset obtained from the Ethiopian Statistical Agency. From this dataset, we have information on the expenditure weights of 55 product categories covering the products and services in the Ethiopian CPI. The weights are specific to each Ethiopian region.[12] 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 categories are presented in Table OA.3.

[12] 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.

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		

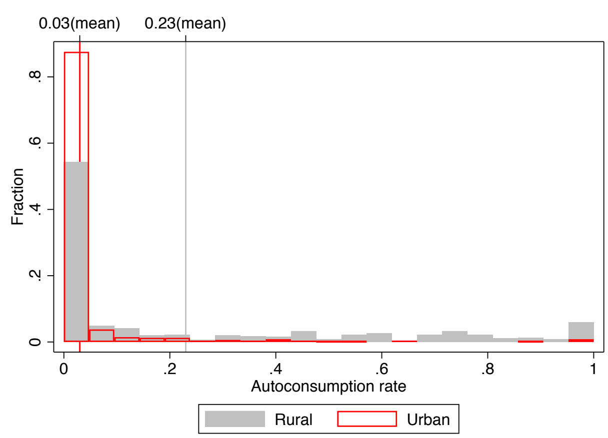
Figure OA.3: ***Weights in terms of expenditures***



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 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.[13] 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).[14]

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 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.

[13] 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.

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