UCL Journal of Economics

Article

Demand for Informal goods in Africa: Demand System Estimation with Quality Effects and its Limitations

Kexin Wu[1]

[1] Student (Bsc Economics, 2020-23), University College London, United Kingdom; kexin.wu@ucl.ac.uk

Submission Date: 1st April 2023; Acceptance Date: 11th July 2023; Publication Date: 2nd October 2023

How to cite
Wu, K. (2023). Demand for Informal goods in Africa: Demand System Estimation with Quality Effects and its Limitations. UCL Journal of Economics, vol. 2, pp. 85-110.
DOI: https://doi.org/10.14324/111.444.2755-0877.1590

Peer review
This article has been peer-reviewed through the journal's standard double-blind peer review, where both the reviewers and authors are anonymised during review

DOI: https://doi.org/10.14324/111.444.2755-0877.1590

Open access
UCL Journal of Economics is a peer-reviewed open-access journal

Abstract

The growth of many African Economies is accompanied by an unusual expansion of the informal sector but the demand for informal goods is rarely tested. This study examines the demand for informal goods in Africa, specifically in the Democratic Republic of the Congo in 2012, using household-level survey data. This paper reviews and summarizes that income inequality, quality preference, and preference for some product-specific characteristics are the main drivers of the demand for informal goods. This paper hypothesizes three channels of demand for informal goods and uses the empirical method in Angus Deaton (2018) to test the hypothesis by estimating expenditure elasticity, price elasticity, and quality elasticity. The results show that the demand for informal goods generally decreases as income increases and income inequality reduces, but this pattern can differ for different categories of goods. The results show that the three channels are likely to be plausible and informal and formal goods can be both substitutes or complements in the final markets. Policies to improve formality may consider the different demand patterns across markets to be more effective. Potential limitations of the method in this specific context and implications for future improvements are discussed.

Keywords: Demand Estimation, Quality effect, Informal Sector

1. Introduction

Since 2000, the gradual take-off of total employment and economic growth in many African countries like Ethiopia, Ghana, and Kenya is largely contributed by the expansion of informal employment, with formal employment remaining relatively stagnant (McMillan and Zeufack, 2022; Diao et al., 2021). Medina, Jonelis, and Cangul (2017) estimate that the size of the informal sector in Africa remained the highest in the world from 2010 to 2014 and varied from around 25 per cent in Mauritius and Namibia to over 40 per cent in the Democratic Republic of the Congo, Benin, and Nigeria. This is contradictory to patterns of Asian developing countries, such as Vietnam, where the take-off of total employment was accompanied by flat or even declining informal employment (McCaig and Pavcnik, 2018).

Most studies about expanding informal sectors in Africa focus on the supply side, like capital or labour costs. However, Diao et al. (2021) suggest that a supply-driven structural change should coincide with the expansion of formal activities in the manufacturing sector, while a demand-driven structural change is more likely to be observed with the marginal entrance of informal firms in Africa. Kruse et al. (2022) suggest that industrialization in Sub-Saharan Africa is accompanied by proliferating unregistered firms because there is a significant demand for informal manufactured goods. Few studies provide evidence about why there is substantial demand for informal goods and how the demand-side factors can relate to supply-side factors. If we ignore demand-side factors, we may fail to understand the full potential impact of policies to improve formality. Banerji and Jain (2007) suggest that some policies to improve formality from the supply side, like lowering the capital costs for informal firms, may help informal firms produce a wider range of products, compete with the formal sector, and promote informality. To understand the welfare implication of the expanding demand for informal goods, it is important to understand the demand pattern of informal goods in more detail.

This paper explores the demand for informal goods from three aspects. First, previous studies suggest that income inequality and consumers' preference for quality are necessary conditions for demand for informal goods in the final goods markets (Mishra and Ray, 2010; Mishra, 2022). There is also evidence about consumers choosing informal goods for other features like accessibility. This paper reviews interactions between income, preference for quality, and preference for other characteristics in determining the demand for informal goods, and summarizes three channels of the demand: First, informal goods are preferred only as cheaper and low-quality substitutes to formal goods; Second, informal goods are cheaper substitutes but are also preferred by their own traits except for the lower price; Third, informal goods are complements for formal goods.

Next, this paper uses the household-level expenditure data from the Democratic Republic of the Congo in 2012 and the aggregate demand system with quality effects from Deaton (2018) to estimate the demand for informal goods, providing estimates for income, price, quality elasticity and cross-price elasticity between informal and formal goods. The results show that the three channels summarized are likely to be plausible. The budget share for informal goods generally decreases as total expenditure increases, but this pattern can vary across categories of goods. For example, the budget share of informal transportation is increasing with income. Informal and formal goods can be imperfect substitutes or complements in the final markets. Thus, policymakers may customize policies in different markets to be more efficient.

DOI: https://doi.org/10.14324/111.444.2755-0877.1590

Finally, since the demand for informal goods is rarely tested empirically and the method in this paper is applied under assumptions, this paper discusses the limitations in estimation and how future works may improve upon it.

The paper is organized as follows: Section II reviews evidence about the demand for informal goods in Africa. Section III hypothesizes the three channels of demand for informal goods. Sections IV and V introduce the dataset and the empirical strategy respectively. Section VI discusses the potential limitations of the method in the specific context and adjustments to data. Section VII discusses the empirical results.

2. Literature Review

2.1. Disparity in Costs and Quality of Products between Formal and Informal Sectors in Africa

The informality of firms is generally defined by small size, lack of registration with government agents, lack of financial statements, and constrained access to credit (Benjamin and Mbaye, 2012). Constraints in production like high entry costs and financial constraints to accessing capital are exacerbated by constraints on the institution side like the weak legal framework (Dabla-Norris, Gradstein and Inchauste, 2008; Ulyssea, 2010).

Barriers between formal and informal sectors lead to a disparity in production, especially the lower capital-labour ratio in the informal sector (de Paula and Scheinkman, 2007). The informal sector in Africa is much more labour-intensive than the formal sector and is the main source of absorbing employment, especially among disadvantaged groups (Etim and Daramola, 2020). The wage of informal workers is on average lower due to the lower observable or unobservable skills and the lack of minimum wage enforcement (Bargain and Kwenda, 2014; Rauch, 1991). While the formal sector in Africa is not absorbing employment effectively, facing much higher labour costs relative to their productivity and better access to capital, making them more capital-intensive relative to their income level (Hernando De Soto, 2006; McMillan and Zeufack, 2022; Diao et al., 2021).

One implication of the disparity is the difference in the quality of products. Producing high-quality products requires more capital or high-skilled labour, and the costs of quality function are decreasing in the capital-labour ratio (Copeland and Kotwal, 1996). Thus, formal firms have advantages in producing high-quality goods, while informal firms have advantages in producing low-quality but cheaper counterparts. Consumers' choices depend on their preferences for quality and income (Banerji and Jain, 2007). The disparity in inputs can also lead to horizontal differentiation. The higher labour intensity of informal sectors allows their products to be more personalized, cultural, and traditional, while products from the formal sector are generalized and industrialized. In a market with both sectors, consumers' choices can depend on their preference for product-specific characteristics.

2.2. Empirical Evidence about Consumer Bases for Informal Goods in Africa

Empirical evidence about the demand for informal goods is quite rare in Africa. Böhme and Thiele (2012) apply household-level survey data from six African countries to show that households working in both the informal and formal sectors have a significant demand for informal goods. Besides, the income elasticity of informal goods is consistently less than one across countries. Their results suggest that the expenditure

DOI: https://doi.org/10.14324/111.444.2755-0877.1590

share on informal goods is decreasing in income though there are overlapping consumer bases between informal and formal goods. These estimates classify most foods and non-alcoholic beverages as informal goods while classifying electricity, fuel, clothing, and footwear as formal goods. The estimates aggregate across categories of goods, and the heterogeneity between categories was untested due to data limitations. Bachas, Gadenne, and Jensen (2020) define formality based on the probability that the consumption taxes are included in the consumer prices in a particular distribution channel. The informal Engel curve (IEC) is on average flatter and higher for African countries like the Democratic Republic of the Congo. This result implies budget shares for informal goods are decreasing but persistent across income groups.

2.3. Important Building Blocks for the Demand for Informal Goods

The first factor to explain the demand for informal goods is income inequality. As discussed in Section II.1, due to the cost dualism, formal and informal sectors have advantages in producing goods of different quality. The mechanism through which income inequality affects the demand for informal goods is closely related to people's preference for quality. Mishra and Ray (2010) develop a model specifying informal and formal goods as close substitutes, showing that a sufficiently high inequality of income is necessary to generate demand for low-quality informal products. The persistent demand for informal goods is mainly because income inequality is too high and the demand for informal goods is decreasing with income. Mishra (2022) shows that the coexistence of the informal and formal sectors is less likely to reach equilibrium if consumers' preference for high-quality formal goods over low-quality informal goods is not large enough.

Banerji and Jain (2007) model the disparity on the supply side between informal and formal sectors and propose that this would lead to the quality dualism between informal and formal goods at equilibrium. At the equilibrium, high-quality formal goods are chosen by consumers with incomes higher than a certain threshold, while low-quality informal goods are chosen by consumers with incomes lower than the threshold.

Böhme and Thiele (2012) suggest that if there is a significant quality dualism, we should observe a lower income elasticity for informal goods, which is consistent with their estimation of elasticity in African countries. Nearly half of the subcategories exhibit significant differences in unit prices between informal and formal goods, and 96 per cent of them show a lower value for informal goods. Bachas, Gadenne, and Jensen (2020) show that in the Democratic Republic of the Congo, 34.4 per cent of the purchases of informal goods are explained by their lower prices. However, except for prices and quality, half of the purchases of informal goods are for other reasons, like access or store attributes, affecting people's preference for quality. Preference for some product characteristics may interact with quality preference and income. Besides, models generally assume that informal and formal goods are imperfect substitutes while ignoring the possibility that they can be complements in some markets.

DOI: https://doi.org/10.14324/111.444.2755-0877.1590

3. Hypothesis: Three Channels of Demand for Informal Goods

First, informal goods are preferred by low-income households as cheaper and low-quality substitutes for formal goods. The demand for informal goods is mainly attributed to income inequality and consumers' preference for quality. The budget share on informal goods decreases as income increases. The expenditure elasticity for informal goods is expected to be less than one and lower than that of formal goods.

Second, informal goods are imperfect substitutes for formal goods but are appreciated by a wider range of households for some other characteristics, which can give them more power to compete with formal goods as consumers' income increases. We may expect the income elasticity of informal and formal goods to be similar or even higher for informal goods.

Third, informal goods can be complementary goods to formal goods. Tucho (2022) suggests that in many Sub-Saharan cities, informal non-motorized transportation is complementary to formal motorized transportation. As the country develops, imported or industrialized goods begin to appear in the market, but it can take time for consumers to fully accept them, and consumers may still need informal goods to complement their formal purchases.

4. Data

The data used in this paper is from the National Household Survey of the Democratic Republic of the Congo (DRC) in 2012 by the National Institute of Statistics, covering 21454 households (National Institute of Statistics, 2012). This is a standardized 1-2-3 survey. Phase 1 is a general employment survey, providing detailed sociodemographic characteristics of households and employment. Phase 2 targets heads of informal production units, investigating their economic performance. Phase 3 is an expenditure survey, including aggregate data on the total expenditure of 12 categories of goods. Households are identified by the site number (national identification number for all districts and villages from 1 to n) and the number of households (coding households in one site).

The most important dataset for this study is module 11 of phase 3, which recorded households' daily expenditure on final goods within 15 days of the survey. Households can have more than one record of the purchase in module 11. Module 11 is a representative sample of 15005 households of the phase 1 sample. The classification of goods follows the COICOP12 nomenclature. Module 11 includes 9 categories of goods only, excluding education expenditure, hotel or restaurant expenditure, and miscellaneous goods and services. Data from module 11 is matched with demographic variables in phase 1. The summary statistic for the main demographic variables is shown in Table 1.

Table 2 below shows the share of expenditure of 12 categories of goods in phase 3 according to the quintile of total expenditure. If we classify informal and formal goods by categories as in Böhme and Thiele (2012), then the expenditure on informal goods like food and non-alcoholic beverages is decreasing in income, while formal goods like communication and transportation are increasing in income. The second way of classifying informal goods is by the distribution channel (Bachas, Gadenne, and Jensen, 2020). Since this paper explores the demand pattern for informal goods both in general and for decomposed categories, and substitutability is one of the main interests, classifying informal and formal goods by categories is not justified. Thus, this paper classifies informal and formal goods by distribution channels, using the indicator in module 11 shown in Table 3. Besides, to eliminate the possible subcontracting of formal goods to informal distribution channels, I adjust the channel indicator

DOI: https://doi.org/10.14324/111.444.2755-0877.1590

by classifying industrialized, imported goods, and consumption involving the usage of public infrastructures as formal goods. The full adjustment by product codes is included in Appendix 1.

Table 1: Summary Statistics for Selected Demographic Variables in Phase I

	Mean	Sd	Observations	Min	Max
Site	3504.126	2144.081	15005	10	7020
Menage	12.20187	8.445378	15005	1	40
Total Expenditure(in Congolese Franc)	1915017	4959132	15005	8200	5.48e+08
Number of Adults	4.262338	2.208481	15005	.72	17.88
Household Size	5.282706	2.73136	15005	1	22
Number of Members Aged Above 5	4.308697	2.37705	15005	1	18
Province	5.286638	3.154936	15005	1	11
Age (Head)	44.2217	14.24288	14989	15	98
Sex (Head) (1=Male)	.8125958	.3902487	15005	0	1
Years of Study (Head)	8.099571	5.001173	14462	0	21
Working Sector (Head)(1=formal)	.2095968	.4070344	15005	0	1
No Education (=1)	.1568144	.3636378	15005	0	1
Primary Education (=1)	.2069977	.4051673	15005	0	1
No formal Education (=1)	.006931	.0829665	15005	0	1
Secondary Education (=1)	.4893036	.4999022	15005	0	1
High Education (=1)	.1133622	.3170456	15005	0	1
Vocational Education (=1)	.0097967	.0984957	15005	0	1

Note: Only the sample of 15005 households is included here. Variables included in this table are used as control variables in later regression or used to generate control variables for later regressions. HSECTOR indicates the working sector of the household head, which equals 0 if the household head is working in the agricultural or non-agricultural informal sector, unemployed, retired, or inactive, and equals 1 if the head is working in the public sector or formal private sector. Details for generating other variables are shown in the do files provided. Total expenditure in this table includes the total household expenditure of 12 categories recorded in phase 3, including expenditure in memory and non-daily expenditure like education and hotel. Column Sd means standard deviation.

DOI: https://doi.org/10.14324/111.444.2755-0877.1590

Table 2: Average Budget Shares by Categories

	(1) First Quartile mean	(2) Second Quartile mean	(3) Third Quartile mean	(4) Fourth Quartile mean
1 Food/Non-alcoholic Beverages	.7443695	.7240466	.6870451	.597793
2 Alcoholic Beverages	.0185005	.0154278	.0152805	.017297
3 Clothing	.0205116	.024333	.0277453	.0307259
4 Housing	.1345526	.138374	.1606838	.1857187
5 Furnishing	.030115	.0296118	.0280484	.0342221
6 Health	.0121706	.0156174	.0145719	.0151105
7 Transportation	.004885	.0091243	.015843	.0356724
8 Communication	.002439	.0057898	.0090357	.0191471
9 Leisure	.0049803	.0054028	.0055001	.0085978
10 Education	.0092385	.0099588	.0119747	.018898
11 Hotel/Restaurant	.0033694	.0053112	.0066497	.0151312
12 Miscellaneous Goods/Services	.0148681	.0170025	.0176218	.0216863
Observations	3298	3772	3859	4076

Note: The quartile of total expenditure is classified using the variable QUART in phase 3. Details of codes and labels are available in the do files provided.

Table 3: Labels of the Channel Indicator in Module 11

Informal	Formal
01 Gift Given	07 Purchase in Supermarket
02 Gift Received	08 Purchase in a formal shop or Workshop(Company)
03 Self-produced Goods/Services	09 Purchase from the Public or Semi-public Sector
04 Purchase at Sellers' Home/ Small Shops/Informal Workshops	10 Other Formal Purchasing Place
05 Purchase on Public Markets	11 Purchasing Abroad
06 Other Informal/Independent Shops	

Note: The channel indicator is given in the original data of module 11. Details are available in the codebook and do files.

DOI: https://doi.org/10.14324/111.444.2755-0877.1590

5. Empirical Strategy

5.1. Quality, Price, and Unit Values

In practice, the most prevalent measure of quality for a group of goods is the unit value, which is the ratio of expenditure on a group of goods and the quantity purchased. In the household survey, price data is hard to collect, and there is a variation of price across villages which is particularly pervasive in most African countries, where markets are not integrated due to underdeveloped transportation (Deaton, 1987). The data of unit value is easier to obtain but cannot be a perfect substitute for price because it contains the effect of both price and quality. Unit values can be seen as the proxy for quality and price, but if we want to estimate price elasticity matrices and expenditure elasticity, it is crucial to rule out the quality effect and control for the variation across clusters.

The empirical method and codes are from Deaton (2018). The following sections show how this method can help separate the quality effect and recover price elasticities, control for regional variation of prices, and alleviate measurement error, simultaneity bias, and data limitation.

5.2. Modelling of Quality

According to Deaton (2018), the relationship between quality, quantity and price is defined as follows.

For a group of good G, the group quantity index QG is expressed by the household's vector of consumption levels of each item within the group, qG, and adjusted incommensurate items by kG,

$$Q_G = \boldsymbol{k_G q_G} \quad (1)$$

The corresponding unobserved price vector p_G is,

$$\boldsymbol{p_G} = \pi_G \boldsymbol{p_G^0} \quad (2)$$

where p_G^0 is the reference price vector which is constant across clusters. π_G is a scalar measuring the level of prices within group G, varying across clusters. The expenditure on group G is,

$$x_G = \boldsymbol{p_G q_G} = Q_G \pi_G \left(\frac{\boldsymbol{p_G^0 q_G}}{\boldsymbol{k_G q_G}} \right) \quad (3)$$

DOI: https://doi.org/10.14324/111.444.2755-0877.1590

The quality ξ is the expenditure of a bundle at p_G^0 relative to the physical volume Q_G,

$$\xi_G = \frac{\boldsymbol{p_G^0 q_G}}{\boldsymbol{k_G q_G}} \quad (4)$$

Thus,

$$x_G = Q_G \pi_G \xi_G \quad (5)$$

and,

$$\ln(x_G) = \ln(Q_G) + \ln(\pi_G) + \ln(\xi_G) \quad (6)$$

The unit value v_G is defined as the ratio of the expenditure of the goods and the physical quantity,

$$v_G = \frac{x_G}{Q_G} = \frac{\boldsymbol{p_G q_G}}{Q_G} = \xi_G \pi_G \quad (7)$$

Assuming separable utility across groups and subgroup demand functions are homogeneous of degree zero,

$$\boldsymbol{q_G} = f_G(x_G, \boldsymbol{p_G}) = f_G\left(\frac{x_G}{\pi_G}, \boldsymbol{p_G^0}\right) \quad (8)$$

Thus, the quality ξ_G depends on price π_G only through q_G, which in turn depends on the ratio x_G/π_G, by chain rule and rearrangements,

$$\frac{\partial ln \xi_G}{\partial ln \pi_G} = \frac{\partial ln \xi_G}{\partial ln x_G}\left[\frac{\partial ln x_G}{\partial ln \pi_G} - 1\right] \quad (9)$$

The last term in (9) in the bracket is the price elasticity of Q_G with respect to π_G, denoted by ϵ_P.

DOI: https://doi.org/10.14324/111.444.2755-0877.1590

Specify the basic clustered aggregate demand model for one good,

$$w_{Ghc} = \alpha^0 + \beta^0 lnx_{hc} + \gamma^0 z_{hc} + \theta ln\pi_{Ghc} + f_{Gc} + u^0_{Ghc} \quad (10)$$

$$lnv_{Ghc} = \alpha^1 + \beta^1 lnx_{hc} + \gamma^1 z_{hc} + \psi ln\pi_{Ghc} + u^1_{Ghc} \quad (11)$$

where w_{Ghc} and v_{Ghc} are the budget share and unit values of goods for household h in cluster c respectively. x_{hc} is the total expenditure for the household, π_{Ghc} is the unobserved price faced by the household and f_{Gc} is the cluster fixed effect. The assumption is that market prices vary across clusters but do not vary within them over the survey period. z_{hc} includes other demographic variables.

By (7), given π_G is the scalar vector on a cluster level, which is independent of the household's total income x, ignoring the household identifier subscript of hc for simplicity,

$$\frac{\partial lnv_G}{\partial lnx} = \frac{\partial ln\xi_G}{\partial lnx} + \frac{\partial ln\pi_G}{\partial lnx} = \frac{\partial ln\xi_G}{\partial lnx} \quad (12)$$

Then by (11) and (12), we can get the Prais-Houthakker income elasticity of quality, β^1 (Prais and Houthakker, 1971),

$$\beta^1 = \frac{\partial lnv_G}{\partial lnx} = \frac{\partial ln\xi_G}{\partial lnx} = \frac{\partial ln\xi_G}{\partial lnx_G}\frac{\partial lnx_G}{\partial lnx} \quad (13)$$

which identifies the relationship between income and quality. Intuitively, they should be positive since people are more willing to pay for quality as their income increases.

The last term in (13) is the expenditure elasticity of group G, ϵ_x, so we can express the first term in (13) by β^1 and ϵ_x and then substitute back to (9), and by (11), we can get the relationship between quality, price, and expenditure elasticity,

$$\psi = \frac{\partial lnv_G}{\partial ln\pi_G} = \frac{\partial ln\xi_G}{\partial ln\pi_G} + 1 = \frac{\beta^1 \epsilon_p}{\epsilon_x} + 1 \quad (14)$$

DOI: https://doi.org/10.14324/111.444.2755-0877.1590

By (10), the budget share elasticity with respect to *x* is β^0/w, which includes both the quality and quantity effect of income. By the definition of the budget share, (4), (6) and (12),

$$\frac{\beta^0}{w_G} = \frac{\partial lnw_G}{\partial lnx} = \frac{\partial \ln (v_G Q_G/x)}{\partial lnx} = \beta^1 + \epsilon_x - 1 \quad (15)$$

Then similarly, by (11) and (14), and the definition of price elasticity ϵ_p,

$$\epsilon_p + \psi = \frac{\theta}{w_G} = \frac{\partial lnw_G}{\partial ln\pi} = \frac{\partial \ln (v_G Q_G/x)}{\partial ln\pi} \quad (16)$$

Then by (14), and by substituting (15) and (16) in,

$$\psi = 1 - \frac{\beta^1 (w_G - \theta)}{\beta^0 + w_G} \quad (17)$$

Suppose $\phi = \theta/\psi$, which can be estimated. Given that θ can be recovered by,

$$\theta = \frac{\phi}{1 + (w_G - \phi)\zeta} \quad (18)$$

In which,

$$\zeta = \frac{\beta^1}{\beta^0 + w_G(1 - \beta^1)} \quad (19)$$

After parameters in (17)-(19) are estimated, we can recover the elasticity in (15) and (16).

5.3. Two Stages Estimation

The method below from Deaton (2018) is slightly adapted to use budget share and add demographic factors.

There are two stages; in the first stage, we run the regressions (10) and (11) without the price data and then construct the two variables below, ignoring the subscript of *G* for simplicity,

$$\hat{y}^0{}_{hc} = lnw_{hc} - \hat{\beta}^0 lnx_{hc} - \hat{\gamma}^0 z_{hc} \quad (20)$$

DOI: https://doi.org/10.14324/111.444.2755-0877.1590

$$\hat{y}^1{}_{hc} = lnv_{hc} - \hat{\beta}^1 lnx_{hc} - \hat{\gamma}^1 z_{hc} \quad (21)$$

In (20) and (21), the effect of expenditure is filtered out while the cluster fixed effect and price effect are kept. We can get estimates for the variance of u^1_{hc} and covariance of u^1_{hc} and u^0_{hc}, denoted by $\hat{\sigma}^{11}$ and $\hat{\sigma}^{01}$.

In the second stage, we use cluster-level information and the subscript c only, the true average values y_c^0 and y_c^1 can be expressed as

$$y_c^0 = \alpha^0 + \theta ln\pi_c + f_c + u_c^0 \quad (22)$$

$$y_c^1 = \alpha^1 + \psi ln\pi_c + u_c^1 \quad (23)$$

Then,

$$cov(y_c^0, y_c^1) = \theta\psi m + \frac{\sigma^{01}}{n_c} \quad (24)$$

$$var(y_c^1) = \psi^2 m + \frac{\sigma^{11}}{n_c} \quad (25)$$

where m is the variance of log prices in a large sample and n_c is the number of households per cluster.

Given (22)-(25), the estimate for ϕ is denoted by,

$$\hat{\phi} = \frac{cov(\hat{y}_c^0, \hat{y}_c^1) - \frac{\hat{\sigma}^{01}}{n_c}}{var(\hat{y}_c^1) - \frac{\widehat{\sigma^{11}}}{n_c}} \quad (26)$$

Then, substituting this back to (18), we can recover θ and obtain ψ from (17). Finally, substituting them back to (16), we can get estimates of price elasticities.

DOI: https://doi.org/10.14324/111.444.2755-0877.1590

5.4. Final Regressions

The theoretical basis for the multi-good utility model with quality effects is that the consumer maximizes the sub-utility group subject to the amount spent on each group, assuming the overall utility function is separable in every group. The main stages of the estimation follow the same steps in the last sections, but the estimated parameters and price elasticity are matrices. The smallest unit of groups, the variable SITE in Table 1 is used as clusters.

Our final aggregate demand system for *M* groups of goods is,

$$w_{Ghc} = \alpha_G^0 + \beta_G^0 lnx_{hc} + \gamma_G^0 z_{hc} + \sum_{H=1}^{M} \theta_{GH} ln\pi_{Hc} + f_{Gc} + u_{Ghc}^0 \quad (27)$$

$$lnv_{Ghc} = \alpha_G^1 + \beta_G^1 lnx_{hc} + \gamma_G^1 z_{hc} + \sum_{H=1}^{M} \psi_{GH} ln\pi_{Hc} + u_{Ghc}^1 \quad (28)$$

The description of variables is shown in Table 4. Demographic variables z_{hc} includes household sizes, sex, age, years of schooling, and working sector of household heads, which are controlled to alleviate the potential correlated heterogeneity between total expenditure and budget shares or unit values. The fixed effect of provinces is removed in the estimation and the demand system is estimated for the full sample of module 11 and decomposed categories. Details are shown in Table 5.

DOI: https://doi.org/10.14324/111.444.2755-0877.1590

Table 4: Description of Variables in Regressions

Variable	Description
w_{Ghc}	Budget share of group G for household h in cluster c
lnv_{Ghc}	Unit values of group G for household h in cluster c in natural logarithm
lnx_{hc}	Total expenditure of household h in cluster c in natural logarithm
$lhhs_{hc}$	Household size of household h in cluster c in natural logarithm
SEXHEAD	Sex of the household head: Males=1, Female=0
YEAR_STUDY	Years of the household head in school or vocational training
HSECTOR	Working sector of the household head
AGEHEAD	Age of the household head
Province	Province of the household, labeled from 1 to 11

Note: The summary statistics of all demographic variables can be found in Table 1. The summary statistics for budget share and unit values are not provided here because their statistics vary across different categories. $lhhs$ is the natural logarithm of the variable 'Household Size' in table 1. lnx_{hc} is the natural logarithm of the total expenditure in table 1.

DOI: https://doi.org/10.14324/111.444.2755-0877.1590

Table 5: Description of Groups of Goods for Estimations

	Product Code	Description
A	All	Full sample of module 11 using adjusted product codes
B	11***	All foods
C	111**	Cereals, pastas, staples
D	112**, 113**	Meat, seafood
E	114**, 115**, 118**, 119**	Diary products, fat, sugar, sauce
F	116**, 117**	Fruit, vegetable
G	12***	Non-alcoholic beverage
H	2****	Alcoholic beverage
I	3****	Clothing
J	4****, 5****s	Furnishing, housing/decoration
K	6****	Health
L	7****	Transportation
M	9****	Communication (postal/telephone), leisure
N	All	Full sample using the original product code

Note: The channel indicator is given in the original data of module 11. Details are available in the codebook and do files. The adjusted channel indicator is used for all estimations, except N. The description of the adjusted channel indicator is available in section V and Appendix 1. Product codes provided gave 5 digits. The first digit from the left-hand side represents the category indicated in table 1. For example, 2**** in this table corresponds to 2 in table 1 and indicates alcoholic beverages. The second and third digit indicates narrower categories. For example, 11*** means food only and 111** indicates cereals, pasta, and staples only.

DOI: https://doi.org/10.14324/111.444.2755-0877.1590

6. Limitations and Further Adjustment of Data

6.1. Unit Values, Services, and Groups of Goods for Regressions

First, unit values are considered a better proxy for prices of foods than other categories of goods. To alleviate the measurement error in the calculation of unit values, I remove records of nonquantifiable services in module 11 because the definition of physical quantity for services is vague. The code and descriptions for services are available in Appendix 1. Besides, in Table 3, the informal channels recorded in module 11 include gifts given and gifts received, and to alleviate measurement errors, records from these two channels are removed.

Second, in household survey data, zero purchases of certain groups of goods are common. If both zero and non-zero purchases are included, then the method can be thought of as a linear approximation averaging over zero and non-zero purchases. However, in this study, we include non-zero purchases only because the data on the unit value of zero purchases is omitted. In this case, the results are estimated conditional on non-zero purchases. Thus, here we are interested in why households buy more or less of goods rather than why make a purchase, assuming that the purchases are randomly distributed with time. Using the median of unit values and budget shares for the missing records is usually a way to alleviate the problem of zero purchases. I impute the median values conditional on the demographic variables shown in Table 4, assuming that the recording is missing randomly conditional on those variables. This assumption is untestable. Both estimates with imputation and without imputation are provided and compared, but the results without imputation are the key to the discussion.

Besides, the functional form of the budget share regression assumes a linear Engel curve, but it can be quadratic. Furthermore, Gibson and Kim (2019) suggest that this method still underestimates the quality effect and overestimates the price elasticity of quantity. To further alleviate the simultaneity and measurement errors, instrumental variables for the expenditure terms can be considered. A comparison of existing empirical strategies for the demand systems is shown in Appendix 2.

6.2. Symmetry Restriction, Completing the System, and Statistical Inference

First, in survey data, given zero purchases and limited time of collection, there can be hundreds of different purchase combinations of goods, thus empirical specification cannot fully comply with the theoretical restriction of symmetry. Nonetheless, Deaton (2018) proposed a method to approach the symmetry constraint. Assuming that the quality effects are small, there is a way to obtain a price elasticity matrix that is symmetric in signs.

Second, we can hardly obtain full consumption lists of consumers. Deaton (2018) suggests that the adding-up and homogeneity restriction of the demand system allows us to add a fictitious good to complete the system. If we want to focus on several groups of goods only, all other goods are included in the fictitious good.

The expenditure elasticity should be interpreted cautiously because the method provides no standard error for it, but if estimates for β^0 or β^1 are not significant, it is hardly possible that the estimate for expenditure elasticity is significant. In the second stage, the standard errors for price elasticity matrices are obtained by bootstrapping, making 1000 draws from the clusters in the second stage, which is defined

DOI: https://doi.org/10.14324/111.444.2755-0877.1590

as half of the interval that contains 68.3 per cent of the bootstrapped estimates. The estimate for price elasticity is regarded as significant if it is larger than twice the bootstrapped standard error.

7. Results

7.1. Empirical Results

This section discusses the result of groups A, C to H, I, and L in Table 5. Other results are shown in Appendix 3.

A: Full Sample

First, full sample results can give insight into the general linkage between informal and formal goods in the final product market. Table 6 shows the first-stage regression results with total expenditure elasticity, ϵx, and quality elasticity of income, $\beta 1$, for the estimation without and with the imputation of missing values.

According to equation (15), the positive quality elasticity of income is filtered out from the budget share elasticity, so the expenditure elasticity can be lower than estimates without considering the quality effects. The coefficient of lnx in both estimations is positive for formal goods while negative for informal goods, suggesting that the consumers would spend less budget on informal goods as total expenditure increases while more budget on formal goods. In Appendix 4, the linear Engel curves are estimated using the first-stage results. Ignoring quality effects, if the coefficient lnx is positive, we may expect that the expenditure elasticity is above the unity and the quantity demanded is also increasing in income, but the expenditure elasticity for formal and informal goods are all less than unity. The reason behind this is that the positive and significant quality effects of income, $\beta 1$, are filtered out. The quality elasticity, $\beta 1$, of income is positive for both informal and formal goods and higher for formal goods, implying that the quality of goods purchased increases as their total expenditure increases and the pace is quicker for the formal goods.

Another interesting coefficient to look at is that of the household size, lhhs. The coefficient of lhhs is negative for formal goods but positive for informal ones, suggesting that ceteris paribus on average, households with larger size is more likely to choose informal goods.

Table 7 shows the estimations of price elasticity. Since the quality elasticity is significant, as discussed in section VI.2, it is better to look at the unconstrained estimates of price elasticity. For all the table of price elasticity, the figures in bold are significant. The price elasticity for both formal and informal goods is negative and significant. The cross-price elasticity shows that informal and formal goods are substitutes but those estimates are almost significant. The pattern is consistent for estimates with and without the symmetry constraint. To summarize, in general, the budget share for informal goods decreases as the total expenditure level increases, and informal and formal goods are likely to be imperfect substitutes, as suggested by the first channel.

DOI: https://doi.org/10.14324/111.444.2755-0877.1590

Table 6: A: Results of Full Sample (Adjusted Channel Indicator): First Stage

	Budget Share Regressions			
	Without Imputation		With Imputation	
	Formal	Informal	Formal	Informal
lnx	0.000640***	-0.00307***	0.000583***	-.00306***
	(0.000)	(0.000)	(0.000)	(0.000)
$lhhs$	-0.000225***	0.000649***	-.000273***	0.000639 ***
	(0.001)	(0.000)	(0.000)	(0.000)
R^2	0.3669	0.5169	0.3452	0.5160
N	8895	14313	9494	14319
	Unit Value Regressions			
	Without Imputation		With Imputation	
β^1	0.585***	0.392***	0.557***	0.392***
	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.4137	0.7049	0.3772	0.7047
N	8895	14313	9494	14319
ϵ_x	0.631	0.490	0.640	0.490

p-values in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: In this section, the result tables will include only the goods we are interested in, the 'other goods' generated for completing the system are not shown here but full results can be generated by running the code. Estimates of other coefficients are available on requests.

Table 7: A: Results of Full Sample: Price Elasticity

	No Symmetry Restriction			
	With Imputation		Without Imputation	
	Formal	Informal	Formal	Informal
Formal	**-0.661**	-0.115	**-0.621**	-0.157
	(0.0586)	(0.120)	(0.0627)	(0.119)
Informal	0.0752	**-0.771**	0.0819	**-0.769**
	(0.0384)	(0.0831)	(0.0473)	(0.0816)
	With Symmetry Restriction			
	Formal	Informal	Formal	Informal
Formal	**-0.766**	0.515	**-0.745**	0.530
	(0.0740)	(0.295)	(0.0890)	(0.340)
Informal	0.0605	**-0.759**	0.0619	**-0.755**
	(0.00341)	(0.0856)	(0.0392)	(0.0815)

Note: Bootstrapped standard error in parentheses. Price elasticity is significant if in bold (The standard errors for price elasticity matrices are obtained by bootstrapping, making 1000 draws from the clusters in the second stage, which is defined as half of the interval that contains 68.3 percent of the bootstrapped estimates. The estimate for price elasticity is regarded as significant if it is larger than twice the bootstrapped standard error.). There are four two-times-two matrices of price elasticity in the table. The entries with different names on rows and columns are cross-price elasticity, which is positive for substitutes and negative for complements.

DOI: https://doi.org/10.14324/111.444.2755-0877.1590

C: Staples

Table 8 shows both informal and formal staples are necessities while the expenditure elasticity is slightly higher for the informal goods, which may contribute to the higher quality elasticity of formal staples. This may suggest that consumers choose formal staples for quality more but informal staples for quantity more as total expenditure increases. The cross-price elasticity is not pinned down without imputation; the results with imputation show that formal and formal goods are complements as the third channel.

Table 8: C: Results by Categories: Staples

Budget Share Regressions				
	No Imputation		With Imputation	
	Formal	Informal	Formal	Informal
lnx	-0.004523***	-0.000902***	-0.00432***	-0.000896***
	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.4670	0.5917	0.4402	0.5589
N	4884	12462	5883	13635
Unit Value Regressions				
	No imputation		With Imputation	
β^1	0.209***	0.167***	0.187***	0.158***
	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.5183	0.5708	0.4757	0.5374
N	4884	12462	5883	13635
ϵ_x	0.544	0.629	0.583	0.639

p-values in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Price Elasticity				
No Symmetry Restriction				
	No Imputation		Without Imputation	
	Formal	Informal	F	I
Formal	**-0.607**	-0.0431	**-0.584**	**-0.183**
	(0.057)	(0.046)	(0.12)	(0.060)
Informal	-0.0264	**-0.670**	**-0.241**	**-0.378**
	(0.05)	(0.048)	(0.059)	(0.064)
With Symmetry Restriction				
	Formal	Informal	Formal	Informal
Formal	**-0.604**	-0.0550	**-0.546**	**-0.370**
	(0.057)	(0.070)	(0.12)	(0.08)
Informal	-0.0228	**-0.691**	**-0.157**	**-0.394**
	(0.029)	(0.046)	(0.03)	(0.06)

Note: In the regressions for decomposed categories, the channel indicator used is the adjusted channel indicator. Bootstrapped standard error in parentheses. Price elasticity is significant if in bold (The price elasticity is larger than twice the bootstrapped standard errors in magnitude).

DOI: https://doi.org/10.14324/111.444.2755-0877.1590

D: Meat and Seafood

Table 9 shows both informal and formal meat are necessities while the expenditure elasticity is slightly higher for formal goods. The higher quality elasticity for informal goods may suggest that consumers choose informal meat for quality more as total expenditure increases. The negative cross-price elasticity with symmetry constraint suggests that informal and formal goods are complements but they are not significant.

Table 9: D: Results by Categories: Meat and Seafood

	Budget Share Regressions			
	No Imputation		With Imputation	
	Formal	Informal	Formal	Informal
lnx	-0.000312	-0.000557***	-0.000286	-0.000507***
	(0.255)	(0.000)	(0.256)	(0.000)
R^2	0.6645	0.4260	0.6936	0.4090
N	620	13911	787	14971
	Unit Value Regressions			
	No Imputation		With Imputation	
β^1	0.359***	0.478***	0.338***	0.450***
	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.7878	0.6738	0.7726	0.6520
N	620	13911	787	14971
ϵ_x	0.538	0.442	0.563	0.478

p-values in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

	Price Elasticity			
	No Symmetry Restriction			
	No Imputation		With Imputation	
	Formal	Informal	Formal	Informal
Formal	**-0.915**	0.0743	**-0.974**	0.0186
	(0.071)	(0.069)	(0.095)	(0.010)
Informal	-0.0389	**-0.567**	-0.0252	**-0.659**
	(0.049)	(0.094)	(0.041)	(0.067)
	With Symmetry Restriction			
	Formal	Informal	Formal	Informal
Formal	**-0.894**	-0.0451	**-0.967**	-0.0350
	(0.069)	(0.082)	(0.10)	(0.074)
Informal	-0.0196	**-0.576**	-0.0146	**0.662**
	(0.035)	(0.10)	(0.03)	(0.067)

In the regressions for decomposed categories, the channel indicator used is the adjusted channel indicator. Bootstrapped standard error in parentheses. Price elasticity is significant if in bold (The price elasticity is larger than twice the bootstrapped standard errors in magnitude).

DOI: https://doi.org/10.14324/111.444.2755-0877.1590

E: Dairy, Oil, Sugar and Sauce

Table 10 shows that, for dairy products, oil, sugar, and sauce, informal goods have higher expenditure elasticity and lower quality elasticity. One unconstrained estimate of price elasticity without imputation shows that they are complements as in channel three. This pattern may indicate that informal goods in this market may be preferred by other characteristics like traditional tastes or accessibility which helps sustain their demand as total expenditure increases.

Table 10: E: Results by Categories: Dairy, Oil, Sugar, Sauce

Budget Share Regressions				
	No Imputation		With Imputation	
	Formal	Informal	Formal	Informal
lnx	-0.0003625***	-0.00105***	-0.000323***	-0.000981***
	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.4156	0.5057	0.4136	0.4864
N	3546	14222	4305	15259
Unit Value Regressions				
	No Imputation		With Imputation	
β^1	0.494***	0.300***	0.463***	0.283***
	(0.000)	(0.000)	(0.0000)	(0.000)
R^2	0.5026	0.5885	0.4808	0.5673
N	3546	14222	4305	15259
ϵ_x	0.203	0.418	0.269	0.454

p-values in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Price Elasticity				
No Symmetry Restriction				
	No imputation		With Imputation	
	Formal	Informal	Formal	Informal
Formal	**-0.313**	**-0.411**	**-0.614**	-0.120
	(0.14)	(0.13)	(0.15)	(0.11)
Informal	0.0351	**-0.763**	-0.0702	**-0.839**
	(0.062)	(0.08)	(0.052)	(0.059)
With Symmetry Restriction				
	Formal	Informal	Formal	Informal
Formal	**-0.416**	-0.00933	**-0.600**	-0.189
	(0.12)	(0.15)	(0.122)	(0.12)
Informal	-0.00304	**-0.729**	-0.0614	**-0.844**
	(0.05)	(0.066)	(0.038)	(0.055)

In the regressions for decomposed categories, the channel indicator used is the adjusted channel indicator. Bootstrapped standard error in parentheses. Price elasticity is significant if in bold (The price elasticity is larger than twice the bootstrapped standard errors in magnitude).

DOI: https://doi.org/10.14324/111.444.2755-0877.1590

F: Vegetables and Fruits

Table 11 shows that for vegetables and fruits, formal goods have higher expenditure elasticity, and the quality elasticity is similar for informal and formal goods. The positive estimates of cross-price elasticity in all estimations show that they are imperfect substitutes, though they are insignificant. This may be consistent with the first channel, suggesting that informal goods may be preferred as imperfect substitutes because of income inequality and their budget share decreases much more quickly than formal goods as total expenditure increases.

Table 11: F: Results by Categories: Vegetable and Fruits

	Budget Share Regressions			
	No Imputation		With Imputation	
	Formal	Informal	Formal	Informal
lnx	-0.000235	-0.00279***	-0.000117	-0.00263***
	(0.18)	(0.000)	(0.481)	(0.000)
R^2	0.7392	0.5947	0.7334	0.5555
N	505	13829	621	14908
	Unit Value Regressions			
	No Imputation		With Imputation	
β^1	0.166*	0.185***	0.224**	0.173***
	(0.073)	(0.000)	(0.012)	(0.000)
R^2	0.7445	0.6741	0.7419	0.6443
N	505	13829	621	14908
ϵ_x	0.649	0.442	0.681	0.475

p-values in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

	Price Elasticity			
	No Symmetry Restriction			
	No imputation		With Imputation	
	Formal	Informal	Formal	Informal
Formal	**-0.858**	0.136	**-0.867**	0.145
	(0.14)	(0.15)	(0.18)	(0.21)
Informal	0.0628	**-0.594**	0.109	**-0.752**
	(0.092)	(0.087)	(0.11)	(0.082)
	With Symmetry Restriction			
	Formal	Informal	Formal	Informal
Formal	**-0.882**	0.322	**-0.917**	0.523
	(0.17)	(0.40)	(0.21)	(0.42)
Informal	0.0553	**-0.590**	0.0872	**-0.744**
	(0.069)	(0.081)	(0.07)	(0.075)

In the regressions for decomposed categories, the channel indicator used is the adjusted channel indicator. Bootstrapped standard error in parentheses. Price elasticity is significant if in bold (The price elasticity is larger than twice the bootstrapped standard errors in magnitude).

DOI: https://doi.org/10.14324/111.444.2755-0877.1590

G: Non-alcoholic Beverage

Table 12 shows that informal and formal goods have similar expenditure elasticity and constrained estimates with imputation suggest that they are imperfect substitutes. The coefficient of lnx is less negative for informal goods, suggesting their budget share decreases more slowly than formal goods as total expenditure increases. These may suggest that demand for informal goods may come from the second channel and informal goods may be preferred for other properties. The results may be interpreted with caution given the limitations of imputation.

Table 12: G: Results by Categories: Non-alcoholic Beverage

Budget Share Regressions				
	No Imputation		With Imputation	
	Formal	Informal	Formal	Informal
lnx	-0.000231***	-0.000106***	-0.000219***	-0.0000665***
	(0.000)	(0.000)	(0.000)	(0.021)
R^2	0.3148	0.3529	0.3463	0.3277
N	1872	9873	2284	11156
Unit Value Regressions				
	No Imputation		With Imputation	
β^1	0.264***	0.426***	0.236***	0.394***
	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.4191	0.4543	0.4306	0.4217
N	1872	9873	2284	11156
ϵ_x	0.465	0.473	0.515	0.543

p-values in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Price Elasticity				
No Symmetry Restriction				
	No imputation		With Imputation	
	Formal	Informal	Formal	Informal
Formal	-0.157	0.0906	**-0.524**	0.114
	(0.20)	(0.12)	(0.24)	(0.11)
Informal	0.169	**-0.189**	**0.406**	**-0.453**
	(0.16)	(0.094)	(0.21)	(0.12)
With Symmetry Restriction				
	Formal	Informal	Formal	Informal
Formal	-0.170	0.157	**-0.586**	**0.286**
	(0.18)	(0.14)	(0.24)	(0.13)
Informal	0.128	-0.177	**0.238**	**-0.402**
	(0.11)	(0.09)	(0.11)	(0.10)

In the regressions for decomposed categories, the channel indicator used is the adjusted channel indicator. Bootstrapped standard error in parentheses. Price elasticity is significant if in bold (The price elasticity is larger than twice the bootstrapped standard errors in magnitude).

DOI: https://doi.org/10.14324/111.444.2755-0877.1590

H: Alcoholic Beverage

Table 13 shows a pattern similar to that of non-alcoholic beverages, except that the coefficient of lnx for formal beverages is positive but not significant. The expenditure elasticity without quality effects is similar for informal and formal goods. The unconstrained estimates with imputation show a significant and positive cross-price elasticity, suggesting that they are imperfect substitutes.

Table 13: H: Results by Categories: Alcoholic Beverage

	Budget Share Regressions			
	No Imputation		With Imputation	
	Formal	Informal	Formal	Informal
lnx	0.0000781	-0.000137**	0.0000917	-0.0000785
	(0.675)	(0.030)	(0.627)	(0.187)
R^2	0.5684	0.3341	0.5821	0.2862
N	584	4696	687	5439
	Unit Value Regressions			
	No Imputation		With Imputation	
β^1	0.585***	0.471***	0.548***	0.429***
	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.7791	0.5696	0.7472	0.5082
N	584	4696	687	5349
ϵ_x	0.466	0.449	0.508	0.525

p-values in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

	Price Elasticity			
	No Symmetry Restriction			
	No imputation		With Imputation	
	Formal	Informal	Formal	Informal
Formal	**-0.674**	0.166	**-0.695**	**0.339**
	(0.12)	(0.11)	(0.10)	(0.12)
Informal	0.0176	**-0.762**	0.00787	**-0.749**
	(0.087)	(0.08)	(0.073)	(0.073)
	With Symmetry Restriction			
	Formal	Informal	Formal	Informal
Formal	**-0.656**	0.0722	**-0.698**	0.121
	(0.011)	(0.074)	(0.10)	(0.067)
Informal	0.0654	**-0.778**	0.116	**-0.745**
	(0.066)	(0.08)	(0.064)	(0.072)

In the regressions for decomposed categories, the channel indicator used is the adjusted channel indicator. Bootstrapped standard error in parentheses. Price elasticity is significant if in bold (The price elasticity is larger than twice the bootstrapped standard errors in magnitude).

DOI: https://doi.org/10.14324/111.444.2755-0877.1590

I: Clothing

Table 14 shows that the coefficient of lnx for informal goods is positive, and expenditure and quality elasticities of income are higher for informal goods. We may expect informal goods to be preferred for other attributes, helping their demand to be more sustainable. However, estimates of price elasticity are not significant due to the limitations in using unit values for non-foods as a proxy for prices.

Table 14: I: Results by Categories: Clothing

	Budget Share Regressions			
	No Imputation		With Imputation	
	Formal	Informal	Formal	Informal
lnx	-0.000848*	0.000414	-0.000659	0.00298
	(0.072)	(0.168)	(0.128)	(0.300)
R^2	0.6355	0.4452	0.6898	0.4261
N	450	2327	540	2743
	Unit Value Regressions			
	No Imputation		With Imputation	
β^1	0.386***	0.516***	0.388***	0.434***
	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.6907	0.4033	0.7029	0.3590
N	450	2327	540	2743
ϵ_x	0.421	0.579	0.466	0.630

p-values in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

	Price Elasticity			
	No Symmetry Restriction			
	No imputation		With Imputation	
	Formal	Informal	Formal	Informal
Formal	-0.511	0.0536	-0.290	-0.218
	(0.55)	(0.25)	(0.66)	(0.48)
Informal	0.309	-0.448	0.241	-0.253
	(0.91)	(0.43)	(0.54)	(0.45)
	With Symmetry Restriction			
	Formal	Informal	Formal	Informal
Formal	-0.554	0.173	-0.348	0.0407
	(1.13)	(0.58)	(1,12)	(0.55)
Informal	0.172	-0.404	0.0383	-0.197
	(0.58)	(0.39)	(0.53)	(0.41)

In the regressions for decomposed categories, the channel indicator used is the adjusted channel indicator. Bootstrapped standard error in parentheses. Price elasticity is significant if in bold (The price elasticity is larger than twice the bootstrapped standard errors in magnitude).

DOI: https://doi.org/10.14324/111.444.2755-0877.1590

L: Transportation

Table 15 shows that the coefficient of lnx is positive for both informal and formal goods, suggesting that the budget for transportation is increasing in total expenditure. The lower quality elasticity of income and higher expenditure elasticity for informal goods suggest that consumers may prefer informal transportation more for quantity as total expenditure increases. The estimates indicate that informal and formal transportation are complements, but these are not significant.

Table 15: L: Results by Categories: Transportation

	Budget Share Regressions			
	No Imputation		With Imputation	
	Formal	Informal	Formal	Informal
lnx	0.000535*	0.000245*	0.000339	0.000299**
	(0.089)	(0.082)	(0.277)	(0.024)
R^2	0.4938	0.5743	0.5000	0.5388
N	366	2219	432	2639
	Unit Value Regressions			
	No Imputation		With Imputation	
β^1	0.518***	0.306***	0.451***	0.323***
	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.8092	0.6151	0.7748	0.5841
N	366	2219	432	2639
ϵ_x	0.772	0.811	0.739	0.817

p-values in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

	Price Elasticity			
	No Symmetry Restriction			
	No imputation		With Imputation	
	Formal	Informal	Formal	Informal
Formal	**-0.472**	-0.010	**-0.596**	-0.0310
	(0.14)	(0.056)	(0.17)	(0.10)
Informal	-0.304	0.324	-0.155	0.0493
	(0.22)	(0.25)	(0.16)	(0.24)
	With Symmetry Restriction			
	Formal	Informal	Formal	Informal
Formal	**-0.411**	-0.159	**-0.564**	-0.108
	(0.14)	(0.10)	(0.15)	(0.10)
Informal	-0.139	0.270	-0.0901	0.0226
	(0.087)	(0.24)	(0.09)	(0.21)

In the regressions for decomposed categories, the channel indicator used is the adjusted channel indicator. Bootstrapped standard error in parentheses. Price elasticity is significant if in bold (The price elasticity is larger than twice the bootstrapped standard errors in magnitude).

DOI: https://doi.org/10.14324/111.444.2755-0877.1590

7.2 Discussion

Based on the result for the full sample, we can know that on average, the informal sector shrinks as the aggregate income level increases. Before the income level increases to a certain threshold, policies to promote formality from the supply side may not be as effective as they were expected to be because the large demand for low-quality and cheaper informal goods resulting from income inequality attract potential entrepreneurs into the informal sector. For markets where demand for informal goods follows the second and third channels, the demand may sustain for a longer time as the economy develops. In this case, the policy to improve formality can focus more on reducing the registration costs and regulation costs of traditional shops or informal firms, allowing them to maintain their production as formal firms, rather than pushing their transformation into industrialized producers.

8. Results

This paper studies the demand pattern for informal goods in Africa using household-level survey data from the Democratic Republic of the Congo in 2012. There are three major factors driving the demand for informal goods: income inequality, quality preference, and preference for product-specific characteristics. I hypothesize that there are mainly three channels of demand for informal goods and use the empirical method proposed by Deaton (2018) to test them by estimating expenditure elasticity, price elasticity, and quality elasticity. The empirical results suggest that budget shares for informal goods decrease as total expenditure increases. The pattern can be different for different categories of goods. Informal goods and formal goods can be either imperfect substitutes or complements in the final good markets, and the demand for some informal goods can sustain as income increases because they are preferred for their own features. Policies to improve formality can adjust according to the demand pattern in different markets to improve policy effectiveness. This paper also discussed how the empirical method is conducted under assumptions and is subject to several limitations, which may be improved in the future.

DOI: https://doi.org/10.14324/111.444.2755-0877.1590

References

Bachas, P., Gadenne, L. and Jensen, A. (2020). Informality, Consumption Taxes and Redistribution. National Bureau of Economic Research, Working Paper Series, 27429. doi:https://doi.org/10.3386/w27429.

Banerji, A. and Jain, S. (2007). Quality dualism. Journal of Development Economics, 84(1), pp.234–250. doi:https://doi.org/10.1016/j.jdeveco.2005.09.010.

Bargain, O. and Kwenda, P. (2014). The Informal Sector Wage Gap: New Evidence Using Quantile Estimations on Panel Data. Economic Development and Cultural Change, 63(1), pp.117–153. doi:https://doi.org/10.1086/677908.

Benjamin, N.C. and Mbaye, A.A. (2012). The Informal Sector, Productivity, and Enforcement in West Africa: A Firm-level Analysis. Review of Development Economics, 16(4), pp.664–680. doi:https://doi.org/10.1111/rode.12010.

Böhme, M. and Thiele, R. (2012). Is the Informal Sector Constrained from the Demand Side? Evidence for Six West African Capitals. World Development, 40(7), pp.1369–1381. doi:https://doi.org/10.1016/j.worlddev.2011.12.005.

Copeland, B.R. and Kotwal, A. (1996). Product quality and the theory of comparative advantage. European Economic Review, 40(9), pp.1745–1760. doi:https://doi.org/10.1016/0014-2921(95)00061-5.

Dabla-Norris, E., Gradstein, M. and Inchauste, G. (2008). What causes firms to hide output? The determinants of informality. Journal of Development Economics, 85(1-2), pp.1–27. doi:https://doi.org/10.1016/j.jdeveco.2006.05.007.

de Paula, Á. and Scheinkman, J.A. (2007). The Informal Sector. [online] National Bureau of Economic Research. Available at: http://www.nber.org/papers/w13486 [Accessed 18 Feb. 2023].

Deaton, A. (1987). Estimation of own- and cross-price elasticities from household survey data. Journal of Econometrics, 36(1-2), pp.7–30. doi:https://doi.org/10.1016/0304-4076(87)90041-8.

Deaton, A. and Muellbauer, J. (1980). An Almost Ideal Demand System. The American Economic Review, [online] 70(3), pp.312–326. Available at: http://www.jstor.org/stable/1805222 [Accessed 19 Feb. 2023].

Deaton, A. (2018). The analysis of household surveys: a microeconometric approach to development

policy, World Bank Group

Diao, X., Ellis, M., McMillan, M. and Rodrik, D. (2021). Africa&Apos;S Manufacturing Puzzle: Evidence from Tanzanian and Ethiopian Firms. SSRN Electronic Journal. doi:https://doi.org/10.2139/ssrn.3772595.

Etim, E. and Daramola, O. (2020). The Informal Sector and Economic Growth of South Africa and Nigeria: A Comparative Systematic Review. Journal of Open Innovation: Technology, Market, and Complexity, 6(4), p.134. doi:https://doi.org/10.3390/joitmc6040134.

Gibson, J. and Kim, B. (2019). Quality, quantity, and spatial variation of price: Back to the bog. Journal of Development Economics, 137, pp.66–77. doi:https://doi.org/10.1016/j.jdeveco.2018.11.008.

Hernando De Soto (2006). The mystery of capital : why capitalism triumphs in the West and fails everywhere else. New York: Basic Books, [], Cop.

Kruse, H., Mensah, E., Sen, K. and de Vries, G. (2022). A Manufacturing (Re)Naissance? Industrialization in the Developing World. IMF Economic Review. doi:https://doi.org/10.1057/s41308-022-00183-7.

Lecocq, S. and Robin, J.-M. (2015). Estimating Almost-ideal Demand Systems with Endogenous Regressors. The Stata Journal: Promoting communications on statistics and Stata, 15(2), pp.554–573. doi:https://doi.org/10.1177/1536867x1501500214.

McCaig, B. and Pavcnik, N. (2018). Export Markets and Labor Allocation in a Low-Income Country. American Economic Review, 108(7), pp.1899–1941. doi:https://doi.org/10.1257/aer.20141096.

McMillan, M. and Zeufack, A. (2022). Labor Productivity Growth and Industrialization in Africa. Journal of Economic Perspectives, 36(1), pp.3–32. doi:https://doi.org/10.1257/jep.36.1.3.

DOI: https://doi.org/10.14324/111.444.2755-0877.1590

Medina, L., Jonelis, Mr.Andrew.W. and Cangul, M. (2017). The Informal Economy in Sub-Saharan Africa. Working Paper No. 2017/156 ed. International Monetary Fund.

Mishra, A. (2022). What Sustains Informality? A Study of the Interactions between Formal and Informal Sector Firms. The Journal of Development Studies, Vol.58(No.7), pp.1–13. doi:https://doi.org/10.1080/00220388.2022.2061853.

Mishra, A. and Ray, R. (2010). Informality, Corruption, and Inequality. Bath Economics Research Working Papers, no.13/10. Department of Economics, University of Bath, Bath, U.K.

National Institute of Statistics (2012). National Household Surveys (2005 & 2012) | Great Lakes of Africa Centre | University of Antwerp. [online] www.uantwerpen.be. Available at: https://www.uantwerpen.be/en/projects/great-lakes-africa-centre/national-datasets-livelihoods-drc/national-household-s/ [Accessed 19 Feb. 2023].

Nevo, A. (2011). Empirical Models of Consumer Behavior. Annual Review of Economics, [online] 3(1), pp.51–75. doi:https://doi.org/10.1146/annurev-economics-061109-080402.

Poi, B.P. (2012). Easy Demand-System Estimation with Quaids. The Stata Journal: Promoting communications on statistics and Stata, 12(3), pp.433–446. doi:https://doi.org/10.1177/1536867x1201200306.

Prais, S.J. and Houthakker, H.S. (1971). The Analysis of Family Budgets. CUP Archive.

Rauch, J.E. (1991). Modelling the informal sector formally. Journal of Development Economics, 35(1), pp.33–47. doi:https://doi.org/10.1016/0304-3878(91)90065-4.

Tucho, G.T. (2022). A review on the socio-economic impacts of informal transportation and its complementarity to address equity and achieve sustainable development goals. Journal of Engineering and Applied Science, 69(1). doi:https://doi.org/10.1186/s44147-022-00074-8.

Ulyssea, G. (2010). Regulation of entry, labor market institutions and the informal sector. Journal of Development Economics, 91(1), pp.87–99. doi:https://doi.org/10.1016/j.jdeveco.2009.07.001.

DOI: https://doi.org/10.14324/111.444.2755-0877.1590

Appendix

Appendix 1

Appendix 1.1 below shows the product code for goods in module 11 data that are adjusted to formal goods because of their properties, as discussed in section V. The classification of goods follows the COICOP12 nomenclature, the 12 categories of goods are divided into 45 groups containing 97 groups. The subgroups contain 143 items. The full codes for other products are available in the data file provided.

Appendix 1.1: Codes of Products Adjusted to Formal Goods

Code	Description	Code	Description	Code	Description
11112	Imported rice	**11521**	Imported vegetable oil	**21312**	Imported lager
21314	Imported brown Briere	**11143**	Industrial square bread	**22022**	Imported cigarettes n.e.s.
11145	Industrial round bread	**11155**	Imported biscuits	**31115**	Imported terry cloth
31215	Imported trousers	**31232**	Imported blouse	**11822**	Imported honey
11268	Canned meat	**11918**	Homogenised infant foods and dietetic formulae of any composition	**12221**	Mineral water, soda
12222	Sweetened drinks (Coke, Fanta, Lemonades)	**12227**	Industrial nonalcoholic beverages n.e.s.	**44111**	R´egideso Water Tariff
44112	Well Water Tariff	**45111**	Snel Electricity Tariff	**45501**	Hot water or steam from a central heating plant
61111	Aspirins, Novalgine	**61112**	Vermox, Mebox in syrup	**61113**	Vitamin C, Vitamin B12 and Other Vitamins
61114	Mercurochrome, iodised alcohol	**61115**	Quinine, Malaxin and other anti-malarials	**61116**	Penicillin, Ampicillin, Aminophylline and other antibiotics
61117	Pulmin syrup, Hemoglobin syrup	**61118**	Kongo bololo, Lemongrass leaf and other traditional	**61119**	Other modern medicines n.e.s.
61215	English condom and other condoms and mechanical contraceptives	**61330**	Various prostheses	**62101**	Generalist consultation in a public environment
62103	Specialist consultation in public areas	**62202**	Extraction of a tooth at the public doctor's office	**62313**	Saddle examination (public hospital)
62314	Saddle examination (private hospital)	**62317**	Chest X-ray, Ultrasound (public hospital)	**62322**	Drawing up a form (public hospital)
63001	Daily charges (general ward), public hospital	**63003**	Daily fees (single room), state hospital	**73101**	Transport of persons by train
73102	Transport of luggage by train	**73204**	Urban train	**73311**	Domestic overhead lines KIN-LSHI
73312	Domestic Airlines KIN-TSHIKAPA	**73314**	International lines	**73315**	Other domestic air transport n.e.s.
73313	Domestic airlines KIN-KISANGANI	**81010**	Postage for letters, postcards and parcels i	**81020**	Postage for letters, postcards and parcels na

DOI: https://doi.org/10.14324/111.444.2755-0877.1590

Code	Description	Code	Description	Code	Description
82001	Purchase of telephones, radiotelephones, fax machines	**82002**	Purchase of a mobile phone	**83004**	Purchase of prepaid communication cards, installation costs
92101	Planes, gliders	**92103**	Outboard motors	**94101**	Football stadium
94102	Swimming pools, golf courses, gyms and other fitness centres	**94104**	Fairgrounds and amusement parks	**94105**	Rides, swings and other playground equipment
94106	Hire of sports and leisure equipment and accessories	**94211**	Cinema	**94212**	Concert, music halls, opera, circus, sound and light shows
94213	Theatre	**94214**	Library, museum, art gallery, exhibition rooms	**94221**	Zoological gardens, national parks, botanical gardens monuments
94224	Subscription to television and radio channels				

The table below shows the codes of products that are removed to alleviate the measurement error in calculating the unit values as discussed in section VI. Codes like '621**' means all codes begin with 621.

Appendix 1.2: Codes of Product Removed

Code	Description	Code	Description	Code	Description
32213	Shoe Shine, Cleaning Services	32214	Shoe rental	32215	Other repair of footwear n.e.c.
31411	Mending and alteration costs for a jacket set	31412	Cost of darning, mending and reworking a SAFARY set	31413	Cost of mending trousers, cost of altering trousers
31414	Darning, mending and alteration costs of a wax shirt	31415	Dry cleaning, bleaching and dyeing of men's clothing	31416	Clothing rental
31417	Other repairs to men's clothing n.e.c.	31421	LIBAYA bodice mending and alteration costs	31422	Mending and alteration of women's dresses and ensembles
31423	Women's clothing rental	31424	Dry cleaning, bleaching and dyeing of women's clothing	31425	Other and repairs to women's clothing n.e.c.
41101	Rent actually paid by the tenant or sub-tenant	41201	Effective rents of secondary residences	42111	Shadow rents for owner-occupiers of their main residence
42211	Shadow rents for households housed free of charge or payingf	43211	Masonry Department	43212	Service Painter, Decorator
43213	Plumbing Service, Glazier Service	43214	Carpenter, Joiner Service	43215	Electrician service
43216	Other housing maintenance services n.e.c.	43300	deposit	44111	R´egideso Water Tariff
44112	Well Water Tariff	44113	Other water supply services (water subletting)	44201	Collection and treatment of household waste

DOI: https://doi.org/10.14324/111.444.2755-0877.1590

Code	Description	Code	Description	Code	Description
45111	Snel Electricity Tariff	45112	Other electricity distribution services ?	56211	Salary of domestic staff (Cook, Gardener)
56212	Worker, Good	56213	Caretaker	56214	Salary of a driver, secretary, tutor, etc...
56215	Other domestic services n.e.c.	56221	Household services (window cleaning, disinfection, fumigation, ...)	56222	Dry cleaning, bleaching and dyeing of household linen
56224	Rental of furniture and household goods, carpets, household appliances,...)	56225	Other services for habitation n.e.c.	56300	Hire of tables, chairs and hangings for festivals
55107	Repair of these items	55108	repair of these items	53311	Repair of household appliances
54128	Repair of these items	51301	Repair of furniture, fixtures and fittings	52008	Repair of household textiles
621**	Generalist consultation	622**	Dental services	623**	Urine examination
624**	other dental services	630**	Daily fees of hospitals	61380	Repair of medical devices
723**	Installation and repair of transportation facilities	736**	Carrier service	830**	telephone services
91109	Repair of audio and video equipment	925**	Repair of audio, visual equipment	923**	Repair of boats
93106	Repair of toys and games	93501	Veterinary Services	93502	Other pet services, grooming, daycare

Appendix 2

The available methods to estimate demand can be classified into two main categories (Nevo, 2011). The first one the estimating an aggregate demand system, which does not allow us to estimate the heterogeneity in consumer behaviour. The second one is the discrete choice model, defining products by groups of characteristics, which allows the identification of the heterogeneity preference of individuals.

One example of the first approach is the Almost Ideal Demand System (AIDS). The original and linear AIDS model is (Deaton and Muellbauer, 1980),

$$w_i = \alpha_i + \sum_{j=1}^{k} \gamma_{ij} * \ln(p_j) + \beta_i * ln\left(\frac{m}{a(p)}\right), \qquad i = 1, \dots, k$$

where w_i and p_i are the budget share and of good *i* respectively and *lna*(*p*) is the price index, its formulation formulated in Deaton and Muellbauer (1980). AIDS is relatively easy to estimate and allows us to estimate the budget share elasticity and price elasticity. There are available STATA commands, allowing flexible functional forms. Poi (2012) developed the command *quaids*, allowing for quadratic terms of expenditure and demographic variables. To deal with the collinearity between prices and endogeneity bias, instruments are recommended for expenditure and price (Nevo, 2011). Lecocq and Robin, (2015) developed the command aidsills, allowing for instruments for expenditure and price, quadratic terms, and demographic terms.

DOI: https://doi.org/10.14324/111.444.2755-0877.1590

This standard AIDS may not be plausible in our settings. First, it requires perfect data on price and consumers' full expenditure lists. More importantly, in this paper, due to data limitations, goods are grouped which makes the quality effect of particular importance, and we are interested in the relationship between income and quality, but the standard AIDS does not specify the quality effects.

The second approach is a discrete choice model, defining products by groups of characteristics. This approach starts by specifying an indirect utility of a consumer and the consumer only buys one unit of the good which give the highest utility. The probability that the consumer chooses the good is the probability that the consumer can get higher utility by purchasing it than any other goods. Advanced versions of the model like the nested Logit model and the random coefficient model can study the heterogeneity in preference between individuals. However, one important assumption here is that the consumer only buys one unit of each product, which is unrealistic in some household-level survey datasets (Nevo, 2011). Besides, in our setting, there is still the problem of lacking price data and specification of quality effect. In all, although the method used in this paper is subject to some limitations, given the available approaches and commands, it fits our research interests and data structure the best.

Appendix 3

This Appendix gives the estimates for groups N, J, K and M in table 5.

Appendix 3.1: N: Results of Full Sample: By Original Channel Indicator

	Budget Share Regressions			
	No Imputation		With Imputation	
	Formal	Informal	Formal	Informal
lnx	0.00112***	-0.00329***	0.00101***	-0.00312***
	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.4860	0.4976	0.4533	0.4672
N	2548	14333	3128	15355
	Unit Value Regressions			
	No Imputation		With Imputation	
β^1	0.769***	0.395***	0.677***	0.377***
	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.5323	0.7046	0.5048	0.6799
N	2548	14333	3128	15355
ϵ_x	0.533	0.484	0.600	0.509

p-values in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

	Price Elasticity			
	No Symmetry Restriction			
	No imputation		With Imputation	
	Formal	Informal	Formal	Informal
Formal	**-0.536**	0.0979	**-0.605**	0.0384
	(0.13)	(0.13)	(0.08)	(0.12)
Informal	0.0330	**-0.725**	-0.0192	**-0.835**
	(0.027)	(0.09)	(0.03)	(0.06)
	With Symmetry Restriction			
	Formal	Informal	Formal	Informal
Formal	**-0.543**	0.213	**-0.596**	-0.128
	(0.13)	(0.18)	(0.08)	(0.19)
Informal	0.0310	**-0.724**	-0.0159	**-0.837**
	(0.25)	(0.09)	(0.03)	(0.06)

Note: This estimation uses the original channel indicator shown in module 11. Bootstrapped standard error in parentheses. Price elasticity is significant if in bold (The price elasticity is larger than twice the bootstrapped standard errors in magnitude).

DOI: https://doi.org/10.14324/111.444.2755-0877.1590

Appendix 3.2: J: Results by Categories: Housing/Furnishing

	Budget Share Regressions			
	No Imputation		With Imputation	
	Formal	Informal	Formal	Informal
lnx	-0.000297	-0.00435***	-0.000455	-0.000430***
	(0.358)	(0.000)	(0.174)	(0.000)
R^2	0.8506	0.3368	0.8471	0.3234
N	300	11999	347	12346
	Unit Value Regressions			
	No Imputation		With Imputation	
β^1	0.337**	0.437***	0.296**	0.427***
	(0.013)	(0.000)	(0.039)	(0.000)
R^2	0.7991	0.4920	0.7996	0.4682
N	300	11999	347	12346
ϵ_x	0.515	0.389	0.481	0.401

p-values in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

	Price Elasticity			
	No Symmetry Restriction			
	No imputation		With Imputation	
	Formal	Informal	Formal	Informal
Formal	0.515	-0.163	0.106	-0.073
	(0.50)	(0.37)	(0.38)	(0.29)
Informal	0.0131	**-0.651**	0.0478	**-0.665**
	(0.057)	(0.085)	(0.05)	(0.088)
	With Symmetry Restriction			
	Formal	Informal	Formal	Informal
Formal	0.508	-0.00816	0.101	0.0447
	(0.47)	(0.072)	(0.38)	(0.062)
Informal	-0.00647	**-0.645**	0.0363	**-0.661**
	(0.058)	(0.089)	(0.05)	(0.094)

In the regressions for decomposed categories, the channel indicator used is the adjusted channel indicator. Bootstrapped standard error in parentheses. Price elasticity is significant if in bold (The price elasticity is larger than twice the bootstrapped standard errors in magnitude).

DOI: https://doi.org/10.14324/111.444.2755-0877.1590

Appendix 3.3: K: Results by Categories: Health

	Budget Share Regressions			
	No Imputation		With Imputation	
	Formal	Informal	Formal	Informal
lnx	-0.000248***	-0.000306	-0.000236***	-0.000306
	(0.000)	(0.195)	(0.000)	(0.195)
R^2	0.3491	0.8618	0.3357	0.8970
N	2902	127	3641	163
	Unit Value Regressions			
	No Imputation		With Imputation	
β^1	0.509***	0.325	0.458***	0.325
	(0.000)	(0.337)	(0.000)	(0.337)
R^2	0.4706	0.9189	0.4317	0.9306
N	2902	127	3641	163
ϵ_x	0.248	-0.0425	0.313	-0.0542

p-values in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

	Price Elasticity			
	No Symmetry Restriction			
	No imputation		With Imputation	
	Formal	Informal	Formal	Informal
Formal	-0.322	0.0855	**-0.730**	0.0292
	(0.29)	(0.49)	(0.28)	(0.31)
Informal	0.0506	-0.456	-0.00952	-0.420
	(0.37)	(0.83)	(0.29)	(0.59)
	With Symmetry Restriction			
	Formal	Informal	Formal	Informal
Formal	-0.317	0.0587	**-0.731**	0.0161
	(0.25)	(0.16)	(0.27)	(0.17)
Informal	0.141	-0.457	0.0399	-0.418
	(0.39)	(0.52)	(0.41)	(0.58)

In the regressions for decomposed categories, the channel indicator used is the adjusted channel indicator. Bootstrapped standard error in parentheses. Price elasticity is significant if in bold (The price elasticity is larger than twice the bootstrapped standard errors in magnitude).

DOI: https://doi.org/10.14324/111.444.2755-0877.1590

Appendix 3.4: M: Results by Categories: Communication and Leisure

	Budget Share Regressions			
	No Imputation		With Imputation	
	Formal	Informal	Formal	Informal
lnx	-0.00168**	-2.01e-06	-0.000806	-4.70e-07
	(0.015)	(0.986)	(0.392)	(0.997)
R^2	0.8893	0.4172	0.7753	0.3962
N	265	1921	314	2173
	Unit Value Regressions			
	No Imputation		With Imputation	
β^1	1.008***	0.440***	1.096***	0.401***
	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.8861	0.4395	0.8321	0.4052
N	265	1921	314	2173
ϵ_x	-0.456	0.558	-0.297	0.599

p-values in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

	Price Elasticity			
	No Symmetry Restriction			
	No imputation		With Imputation	
	Formal	Informal	Formal	Informal
Formal	**-0.264**	0.499	-0.504	0.579
	(0.12)	(0.55)	(0.37)	(2.67)
Informal	-0.0259	0.782	-0.0163	0.775
	(0.13)	(0.87)	(0.55)	(4.18)
	With Symmetry Restriction			
	Formal	Informal	Formal	Informal
Formal	**-0.262**	0.0387	**-0.465**	0.0332
	(0.12)	(0.058)	(0.12)	(0.12)
Informal	0.133	0.775	0.124	0.629
	(0.20)	(1.30)	(0.46)	(2.80)

In the regressions for decomposed categories, the channel indicator used is the adjusted channel indicator. Bootstrapped standard error in parentheses. Price elasticity is significant if in bold (The price elasticity is larger than twice the bootstrapped standard errors in magnitude).

DOI: https://doi.org/10.14324/111.444.2755-0877.1590

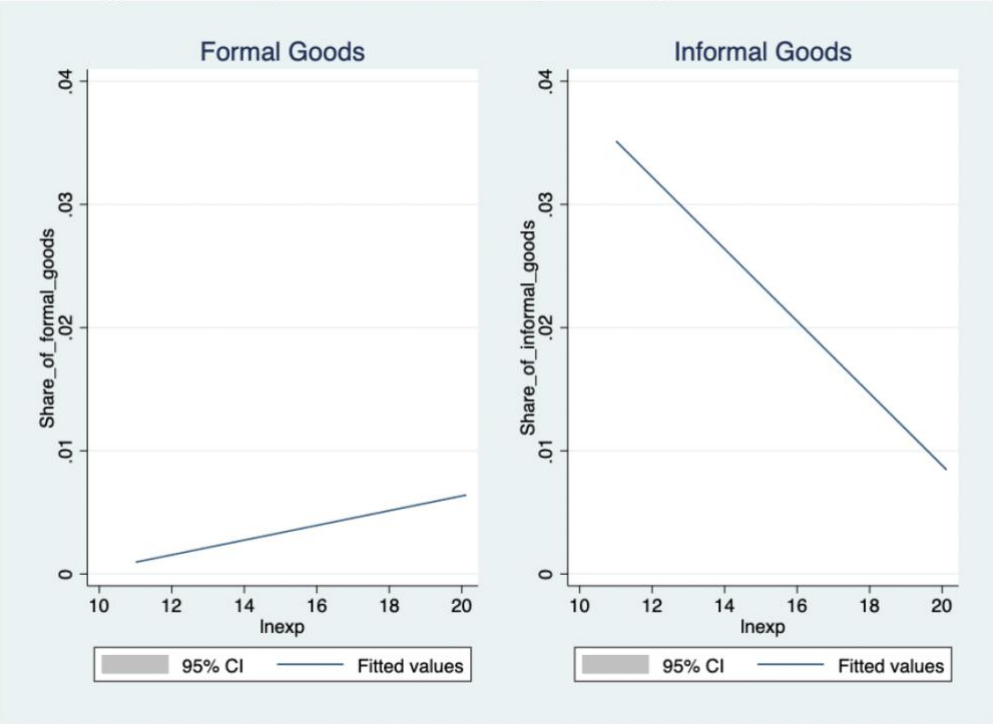
Appendix 4

This Appendix shows the estimated linear Engel curves for all the first stages of estimations in section VII, ignoring the quality effects.

Appendix 4.1 shows the Engel Curve for the Full Sample using the adjusted channel indicator (group A in table 5). Appendix 4.2, Appendix 4.3 and Appendix 4.4 shows the estimated Engel curves for groups B, G, H, I, J, K, L and M. The title of the graphs corresponds to groups in table 5 as follows: 'wc11' and 'wc12' are formal and informal goods for group B respectively. 'wc21' and 'wc22' are formal and informal goods for group G. 'wexp21' and 'wexp22' are formal and informal goods for H respectively. 'wexp31' and 'wexp32' are formal and informal goods for I respectively. 'wexp451' and 'wexp452' are formal and informal goods for J respectively. 'wexp61' and 'wexp62' are formal and informal goods for K respectively. 'wexp71' and 'wexp72' are formal and informal goods for group L respectively. 'wexp891' and 'wexp892' are formal and informal goods for group M respectively.

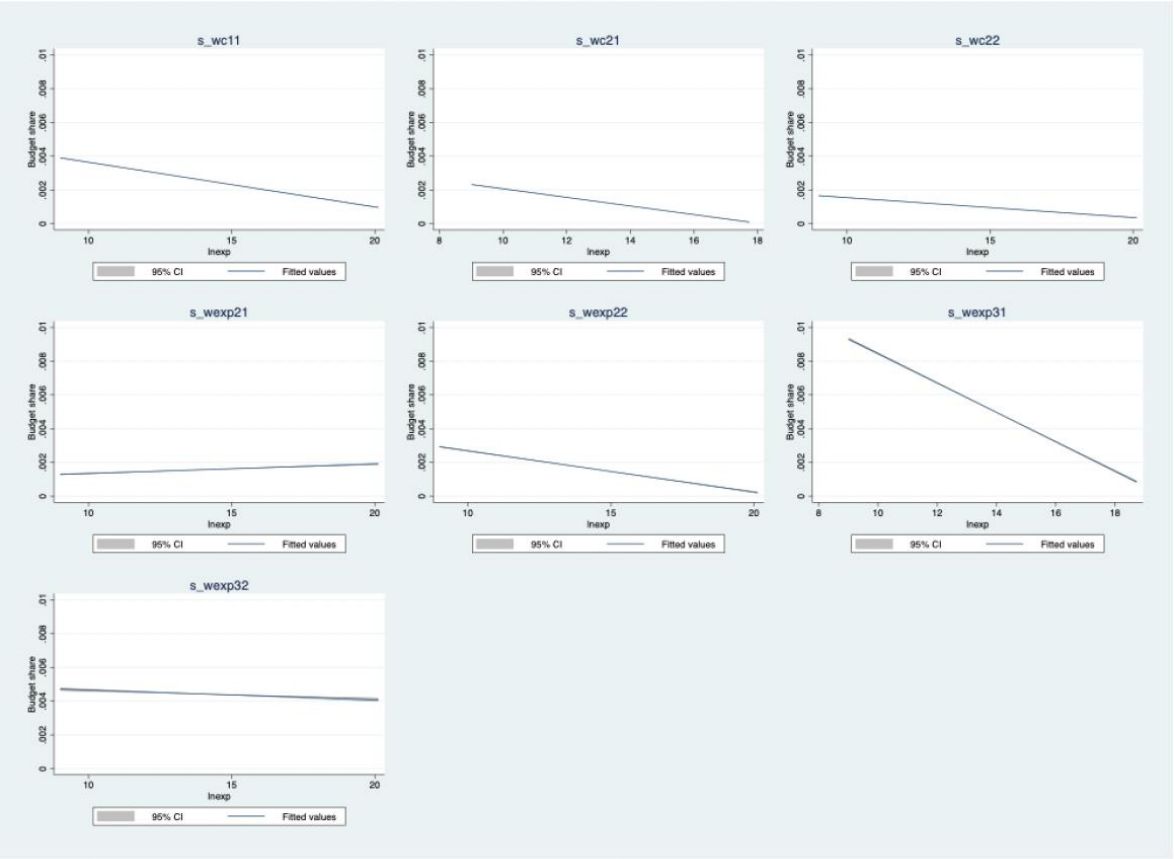
Appendix 4.5 shows the estimated Engel curves for groups C, D, E, and F. The title of the graphs corresponds to groups in Table 5 as follows: 'wr1111' and 'wr1112' are formal and informal goods for C. 'wc231' and 'wc232' are formal and informal goods for D respectively. 'wc481' and 'w482' are formal and informal goods for E respectively. 'wc671' and 'w672' are formal and informal goods for F respectively.

Appendix 4.1: Engel Curve for the Full Sample with Adjusted Channel Indicator



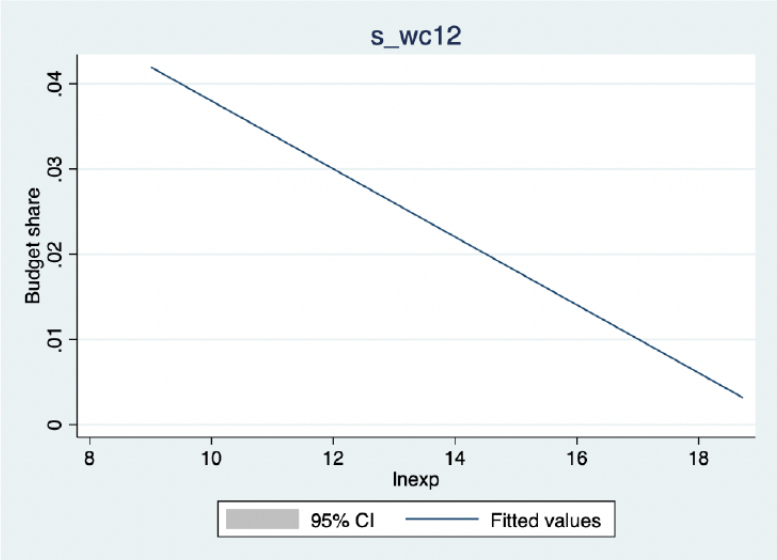
DOI: https://doi.org/10.14324/111.444.2755-0877.1590

Appendix 4.2: Engel Curve for Categories



Note: The title of graphs corresponds to groups in table 5 as follows: 'wc11' and 'wc12' are formal and informal goods for group B respectively. 'wc21' and 'wc22' are formal and informal goods for group G. 'wexp21' and 'wexp22' are formal and informal goods for H respectively. 'wexp31' and 'wexp32' are formal and informal goods for I respectively. 'wexp451' and 'wexp452' are formal and informal goods for J respectively.

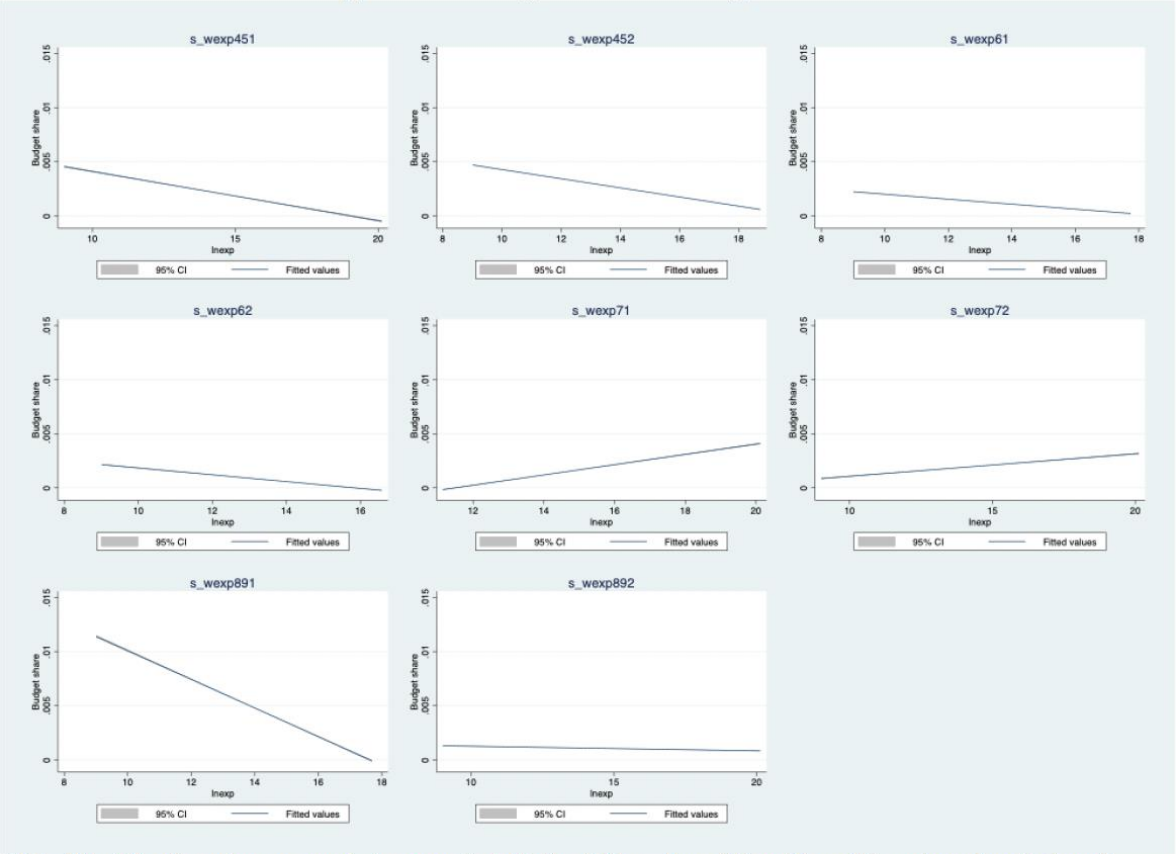
Appendix 4.3: Engel Curve for Categories



Note: The title of graphs corresponds to groups in table 5 as follows: 'wc11' and 'wc12' are formal and informal goods for group B respectively. 'wc21' and 'wc22' are formal and informal goods for group G. 'wexp21' and 'wexp22' are formal and informal goods for H respectively. 'wexp31' and 'wexp32' are formal and informal goods for I respectively. 'wexp451' and 'wexp452' are formal and informal goods for J respectively.

DOI: https://doi.org/10.14324/111.444.2755-0877.1590

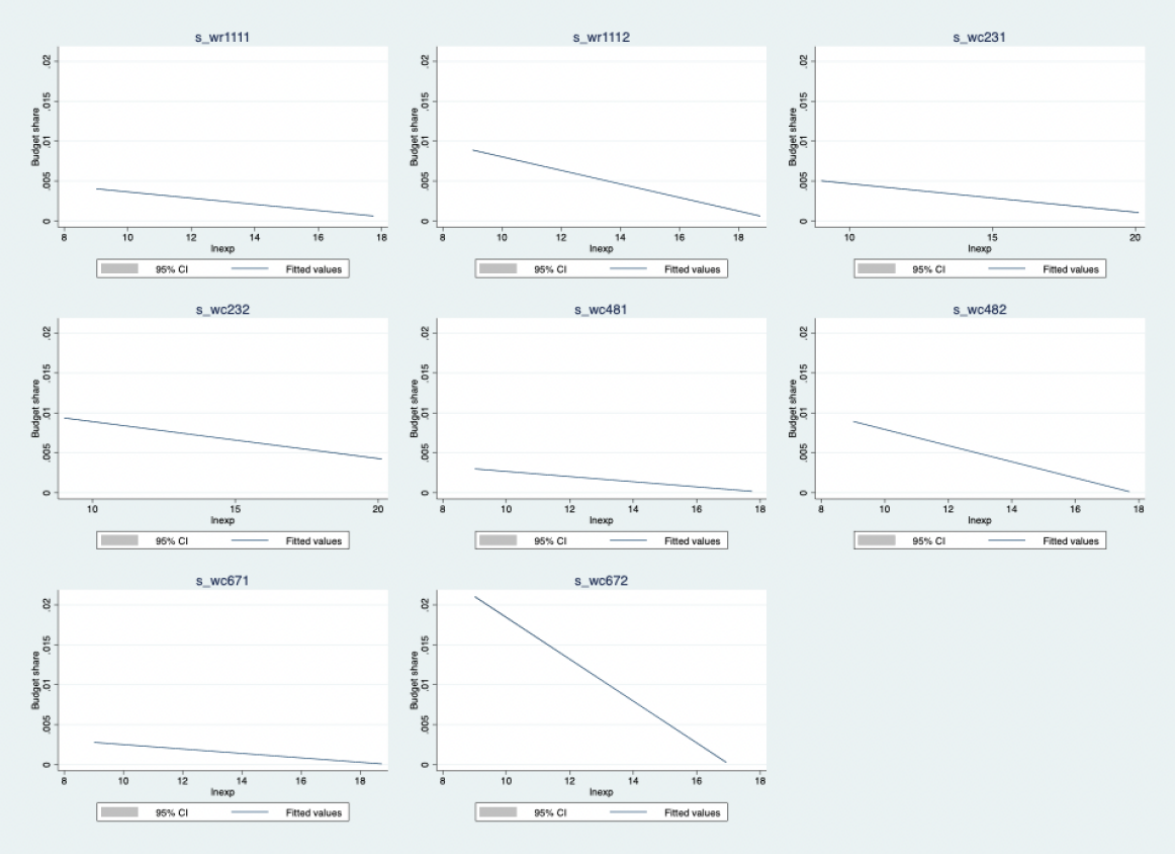
Appendix 4.4: Engel Curve for Categories



Note: The title of graphs corresponds to groups in table 5 as follows: 'wexp61' and 'wexp62' are formal and informal goods for K respectively. 'wexp71' and 'wexp72' are formal and informal goods for group L respectively. 'wexp891' and 'wexp892' are formal and informal goods for group M respectively.

DOI: https://doi.org/10.14324/111.444.2755-0877.1590

Appendix 4.5: Engel Curve for Categories



Note: The title of graphs corresponds to groups in table 5 as follows: 'wr1111' and 'wr1112' are formal and informal goods for C. 'wc231' and 'wc232' are formal and informal goods for D respectively. 'wc481' and 'w482' are formal and informal goods for E respectively. 'wc671' and 'w672' are formal and informal goods for F respectively.

Appendix 5

As discussed in the review, unit values are usually used as proxies for quality. If the informal goods are of lower quality than the formal goods, we should observe that the unit values of informal goods are significantly lower than that of the formal goods.

Appendix 5.1 uses the full sample of all categories; columns A and E are average unit values for classification using adjusted channel indicator and original channel indicator respectively. Both columns show that the average unit value (in natural logarithm) of formal goods is significantly higher than that of the informal goods channels and categories respectively. Appendix 5.2 and Appendix 5.3 shows that the average unit value for the formal is significantly higher for most categories for foods and non-foods. Particularly, the column H of Appendix 5.2 shows that the average difference is negative but insignificant for alcoholic beverages. Overall, we observe a consistent pattern with previous studies, that informal goods may be of lower quality than their formal counterparts.

DOI: https://doi.org/10.14324/111.444.2755-0877.1590

Appendix 5.1: Difference in Unit Values: Full Sample

A	E
Adjusted Channel Indicator	Original Channel Indicator
0.140***	0.823***
(0.000)	(0.000)

p-values in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: The names of this table corresponds to the name of groups in the first column in table 5. For example, column 'A' represent the full sample estimation using adjusted channel indicator.

Appendix 5.2: Difference in Unit Values: By Categories

B	G	H	I	J	K	L	M
0.0955***	0.744***	-0.148	0.694***	0.735***	0.0528	0.195**	1.311***
(0.000)	(0.000)	(0.074)	(0.000)	(0.000)	(0.725)	(0.001)	(0.000)

p-values in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: The names of this table corresponds to the name of groups in the first column in table 5: 'B' for foods, 'G' for non-alcoholic beverage, 'H' for alcoholic beverage, 'I' for clothing, 'J' for housing and furnishing, 'K' for health, 'L' for transportation, 'M' for communication and leisure expenditure. The classification of informal and formal goods is using the adjusted channel indicator. The data is not imputed.

Appendix 5.3: Difference in Unit Values: Food

C	D	E	F
d1	d2	d3	d4
0.208***	0.431***	0.426***	0.263***
(0.000)	(0.000)	(0.000)	(0.000)

p-values in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: The names of this table corresponds to the name of groups in the first column in table 5: 'C' for cereals or staples, 'D' for meat or seafood, 'E' for diary products, fat, sugar and sauce, 'F' for vegetable and fruit. The classification of informal and formal goods is using the adjusted channel indicator. The dats is not impured.

DOI: https://doi.org/10.14324/111.444.2755-0877.1590