Informality, Consumption Taxes and Redistribution

Pierre Bachas, Lucie Gadenne & Anders Jensen*

Abstract

Can taxes on consumption alleviate inequality in developing countries? We assemble expenditure data from 31 countries to study the redistributive potential of consumption taxes. We revisit the role played by rate differentiation across products and consider a new redistribution channel: the de facto tax exemption of the informal sector. Using the place of purchase to proxy for informal sector consumption enables us to characterize Informality Engel Curves: we find that the budget share spent in the informal sector steeply declines with income in all countries. We then extend the standard optimal commodity tax model to allow for an informal sector, which introduces a new equity-efficiency trade-off. When calibrated to our data, we find that the informal sector makes consumption taxes strongly progressive: on average, households in the richest quintile face an effective tax rate that is 2.5 times higher than the poorest quintile's. These progressivity gains fall with development as the informal sector shrinks. On the contrary, the equity gains from subsidizing food relative to non-food products is limited everywhere, and particularly so for poorer countries where low-income households consume most of their food from the informal sector.

JEL Classification: H21, H23, H32, O23

*Pierre Bachas: World Bank Research, pbachas@worldbank.org. Lucie Gadenne: University of Warwick, Institute for Fiscal Studies and CEPR, l.gadenne@warwick.ac.uk. Anders Jensen: Harvard Kennedy School and NBER, Anders_Jensen@hks.harvard.edu. The findings and conclusions are those of the authors and do not represent the views of the World Bank. We thank Eva Davoine, Elie Gerschel, Mariana Racimo, Roxanne Rahnama and Alvaro Zuniga for excellent research assistance. Financial support from the World Bank Development Economics' Research Support Budget and the IFS's TaxDev center is gratefully acknowledged.

1 Introduction

Inequality in developing countries is alarmingly high and is showing no sign of a reduction in the 21st century (Alvaredo et al., 2018). In high income countries, governments seek to curb inequality through direct income taxes and transfers. In lower income countries, however, direct income taxes are constrained (Jensen, 2019). Governments raise revenue primarily through indirect consumption taxes, which are considered regressive (Warren, 2008), and use blunt instruments to introduce progressivity, such as reduced tax rates on necessities – typically food products. The potential negative distributional impact of consumption taxes features prominently in debates on tax policy's role to limit inequality in developing countries (Lustig, 2018).

In this paper we assemble a dataset of expenditure surveys in multiple countries and combine it with theory to study the redistributive capacity of consumption taxes in low and middle income countries. We revisit the role played by rate differentiation across products and consider a new redistribution channel: the de facto exemption of the informal sector from taxation. First, we develop a new method to measure informal (untaxed) consumption from expenditure diaries available in 31 countries, which record the place of purchase for each transaction (e.g. street stalls vs supermarkets). This allows us to construct Informality Engel Curves (IECs): we find that in all countries, the budget share that households spend in the informal sector steeply declines with their income. Second, we consider how optimal tax policy changes as countries develop, by extending the optimal commodity tax model of Diamond (1975) to allow for an untaxed informal sector, and calibrate the model using our data. We show that taking into account the informal sector makes consumption taxes more progressive overall but increases their efficiency cost. Looking at changes in optimal tax policy over the development path, we find that these progressivity gains fall as the informal sector shrinks with development. The redistribution grains from subsidizing food relative to non-food products are however smaller in poor countries, where poorer households consume most of their food from the informal sector.

A major constraint in studying informality is that, by definition, informal sector purchases are hard to observe and to link to consumers' incomes. Our innovation is to use the *places of purchase* reported by households in expenditure surveys to proxy for the share of consumption purchased from the formal sector. Our core dataset includes household expenditure surveys from 31 countries that span a wide range of development levels – from Burundi to Chile – and in which the source of consumption (such as 'home production', 'street stall' or 'supermarket') is reported for each product. This

enables us to construct a detailed taxonomy of places of purchases harmonized across countries. To assign a probability that taxes are paid on purchases from each reported source we build on the literature showing that large, modern retailers are much more likely to remit taxes than smaller, traditional retailers due to information trails on their activities (Kleven et al., 2016; Lagakos, 2016). We provide micro evidence from retail censuses to support our assignment.

Our first contribution is to document how informal consumption varies with income both within and across countries. Unsurpisingly, we find that the aggregate share of informal consumption decreases with development from 80% in low income countries to 20% in upper-middle income countries. We also find a sizeable negative correlation between informal consumption and log total expenditure *within* countries: Informality Engel Curves are always decreasing and are approximately linear with respect to log household expenditure. On average, a 10% increase in household expenditure corresponds to a 1 percentage point decrease in the informal consumption budget share. We then investigate the potential mechanisms underlying this negative slope. We find that geography and preferences across products both play a role: controlling for household location and detailed (4-digit level) product choices explains 60% of the negative correlation between household income and informal budget shares. IEC slopes remain however statistically significant in most countries once these factors are controlled for. We show suggestive evidence that the remaining variation may be explained by richer households valuing quality more and formal firms producing higher-quality products, in line with evidence in Faber and Fally (2017) for the US and Atkin et al. (2018b) for Mexico.

We provide a similar characterization of the within country food Engel curves from a uniquely large sample of expenditure surveys from 89 low and middle income countries. Consistent with an extensive literature, we confirm that the food Engel curves are downward slopping and close to linear with respect to log household expenditure. We estimate that the food budget share decreases by 1.3 percentage points on average when household expenditure increases by 10%. This exercise also allows us to compare our core sample of 31 countries to this larger sample which is representative of 50% of the world population: food consumption patterns in our core countries mirror those of other countries at similar development levels. This suggests that our results regarding Informality Engel Curves could generalize widely.

Our second contribution is to analyze the implications of the shapes of Informality and Food Engel curves for optimal tax policy and redistribution. We extend the multi-person model of commodity taxation of Diamond (1975) in two directions: we introduce formal and informal (untaxed) varieties of each product and we allow for changes in

consumption patterns over the development path. We then calibrate our model using data from our core sample of countries. Taking into account informal consumption impacts both the efficiency and equity of consumption taxes: it increases the efficiency cost since households can substitute to informal varieties when taxes increase, and it makes consumption taxes progressive as long as IECs are downward-sloping.

Studying how optimal tax policy changes over the development path yields two main insights. First, we find that consumption taxes become less progressive overall as countries become richer. This result is due to the (approximate) log-linearity of the IECs, which implies that when incomes grow the poor consume a larger share of total formal consumption, lowering the progressivity of tax systems in which the informal sector is 'de facto' exempt.[1] Second, we find that the equity gains from subsidizing food relative to non-food products – the 'de jure' rate differentiation policy in place in most countries – are lowest in poorer countries. This is true even if we assume that both the formal and informal varieties can be taxed: the log-linearity of food Engel curves alone implies that the poor consume a larger share of total food consumption as countries develop, making a relative subsidy on food products more attractive. Taking into account informal consumption makes rate differentiation even less progressive in poor countries – where poorer households consume most of their food in the informal sector – but more efficient, since higher informal consumption shares for food increases the efficiency cost of taxing food relative to non-food products. For the lowest income countries our results turn the conventional rationale for rate differentiation on its head: we find that taxing food less than non-food products could be justified on efficiency but not on equity grounds.

We finally quantify the progressivity of different tax scenarios in each country in the core sample. We find that setting a uniform rate on consumption is progressive once the informal sector is taken into account: the effective tax rate paid by the top quintile is 2.5 times higher than those paid by the bottom quintile on average. Progressivity is highest amongst low and lower-middle income countries, where on average the top quintile pays 3 times higher effective rates than the bottom quintile. Optimally differentiating rates on food relative to non-food products only leads to very modest progressivity gains, particularly in poorest countries. Consumption taxes are thus an effective redistribution instrument in developing countries because of the 'de facto' progressivity of informal consumption patterns. The typical 'De jure' progressivity policy through exemption of food products on the other hand only has a limited redistributive potential.

[1] This is true even with constant IEC slopes: with downward-sloping log-linear IECs an increase in the aggregate formal share of consumption is sufficient to decrease optimal consumption taxes. In addition a larger IEC slope increases optimal taxes, but differences in slopes explain only a small share of the variation in optimal tax policy across countries.

Our baseline results rely on several assumptions which can be relaxed. First, we show that our empirical results on the shape of the IECs are robust to changes in our formality assignment assumptions. This is because the correlation between budget shares and household expenditure is particularly strong for sources of consumption which are unambiguously informal (non-market consumption) or formal (large supermarkets). Second, we use total expenditures to proxy for household income, because of data limitations.[2] Allowing for savings rates to increase with income decreases the static redistributive effect of consumption taxes, as expected, but only slightly so. Third, we assume that producer prices are exogenous, in line with the existing literature on optimal consumption taxes. This assumption implies that taxes are fully (not) passed-through to consumer prices in the formal (informal) sector, it may not hold if informal retailers buy inputs from formal suppliers. We observe the share of inputs purchased by informal retailers from formal firms in the census data of one country (Mexico). Taking this into account lowers the progressivity effect of the informal sector by roughly 15%.

The rest of the paper is structured as follows. Section 2 describes our data sources and methodology. Section 3 provides new stylized facts on informal consumption patterns and documents the food Engel curves. Section 4 presents the model used to characterize optimal commodity tax policy in the presence of informal consumption, and Section 5 calibrates the model using our data. Section 6 concludes.

1.1 Related Literature

Our paper makes two main contributions to the literature on tax policy in developing countries. First, we introduce informal consumption as a novel channel through which consumption taxes are progressive. This contrasts with the existing literature which typically does not take into account informal consumption, and conclude that, at best, consumption taxes do not reduce inequality (for recent studies see Lustig, 2018 and Harris et al., 2018). Two exceptions are Jenkins et al. (2006) and Muñoz and Cho (2003) who use retailer information to classify expenditures as formal or informal in respectively the Dominican Republic and Ethiopia. Our focus on the equity implications of the informal sector moreover contributes to the the literature on optimal tax design under imperfect enforcement (Allingham and Sandmo, 1972; Cremer and Gahvari, 1993, Best et al., 2015; Boadway and Sato, 2009) which mainly focuses on the efficiency properties of taxation.

[2]Expenditure surveys in developing countries often do not attempt to directly measure income, see Deaton (1997a) for a discussion of measurement problems.This exercise abstracts from life-cycle considerations by assuming that savings are never consumed. If all savings are eventually consumed from the same places of purchases as current consumption, our baseline assumptions of no savings approximates the progressivity of consumption taxes over the life cycle.

We show that tax systems in developing countries can be progressive even in the absence of direct income taxes, and thus also speak to debates on the optimal mix between direct and indirect taxes in these countries (Burgess and Stern, 1993; Huang and Rios, 2016).

Second, we use new empirical evidence to shed light on the redistributive potential of differentiating consumption tax rates across products.[3] Sah (1983a) uses consumption data from the U.K. and concludes that this potential is limited. This analysis has only been extended to a few countries, primarily India (Ahmad and Stern, 1984; Ray, 1986; Srinivasan, 1989). Our data enables us to revisit this debate using micro-data across 115 countries spanning the development spectrum. We confirm that optimally differentiating consumption taxes cannot achieve much redistribution; we show that this is particularly true in low-income countries, and that taking informal consumption patterns into account re-enforces this conclusion.

We also contribute to the empirical literature on consumption patterns and development. A large body of work documents how expenditure shares on specific goods vary with income, including the well-established Engel curve for food (Deaton and Paxson, 1998, Anker et al., 2011, Pritchett and Spivack, 2013, Almås, 2012). In addition to studying the properties of the food Engel curves in a much larger sample of countries, we document new facts on the correlation between income and place of purchase. This relates to the literature on changes in retailer type with development. Our formal versus informal store classification overlaps with this literature's distinction between traditional and modern retailers. Lagakos, 2016 and Bronnenberg and Ellickson, 2015 document the lower prevalence of modern stores in low-income countries from firm surveys or sector-level data. We confirm their finding using household expenditure data, and show that expenditures in the traditional (informal) sector falls with income within countries. More generally, we participate in a growing literature which compiles multi-country micro-data to study macro changes over the development path (Bick et al., 2018; Jensen, 2019; Donovan et al., 2018).

Finally, our paper contributes to the literature on the informal sector. Existing papers focus on the 'supply side' of informality, by evaluating incentives to become formal either at the firm-level (DeSoto, 1989; De Paula and Scheinkman, 2010; La Porta and Shleifer, 2014), at the worker-level (Gerard and Gonzaga, 2016; Jensen, 2019), or both (Ulyssea, 2018). Our approach is complementary by considering the 'demand side' of informality through consumers' shopping behavior in formal and informal retailers.[4] These studies

[3] We only consider linear differentiated consumption tax, abstracting from the possibility that non-linear consumption taxes could be used to achieve more redistribution - see Gadenne (2019).

[4] This builds on a literature in public finance using data on consumption to infer evasion behavior - see Pissarides and Weber (1989); Feldman and Slemrod (2007); Morrow et al. (2019).

typically focus on the positive efficiency impact of policies which reduce the informal sector size. Our results imply that such policies could also have negative distributional consequences by shifting the burden of taxation towards poorer households.

2 Data and Method

We use two datasets to provide new evidence on consumption patterns in developing countries and to shed light on the redistributive potential of consumption taxes: a core sample of 31 countries and a larger sample of 89 countries. First, we detail the method used to proxy for informal consumption at the household level, which can only be applied to the core sample, given data requirements. Second, we present the large sample which is used to document product specific consumption, and in particular food consumption. This allows us to compare along this dimension the core sample to the large sample, which represents over a half of the world population.

2.1 Core Sample

We assemble our core dataset by combining household expenditure surveys from 31 countries. We use surveys which satisfy three criteria. First, they are nationally representative: this enables us to study the progressivity of different tax scenarios using representative samples. Second, they record consumption from open diaries rather than pre-filled diaries, which only document selected products: thus the data contains information on all household expenditure. Third, they ask households to report the place each product is sourced from – or *place of purchase* – and this variable is rarely missing. This last criteria ensures that we can apply our method to proxy for informal sector consumption, described below.[5]

We combine micro data from 31 countries which satisfy these three criteria, such that our core sample contains information on over 400,000 households. Table 1 lists alphabetically the countries in the core sample, with survey names, years, number of households and average number of expenditure items reported by households. Countries in the sample are principally located in Latin America and Sub-Saharan Africa. Unfortunately, Asian countries' household expenditure surveys do not ask for places of purchase.[6] Nonetheless, the core sample covers a wide range of development levels, from Burundi (GDP per capita of 245 USD) to Chile (15,000 USD). We show in section 3.2 that our core

[5]Evidently, another criteria is data access: in a few cases we identified surveys which appeared to satisfy all criteria, but could not obtain the micro-data with the place of purchase.

[6]Survey design appears strongly correlated across countries within regions, showing the influence of regional development partners and/or historical ties across statistical administrations.

Table 1: Core Sample Household Expenditure Surveys

Country	Code	Survey	Year	GDP per capita	Sample size	Nb items/HH
Benin	BJ	EMICOV	2015	827.8	19902	31.82
Bolivia	BO	ECH	2004	1657.5	9149	49.42
Brazil	BR	POF	2009	10595	56049	47.99
Burkina Faso	BF	EICVM	2009	562.8	8404	161.58
Burundi	BI	ECVM	2014	245.3	6681	90.21
Cameroon	CM	ECAM	2014	1400.4	10303	95.76
Chad	TD	ECOSIT	2003	572.1	6747	92.03
Chile	CL	EPF	2017	14748.5	15237	129.19
Colombia	CO	ENIG	2007	5998.8	42733	79.65
Comoros	KM	EDMC	2013	1373.4	3139	83.52
Congo DRC	CD	E123	2005	300.6	12098	106.93
Congo Rep	CG	ECOM	2005	2569.3	5002	84.76
Costa Rica	CR	ENIGH	2014	8993.5	5705	67.48
Dominican Rep	DO	ENIGH	2007	5121.1	8363	89.07
Ecuador	EC	ENIGHUR	2012	5122.2	39617	88.65
Eswatini	SZ	HIES	2010	4168.5	3167	43.9
Mexico	MX	ENIGH	2014	9839.1	19459	57.36
Montenegro	ME	HBS	2009	6516.4	1223	134.07
Morocco	MA	ENCDM	2001	2095.2	14243	87.5
Mozambique	MZ	IOF	2009	415.7	10809	48.75
Niger	NE	ENCBM	2007	330.3	4000	153.26
Papua New Guinea	PG	HIES	2010	1949.3	3811	111.17
Peru	PE	ENAHO	2017	6314.7	43545	78.51
Rwanda	RW	EICV	2014	689.7	14419	53.62
Sao Tome	ST	IOF	2010	1094.7	3145	105.89
Senegal	SN	EDMC	2008	1277.5	1443	517.81
Serbia	RS	HBS	2015	6155.4	6531	105.97
South Africa	ZA	IES	2011	7454.7	25325	44.2
Tanzania	TZ	HBS	2012	788.4	10186	317.77
Tunisia	TN	ENBCNV	2010	4142	11281	139.12
Uruguay	UY	ENIGH	2005	9078.6	7042	77.47

This tables lists alphabetically the countries in our core sample, the survey names and years. GDP per capita is in PPP USD in the year of the survey, obtained from the *World Development Indicators*. The sample size refers to the number of households in the survey, and the number of items reported is the number of expenditure items reported on average across all households in the survey.

sample is representative of other developing countries along the dimensions which can be observed in the large sample.

Appendix B provides further details on the data sources used. Table B1 shows the geographical coverage of the core sample, table B2 lists the surveys considered for inclusion but ultimately discarded, and table B3 provides further details on the structure of the surveys' for each country in our core sample.

2.2 Method: proxying for informal consumption using places of purchase

Our main methodological innovation is to use the place of purchase reported for each expenditure to assign it a probability that it was obtained from a formal - tax registered - source. We first discuss how we allocate each place of purchase to a broad consumption source category. We then explain how we assign to each category a probability of being in the formal sector and provide empirical support to our assignment method.

Most recorded expenditures can be classified by place of purchase into seven categories. The first five pertain to purchases of goods. Ranked by order of retailer size these are: (1) non-market consumption (e.g. home production), (2) non brick-and-mortar stores (e.g. street stalls, public markets), (3) corner and convenience stores, (4) specialized stores (e.g. clothing stores) and (5) large stores (e.g. supermarkets, department stores). Purchases of services can be allocated to two categories: (6) services provided by an institution (e.g. banks, hospitals) and (7) services provided by an individual (e.g. domestic services). All together these categories account for 85% of total household expenditure. The remaining 15% are items for which no place of purchase is specified in the surveys. Figure A.3 shows that the majority of these unspecified places of purchase expenditures pertain to utilities and telecommunication.[7]

We assign each of these categories to either the formal or informal sector. We classify a category as belonging to the formal sector if consumption taxes are paid on most purchases from that category. In many countries, firms below a certain size are exempt from taxation by law (de jure informality) and even when legally subject to taxation, many small firms do not remit taxes (de facto informality) (see Keen and Mintz, 2004); we do not distinguish between these two forms of informality.[8]

The key assumption behind our assignment method is that larger retailers are more likely to be formal. This is by definition true for de jure informality, and there is a large literature arguing that the correlation between firm size and de facto informality is strong. Theoretically, Kleven et al. (2016) develop a model of tax evasion in which informality must be sustained by collusion between firm managers and their employees. Collusion costs are increasing in firm size, leading to less evasion amongst larger firms. Hsieh and Klenow (2014) and Hsieh and Olken (2014) similarly argue that the burden of taxation in developing countries falls more heavily on larger firms. Empirically, Kumler et al. (2015) find that compliance is indeed increasing in firm size in Mexico. We also

[7] We exclude housing expenditure from our analysis due to limited data on owner-occupied imputed rents.

[8] The prevalence of de jure informality is itself endogenous to the level of de facto informality: the legal decision to exempt small firms from commodity taxes is usually motivated by the fact that enforcement costs for the authorities, and compliance costs for the firms are large relative to potential tax liabilities.

expect larger stores to have more customers, which makes evasion harder to sustain as shown empirically by Naritomi (2018). Our assignment also mirrors Lagakos (2016) and Bronnenberg and Ellickson (2015) which categorize retailers as either 'traditional' (small, labor intensive retailers - our categories 1 to 3) or modern (large capital intensive retailers - our categories 4 and 5) and argue that traditional retailers are mostly informal.[9]

We provide direct evidence that the correlation between formality and firm size is indeed positive, using tax registration by firm size, available in the retail firm census of four core sample countries (Cameroon, Mexico, Peru and Rwanda). Figure 1 Panel (a) shows the share of formal firms as a function of log employment in each country.[10] We see that in all countries retailers with 2 employees or less are overwhelmingly informal, but more than 80% of retailers with 20 or more employees are formal.

In Mexico the census classifies retailers in categories which are broadly similar to our broad place of purchase categories. This enables us to go a step further and directly measure retailer size and the probability that a specific retailer type pays consumption taxes. Figure 1 shows for our categories (1) to (5) the log median number of employees (Panel b) and the share of firms paying Value-Added-Taxes on their sales (Panel c). We observe that non-brick-and-mortar stores and convenience stores are small and rarely formal, whereas nearly all large stores are formal.

Our baseline formality assignment follows that in Lagakos (2016) and assigns categories (1) to (3) to the informal sector, and categories (4) and (5) to the formal sector. Goods in category (1) – non-market consumption – are by definition not purchased in markets and therefore untaxed. Categories (2) and (3) (non brick and mortar stores and corner stores) are likely very small and mostly informal, whereas category (5) (large stores) consists mostly of supermarkets which are unlikely to fly below the radar of the tax authorities. For services, we assume that institutions pay taxes while individual providers do not, and assign category (6) to the informal sector and category (7) to the formal sector. We follow the same logic in assigning expenditures in the unspecified category to the formal sector: the bulk of these expenditures are utilities typically provided by large institutions which cannot evade taxes. Appendix B.2 provides more details on the methodology and table B4 shows for each country the original names of the places of purchase, our formality assignment, and their expenditure shares.

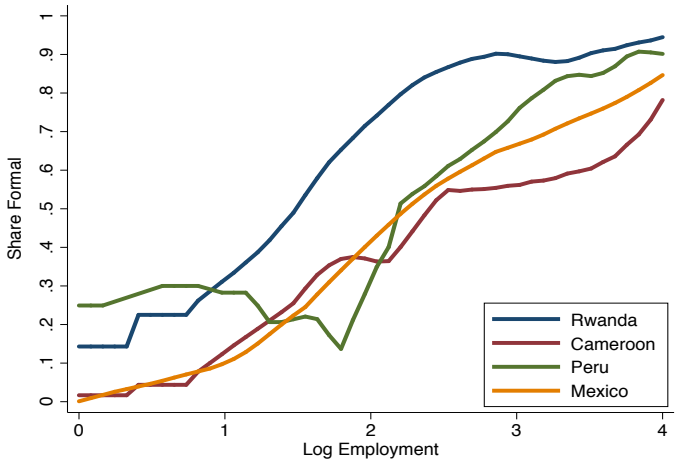
We use two alternative formality assignment rules to examine the robustness of our

[9]Relatedly, Gordon and Li (2009) explain the high shares of taxes on capital (such as corporate income taxes) in developing countries relative to rich countries by the fact that capital is more observable than labor in these countries. This also implies higher compliance rates amongst larger retailers.

[10]Formality is defined as 'being registered with tax authority' in Cameroon and Rwanda, and 'paying Value-Added-Taxes on sales' for Mexico and Peru.

Figure 1: Employment Size, Formality & Store Types

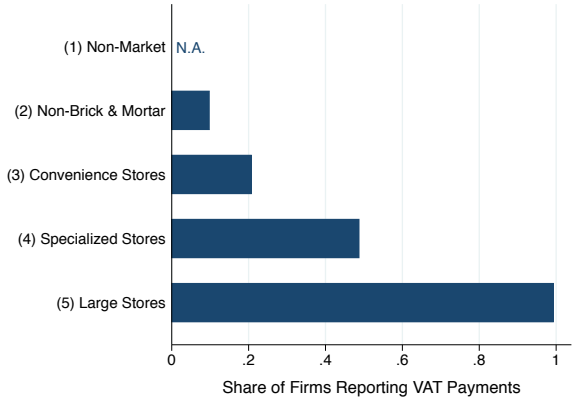
(a) # Employees on Formality in Retail Censuses



(b) # Employees by Store in Mexico



(c) % Paying VAT by Store in Mexico



results. First, we consider an alternative assignment for the place of purchase category for which there is the most uncertainty, category (4) specialized stores. This category is more likely to contain a mix of formal and informal stores. Our baseline scenario classifies it as formal, so our first robustness scenario assigns it instead to the informal sector and leaves all others unchanged. Our second alternative assignment is probabilistic: to take into account the fact that some small retailers pay taxes, we assign to each category of place of purchase a probability that it is formal, using the formality rates observed in the Mexican retail census.[11]

Our formality assignment rule thus enables us to measure, for each household in

[11]Formally we assign the following probabilities that the place of purchase is formal, following the evidence in Figure 1 : 0.1 for category (2), 0.2 for category (3), 0.5 for category (4) and 0.9 for category (5). Other categories are unchanged.

our data, its informal budget share: the share of its total expenditure purchased from the informal sector. In what follows we also consider within-product informal budget shares. We use the UN's detailed COICOP classification of products when considering product level budget shares, this is available at the 2-digit level (12 products), 3-digit level (58 products) and 4-digit level (117 products).[12]

2.3 Large Sample for Product Specific Expenditures

We use the micro data from the World Bank's Global Consumption Database (GCD) to study product-level Engel curves and compare our core sample to a wider sample. This data is the most comprehensive database on consumption patterns in developing countries: it compiles household expenditure surveys across countries and harmonizes product categories across all surveys at the 2-digit COICOP product level.[13] Merging this dataset with our core sample we obtain our large sample with data for 89 countries which represents over 50% of the world population.

We use this large sample to characterize food Engel curves. We define as food products all items falling in the COICOP 2-digit category 'food and non-alcoholic beverages'. This category is a good proxy for the set of products which most governments throughout the world tax at a reduced rate in an attempt to make their consumption tax system more progressive. We compute for each household in the large sample the share of its total expenditure spent on food products and study the correlation between this share and total household expenditure.

3 Informality and Food Engel Curves Across Countries

We start this section by documenting new facts on the correlation between informal consumption and income both across and within countries, using our core sample of countries. We then explore potential explanations for the within-country patterns we observe. We finally provide new evidence on the food Engel curve, confirming patterns already established in the literature, and allowing us to compare along this dimension our core sample of 31 countries to the larger sample of 89 countries.

[12] When COICOP are not available we use crosswalks to convert survey-specific product categories to COICOP categories. This could not be done for three countries: Brazil, Peru and Tunisia. For these countries we use survey-specific product categories at the 3 and 4 digit levels.

[13] Aggregate statistics are available at http://datatopics.worldbank.org/consumption. Appendix B.3 provides further information on the dataset and our merge.

3.1 Informality Engel Curves

3.1.1 Results

Our main object of interest is the within-country relationship between the informal expenditure share and household total expenditure, which we call the Informality Engel Curve (IEC). We follow the literature on Engel curves in developing countries and consider the relationship between expenditure shares and total household expenditure rather than income (Deaton and Paxson, 1998; Atkin et al., 2018a), because total expenditures are measured with less error than income in household expenditure surveys (see Deaton, 1997b, for a detailed exposition of this problem). We plot expenditure share as a function of log total household expenditure, following the literature on product-specific Engel curves (Working, 1943).

Figure 2 presents non-parametric evidence on the IEC for two countries in our core sample, Mexico and Rwanda. The horizontal axis is the log per capita expenditure per person in 2010 constant USD. The solid grey line corresponds to the median of each country's expenditure distribution, and the dotted lines correspond to the 5^{th} and 95^{th} percentiles. We plot kernel-weighted polynomial local regressions of informal expenditure share on log expenditure, the shaded area represents the 95% confidence interval. In both countries IECs are downward-sloping and approximately linear with respect to log household expenditure. In Mexico, the informal expenditure share drops over the income distribution, from 60% to 20%. In Rwanda the IEC is slightly less steep: the informal share falls from over 90% for the poorest households to 65% for the richest. Appendix Figure A.1 plots the relationship between informal shares and household expenditure in the same way for each country in our sample. We observe that IECs are always downward sloping and that approximately linear with respect to log expenditure.

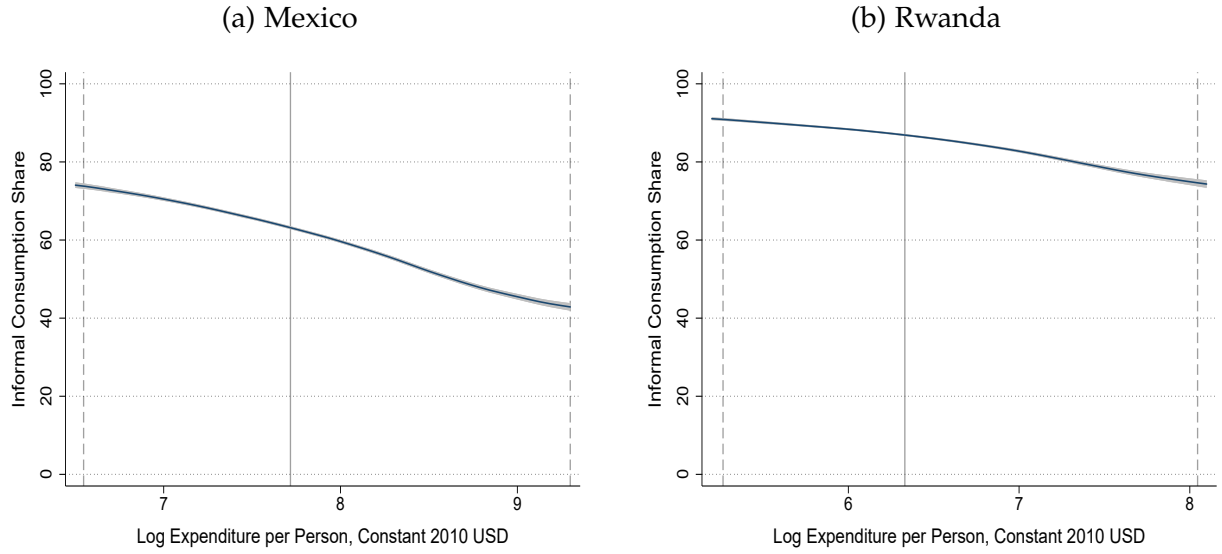
To summarize the information contained in each IEC and compare across countries we focus on two key moments: the informal consumption share of total consumption and the slope of the IECs. To estimate the slope we impose a linear relationship with respect to log income.[14] We obtain the slope from running the following regression separately in each country:

$$Share\ Informal_i = \beta ln(expenditure_i) + \varepsilon_i \tag{1}$$

where $Share\ Informal_i$ is the informal expenditure share of household i, $expenditure_i$ is its total expenditure per person, and we use household weights from each survey.

[14] While the existing literature has found that product-specific Engel curves can be non-linear (Banks et al., 1997; Atkin et al., 2018a), linearity appears to be a reasonable approximation in our setting.

Figure 2: Informality Engel Curves: Examples



Local polynomial fit of the Informality Engel Curves in Mexico and Rwanda. Expenditure on the horizontal axis is measured in log. Informal expenditure shares are on the vertical axis. The shaded area around the polynomial fit corresponds to the 95% confidence interval. The solid grey line corresponds to the median of each country's expenditure distribution, while the dotted lines correspond to the 5th and 95th percentiles. See Appendix Figure A.1 for similar graphs for each country in our core sample.

Results are presented in Figure 3. Panel (a) plots informal consumption as a share of total consumption, on the country's per capita GDP (log of constant 2010 USD). The average informal share is 52%, and we see a large negative correlation with the level of economic development: the average informal expenditures is close to 80% in many Sub Saharan Africa countries, compared to only 20% in upper middle income countries. This negative correlation between economic development and size of the informal sector is consistent with other studies in the literature which attempts to proxy for the size of the informal sector with different methodologies (see in particular Enste and Schneider, 2000; La Porta and Shleifer, 2014; Morrow et al., 2019). These similarities between existing measures of the informal sector and our consumption-based approach suggests that our formality assignment method does a reasonable job at proxying for the size of the informal sector across countries.

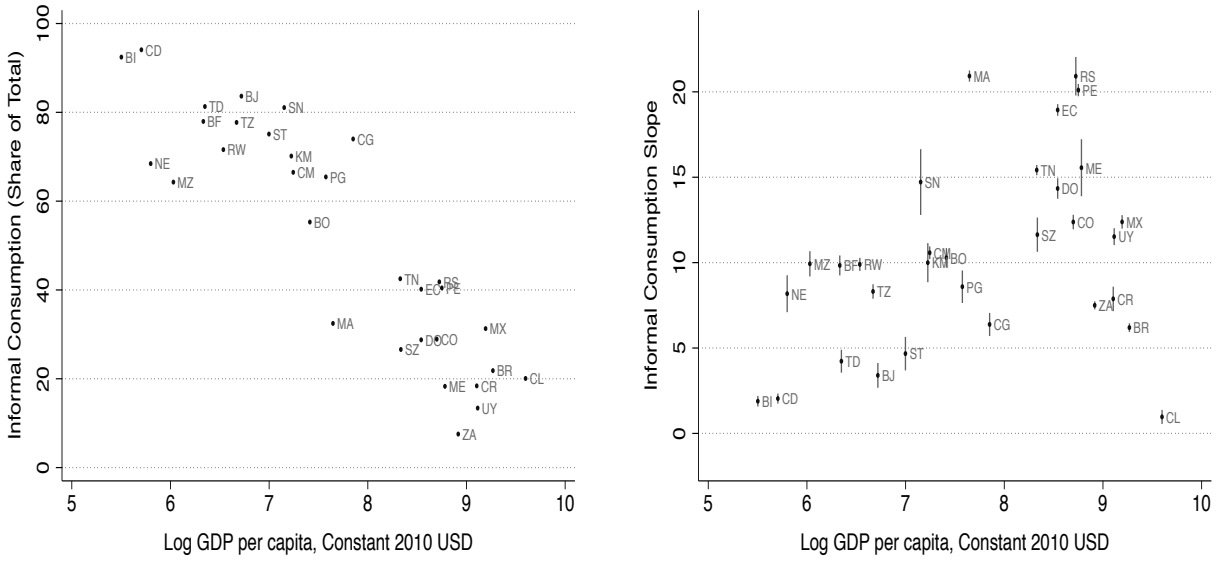
Our method enables us to go further than existing approaches by considering the correlation between informal consumption share and income at the household level, for each country in our sample. Panel (b) plots (minus) the estimated IEC slopes (the $-\beta$ coefficient in equation 1 above) as a function of economic development. In all countries the IEC slopes are negative, and significantly different from zero. The average slope across all countries is 10.3: on average the informal consumption share is reduced by 1.03 percentage points when household expenditure increases by 10%. Beyond the average

slope, we note that the range of slopes varies from 2 in Chile and Burundi to 20 in Morocco and Peru, and that the IECs increase when going from low to middle income countries .

Figure 3: Informal Consumption Across Countries

(a) Average: Informal Consumption Share

(b) Slope: Informality Engel Curve



Panel (a) plots country-level informal consumption as a share of total consumption, on log per capita GDP. The average informal consumption of the core sample composed of 31 countries is 52%. Panel (b) shows the (inverse) slope of informal consumption with total per person household expenditure, on log per capita GDP. The bars correspond to the 95% confidence interval of the slope coefficient. GDP per capita is in constant 2010 USD (Source: World Bank WDI). The average slopes of the core sample of the 31 countries is 10.3.

How do the above results vary based on the assignment of places of purchase to formality status? We present results for two alternative assignment scenarios in Figure A.2. Using these other assignment rules slightly changes the informal consumption averages, but estimates of the IEC slopse are close to those from our baseline assignment rule. The results are robust to changing the assumptions we make to reflect the underlying uncertainty in the formality status of different place of purchase categories. Figure A.5 provides further details: it shows average consumption for each decile by place of purchase, taking the average across countries. We see that slopes with respect to total expenditure are particularly steep for categories for which the formality assignment is more certain, such as non-market purchases, large stores and institutional services.

3.1.2 Explaining differences in informal consumption across households

Why do poorer households consume a higher share of their budget from the informal sector? Here we investigate potential determinants for the observed differences in informal consumption patterns across households. We conduct this investigation of potential mechanisms for two reasons. First, the evidence that IECs are downward sloping in all countries is, to the best of our knowledge, new. Understanding why such consumption patterns arise is therefore of intrinsic interest. Second, studying the mechanisms behind our downward-sloping IECs is relevant from a tax policy perspective, as further discussed in the model of Section 4. If IEC slopes can be explained by characteristics governments can easily observe and target taxes and transfers on (such as preferences across large product categories), then 'de facto' exemption of informal consumption from taxation could also be achieved by alternative policies. If, on the other hand, informal expenditure shares are correlated with household incomes even conditional on such observables, then this exemption enables governments to obtain a more progressive tax system than they would otherwise.[15]

Observable household characteristics

We start by considering how much of the correlation between household income and informal consumption shares can be explained by observable characteristics. To do so we estimate the following regression separately for each country:

$$Share\ Informal_i = \beta * ln(expenditure_i) + \Gamma X_i + \varepsilon_i \tag{2}$$

where we gradually introduce household characteristics X_i as controls.

Table 2 presents the results. We show the average across countries of the coefficients β, the average upper and lower bounds of the 95 confidence intervals, and the number of countries for which the coefficient is statistically significant at the 95% level. Column 1 presents results obtained in a specification with no controls. In column 2 we control for household demographic characteristics: household size and age, education level and gender of the household head.[16] We find no evidence that these characteristics explain

[15]Suppose for example that all the differences in informal shares between poor and rich households can be explained by the fact that the poor spend a higher share of their budget on food, and that food products are a lot more likely to be sold in the informal sector than other products. Then, the existence of an informal sector would affect the progressivity of consumption taxes in the same way as exempting food products from taxation, a policy governments can (and often do) implement through legislation.

[16]The former controls for economies of scale across households of different size which could affect where households choose to shop (Deaton and Paxson, 1998).

the correlation between informal consumption and income, if anything the IEC slopes increase slightly.

Table 2: Average Slopes of the Informality Engel Curves

Specification:	**Main**		**Geography**		**Product Codes**				**All**
Avg. of 26 Countries	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Slope	10.3	11.2	9.6	8.8	6.9	6.3	6.1	5.2	4.2
Confidence Interval	[9.6,11.0]	[10.4,11.9]	[8.8,10.3]	[7.9,9.6]	[6.2,7.4]	[5.7,6.7]	[5.5,6.5]	[4.7,5.6]	[3.6,4.6]
# of p-values < 0.05	31	31	31	30	28	30	30	28	28
R^2 adjusted	0.20	0.22	0.27	0.42	0.46	0.52	0.55	0.55	0.58
Household Characteristics		X	X	X	X	X	X	X	X
Urban/Rural			X						
Census Blocks				X					X
Food Products					X				
COICOP 2-dig						X			
COICOP 3-dig							X		
COICOP 4-dig								X	X

This table shows the average slope of the Informal Engel Curves across countries for different specifications. The slopes are estimated from: $Share\,Informal_i = \beta.ln(expenditure\,pc)_i + \Gamma X_i + \varepsilon_i$. The dependent variable is informal expenditure share and the main explanatory variable is log expenditure per capita. Controls include household characteristics (household size, age, gender, education of head), location indicators (urban/rural, survey enumeration blocks), and product codes for food vs all other purchases, 2nd, 3rd and 4th level of the COICOP classification.

In columns 3 and 4 we add controls for households' location, first by using an indicator for whether the household lives in a rural area (column 3) then by adding survey block fixed effect.[17] This enables us to test whether differences in consumption arise because poorer households live in areas with worse access to formal stores. They may for example be less likely to live in dense urban centers where large modern stores are located. We find that whilst there are large differences in average informal expenditure shares between rural and urban households, as expected, controlling for rural locations explains only 15% of the slope.[18] Controlling for more detailed household location decreases the correlation further, showing that differences in market access by income level explain slightly above 20% of the differences in informal expenditure share.

In columns 5 to 8 we test whether non-homothetic preferences across products play a role: richer households could spend more on products predominantly sold in formal stores. To do this we run a product-level version of specification (2) with product fixed

[17]The survey block is the most granular location information in all surveys and contains on average in our surveys 74 individuals. The median survey block represents 52,900 households, with a country-level minimum of 2,029 households (Rwanda) and maximum of 818,229 households (Tanzania).

[18]The informal budget share is on average 67% in rural areas versus 52% in urban areas and IEC slopes are steeper within urban locations. See Figure A.8 for separate rural/urban level and slopes of informality.

effects and compute an average product-level estimate of β for each country.[19] We gradually consider variations within smaller product types: we first use control for food products, in column 5, then use the COICOP 2-digit level classification in column 6, the 3-digit level classification in column 7 and the most granular 4-digit level classification in column 8. We find that different preferences across products do explain part of the correlation: controlling for food products alone explains 38% of the slopes and controlling for the 12 broad product categories explains 44% of the slope. Controlling for narrow product categories slightly reduces the slope further, to an average of 5.2.[20]

Finally, column 9 shows the average IEC slope when all controls are included. The average IEC slope once all controls are included is 4.2, and it remains statistically significant in all but three countries in our sample. Overall, observable household characteristics explains 62% of the correlation between informal expenditure shares and household income. We reproduce these results for our two alternative assignment scenarios (Tables A2 and A3) and find qualitatively very similar results.

Suggestive evidence on the role of quality

In what follows we investigate whether what remains of the correlation between informal expenditure shares and income once household characteristics are controlled for can be explained by the fact that richer households value high-quality goods more, and that these are more commonly found in formal stores. Faber and Fally (2017) find evidence that this is the case in the US context, whilst Atkin et al. (2018b) find that richer households in Mexico spend more on the high-quality products sold by foreign retailers.

Direct evidence in line with this hypothesis can be found in four countries in which the expenditure modules ask households the main reason they choose a particular place of purchase for each product. In Appendix Table A4 we list the share of households reporting access, price, quality and store attributes as the main reason they choose to shop in a store, separately for formal and informal stores.[21] We find that in all countries households are more likely to report price than quality as the reason for choosing informal stores, and the reverse for formal stores. This price-quality trade-off seems

[19] Formally we run the following regression: $Share\,Informal_{ip} = \beta * ln(expenditure_i) + \alpha_p + \Gamma X_i + \varepsilon_{ip}$ where $Share\,Informal_{ip}$ is the share of household i's informal expenditure on product p, and α_p are product fixed effects. We weigh each observation using household survey weights and the expenditure share of the product.

[20] Figure A.7 displays visually these results for each country by showing the slopes when controlling for increasingly narrow product groups.

[21] Access is defined as a combination of necessity and proximity, and store attributes include quality of reception, credit availability, and homogeneity of products on display.

quantitatively important: in all but one country price (quality) is the main reason given by over a third of households for choosing an informal (formal) stores.[22]

We provide further suggestive evidence on this hypothesis by estimating differences in prices across formal and informal sources of consumption. We use unit values to proxy for prices, these are available for 10 countries in our core sample. We control for narrow product classification, units, and location, so interpret remaining differences in prices as indicative of possible quality differences, following the approach in Atkin et al. (2018b).[23] We limit our analysis to food products, for two reasons. First, while measurement concerns are always present when using unit values, these issues are arguably less prevalent when focusing on food items for which units are well documented in the surveys. Second, most food items are exempt from consumption taxes so price differences between formal and informal stores should not be due to the latter not paying taxes, and should capture the net of tax price differences across varieties. Formally, in each country with available data we estimate the following specification:

$$ln(unit\ value)_{ipmu} = \delta(Formal)_{ipmu} + \mu_{pmu} + \epsilon_{ipmu} \tag{3}$$

where $ln(unit\ value)_{ipmu}$ is the unit value reported by household i in location m, within product category p and sold in units category u, and $Formal_{ipmu}$ is equal to 1 if the product is purchased in a formal store. We control for fixed effects at the product category * unit * category * market level, where products refer to the most detailed product categories available in the survey, markets correspond to a geographical location, and we use the units provided in the survey.[24] This enables us to compare unit values for the same products sold in the same location, and in the same units.

Results are reported in Table 3.[25] Panel A compares prices in formal stores to prices in informal stores, and panel B provides a more detailed comparison between non-brick-and-mortar stores (the omitted category) and other store types. We see that formal stores charge significantly higher prices than informal stores in all but one country. The average price difference is 3.9% , and there is significant heterogeneity across countries, with the

[22]Note that access also seems to play an important role: between one-third and one-half of respondents report access as their main reason for choosing a store. This is consistent with the idea that many households in our sample cannot invest in costly durables, such as cars, which may give them access to a wider variety of stores (Lagakos, 2016).

[23]Product 'quality' should be understood here as a combination of both intrinsic product quality and store-level amenities such as parking, wide aisles, security and hygiene.

[24]An alternative is to create a cross-walk between units of reporting within products. Given limited details in most surveys, this exercise could not be completed with confidence. Consequently, we compare transactions within the same unit-reporting category. A drawback of our approach is that this restricts the overlap of comparison between store-types, as formal, modern stores are more likely to sell in bulk.

[25]Controlling for household demographic characteristics does not affect the results.

Table 3: Unit values across places of purchase

Country	(1) **Bolivia**	(2) **Brazil**	(3) **Burundi**	(4) **DRC**	(5) **Costa Rica**	(6) **Dom.R.**	(7) **Ecuador**	(8) **Morocco**	(9) **Mexico**	(10) **Tanzania**
A. Comparing sectors (omitted category: informal)										
Formal	3.68	-0.02	5.05	5.88	1.58	13.6	1.8	5.18	0.94	2.20
	(0.68)	*(0.129)*	*(2.54)*	*(1.67)*	*(0.3)*	*(0.66)*	*(0.15)*	*(0.25)*	*(0.18)*	*(0.21)*
B. Comparing places of purchase (omitted category: non-brick-and-mortar stores)										
Non-market consumption	-8.19	-3.21	-2.32	-1.02	14.4	2.92	3.29	4.69	*no obs.*	-2.42
	(1.27)	*(0.33)*	*(0.2)*	*(0.185)*	*(0.939)*	*(0.42)*	*(0.197)*	*(0.19)*		*(0.266)*
Convenience stores	4.31	0.108	*no obs.*	0.835	12.1	5.58	4.85	6.69	10.0	*no obs.*
	(3.78)	*(0.2)*		*(2.44)*	*(0.76)*	*(0.33)*	*(0.13)*	*(0.14)*	*(0.25)*	
Specialized stores	1.36	1.33	4.39	20.7	9.72	7.96	4.42	8.37	6.04	1.45
	(0.76)	*(0.21)*	*(5.38)*	*(3.11)*	*(0.8)*	*(1.06)*	*(0.21)*	*(0.27)*	*(0.26)*	*(0.22)*
Large stores	17.4	-0.60	4.7	-15.2	12.3	20.9	6.04	14.9	8.63	4.14
	(1.4)	*(0.19)*	*(3.27)*	*(1.88)*	*(0.73)*	*(0.91)*	*(0.26)*	*(0.84)*	*(0.34)*	*(2.48)*
Nb. of purchases	119,515	698,194	248,714	862,988	115,198	325,820	1,018,776	711,481	442,262	1,053,847
Nb. of geo cells	18	54	33	41	12	62	718	31	127	400
Nb. of product codes	176	968	139	190	308	215	315	245	167	154
Total nb. of fixed effects	1467	9022	2359	5389	1544	4215	11330	3524	5918	13184

Panel A presents results from regressions of log unit values on an indicator for whether the transaction occurred in a formal store. The omitted category is informal stores. Panel B presents results from regressions of log unit values on separate place of purchase indicators, with non-brick-and-mortar stores used as the omitted category. The coefficients are multiplied by 100 to obtain approximate percentage variations. Standard errors are in parentheses. The regressions include fixed effects for each product-market-reporting cell. To increase statistical precision, in every country we exclude cells which have less than 15 observations. More disaggregated geography levels allow for more refined market comparisons, but lead to a larger number of cells being excluded. We therefore choose the lowest geographical level subject to the constraint that non-excluded cells account for at least 75% of aggregate food expenditure. To remove the influence of outliers, we winsorize unit values at 5% and 95% within each country.

price difference ranging from 0 (Brazil) to 13.6% (Dominican Republic). The lower panel provides suggestive evidence that larger stores that are more likely to be formal charge higher prices. Both panels are therefore consistent with the hypothesis that more formal stores offer higher quality products, using prices to proxy for quality.[26]

Overall, the evidence presented in this sub-section suggests that differences in informal expenditure shares across households with different income levels can be explained by differences in preferences across product types, difference in preferences for quality, and market access.

[26]Formal and informal stores may differ in other ways which are reflected in prices, formal stores may in particular be more productive. Our estimates of δ may therefore capture both quality and productivity differences across stores; however this would lead us to under-estimate the true quality differences and make it less likely that we would find positive estimates of δ.

3.2 Food Engel Curves

Most governments set reduced tax rates on food products in an attempt to introduce progressivity to their consumption tax system.[27] This is motivated by a large literature showing that food Engel curves are downward-sloping – the food expenditure share falls with household income – and approximately linear with respect to log total expenditures. Recent papers have documented within-country food Engel curves using household expenditure surveys from several countries: 10 countries in Almås (2012)), 22 countries in Anker et al. (2011), and 38 countries in Pritchett and Spivack (2013). In this subsection we show evidence on food Engel curves on our uniquely large dataset of 89 countries, spanning a wide range of development levels.

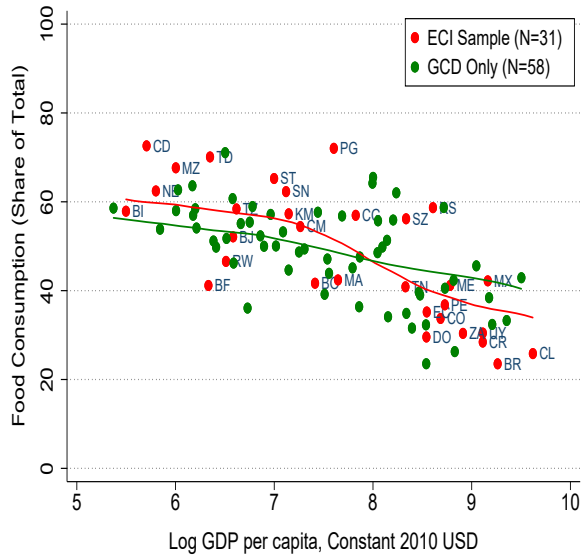
Figure 4 presents our results. Panel (a) plots the country-level aggregate share of food in total expenditures as a function of economic development. Countries that are in our core sample are presented in red, while countries which are only in the large sample are in green. Panel (b) plots minus the slopes of each country's food Engel curve, obtained from regressions of household food expenditure share on log household total expenditure. In Appendix A we plot food Engel curves non-parametrically for all countries in our core sample, we see that this specification is a good approximation of the shape of the Engel curves.

Not surprisingly, the food share of total expenditure falls with economic development (Panel a). In addition, food Engel curves are downward-sloping in all countries. There is no clear relationship between slope size and economic development, in line with the existing literature. We note that countries in our core sample resemble other countries at similar levels of development with respect to both their aggregate food expenditure share and the slopes of their food Engel curves. This is reassuring, as it suggests that these countries are likely to be representative of others of similar development level when it comes to the Informal Engel Curves. We can only document informal consumption patterns and characterize optimal tax policy in our smaller core sample, but this evidence indicates that our conclusions may be relevant more generally.

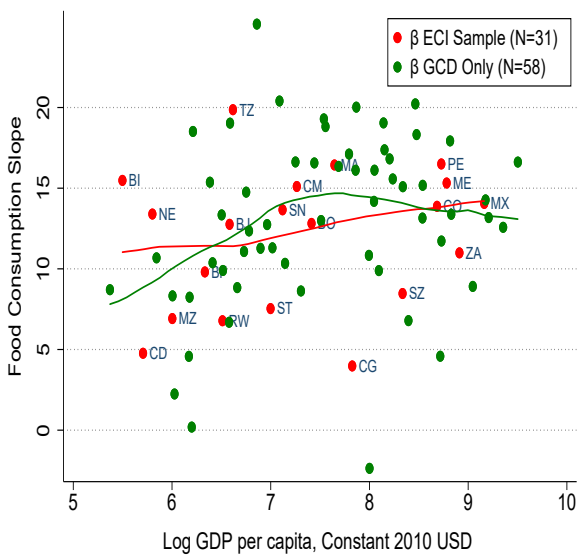
[27] Some countries apply reduced rates or exempt all food items, while other countries target 'basic' food items. In this paper, we consider a tax policy which reduces the rate on all food items. Targeting more narrow items may improve targeting but also allows firms to more easily misreport liable items as non-liable, leading to distortions and losses in tax revenue.

Figure 4: Food Consumption Across Countries

(a) Average: Food Consumption Share



(b) Slope: Food Engel Curve



Panel (a) shows country food consumption as a share of total consumption, plotted against log per capita GDP. The average food consumption of the core sample of 31 countries is 49%, while the average of the GCD sample, composed of 58 countries is 48%. Panel (b) shows the country-specific slope of food consumption with respect to log household expenditure, plotted against log per capita GDP. The slope measures the drop in food consumption for a doubling of households' income. GDP per capita is in constant 2010 USD, and transformed into log base 2 (Source: World Bank WDI). The average slope of food consumption of the core sample of 31 countries is 12.5, while the average of the GCD sample, composed of 58 countries is 13.

4 Optimal Consumption Tax Policy over the Development Path

This section considers the implications for commodity tax policy of our descriptive results. We introduce an informal sector in the multi-person Ramsey model of commodity taxation of Diamond (1975). We assume that there are two varieties of each product and consider what happens when one of these two varieties cannot be taxed. As is standard in this literature we assume full pass-through of taxes to prices of varieties produced in the formal sector and zero pass-through to prices of varieties produced in the informal sector.[28] We then consider how optimal tax policy changes over the development path. We focus throughout on two aspects of optimal tax policy: the optimal uniform commodity tax rate (or optimal level of commodity taxes) and optimal rate differentiation between food and non-food products. Proofs of all results discussed in this section can be found in Appendix C.

[28] Appendix C shows that a simple model in which formal and informal firms compete under monopolistic competition yields these patterns of pass-through.

4.1 Set-up

There is a continuum of mass one of households i with different incomes y^i. Households have preferences over J different products, and for each product j over two varieties v. The subscript $v = 0$ indicates a variety produced in the traditional sector, $v = 1$ a variety produced in the modern sector; we assume that the two varieties $j0$ and $j1$ of product j are substitutes. We consider below how optimal tax policy changes when the traditional sector cannot be taxed.

Producer prices q_{j0}, q_{j1} are exogenous. Consumer prices are given by $p_{jv} = q_{jv}(1+t_{jv})$ if the variety jv is taxed. When variety $j0$ is produced in the informal sector we have $p_{j0} = q_{j0}$. We write $V(p, y^i)$ the indirect utility of household i, s^i_{jv} the budget share that household i spends on variety jv, $s^i_j = s^i_{j0} + s^i_{j1}$ the budget share spent on product j, and ϵ_j the price elasticity of demand for product j, assumed negative. We assume that the product price elasticities ϵ_j and the variety price elasticities ϵ_{jv} are equal across all products (varieties), to focus on the behavioral responses that arise when we introduce an informal sector due to the substitution across varieties within each product. To simplify the exposition we also assume no cross-price elasticity between the different j products.[29] We adopt the following notation to denote average values taken across all households: $\int_i z_i di = \bar{z}$.

The government chooses the tax rates t_j to maximize:

$$W = \int_i G(v(p, y^i))di + \mu \sum_j T_j \tag{4}$$

where T_j is total tax revenues collected by taxing product j. Writing $x_{jv} = \int_i x^i_{jv}(p, y^i)di$ total demand for variety jv, we have $T_j = t_j q_{j1} x_{j1}$ when only modern varieties can be taxed and $T_j = t_j(q_{j0}x_{j0} + q_{j1}x_{j1})$ if all varieties can be taxed. Government preferences are characterized by μ the marginal value of public funds and $G()$ an increasing and concave social welfare function. We write g^i household i's social marginal welfare weight which represents how much the government values giving an extra unit of income to household i, and $\bar{g}$ the average social marginal welfare weight (see Saez and Santcheva, 2016).[30] We assume throughout that g^i is decreasing with household income and $\mu = \bar{g}$, the latter simplifies expressions and corresponds to a government that has no preference for taxation unless it enables redistribution.

[29] This assumption is relaxed in Appendix C.

[30] Formally $g^i = \frac{\partial G(v(p,y^i))}{\partial v(p,y^i)} \frac{\partial v(p,y^i)}{\partial y^i}$.

4.2 Optimal taxation in the absence of an informal sector

We start with a set-up in which both types of varieties are taxed, equivalent to that in Diamond (1975). Consider first the case of a government using a uniform tax rate, $t_j = t, \forall j$. Welfare maximization yields the following expression for the optimal uniform tax rate:

$$\tau^* = \frac{\int_i (\bar{g} - g^i)\phi^i di}{-\epsilon \bar{g}} \tag{5}$$

where $\epsilon = \sum_j \frac{x_j}{x} \epsilon_j$ is the average price elasticity of demand across all products, weighted by each product's share in total consumption, and $\phi^i = \frac{y^i}{\bar{y}}$ is household i's income relative to the average.[31] When all products and varieties are taxed the optimal uniform tax rate is thus positive, increasing in income inequality and in the strength of the government's redistributive preferences, captured by the variance in g_i, and decreasing in the the absolute value of the price elasticity of demand.

When the government sets different rates on products the optimal rate on any product j is given by:

$$\tau_j^* = \frac{\int_i (\bar{g} - g^i)\phi^i \frac{s_j^i}{s_j} di}{-\epsilon_j \bar{g}} \tag{6}$$

where $s_j = \bar{x}_j / \bar{y}$ is the aggregate budget share spent on product j. The optimal tax schedule sets lower rates on products that represent of higher budget share amongst poorer households. When the government simply sets a different rate on food (F) and non-food (NF) products, (6) implies that the optimal policy is to subsidize food relative to non-food products ($\tau_F^* < \tau_{NF}^*$) as long as the food Engel curve is downward-sloping.

4.3 Optimal taxation with an informal sector

Consider now a world with an informal sector, defined as a world in which the varieties $j0$ cannot be taxed. The optimal uniform consumption rate becomes:

$$\tau_1^* = \frac{\int_i (\bar{g} - g^i)\phi^i \frac{s_1^i}{s_1} di}{-\epsilon_1 \bar{g}} \tag{7}$$

This expression differs from expression (5) in two ways. First, the redistributive effect of a uniform rate on consumption is higher in the presence of an informal sector as long as poorer households spend a lower budget share on the formal sector than the

[31] The product price elasticity of demand is a function of the price elasticity of demand of both varieties of this product: $\epsilon_j = \epsilon_{j0} \frac{x_{j0}}{x}_j + \epsilon_{j1} \frac{x_{j1}}{x}_j$.

average ($\int_i (\bar{g} - g^i)\phi^i \frac{s_1^i}{s_1} di > \int_i (\bar{g} - g^i)\phi^i di$). The evidence in section 3.1 indicates that this condition is met: Informality Engel Curves are downward sloping in all countries. This implies that taxing only formal varieties shifts the burden of taxation towards richer households compared to taxing all varieties. Equity considerations therefore lead to higher optimal level of consumption taxes once the informal sector is taken int account.

Second, the efficiency cost of taxing consumption is now governed by the parameter $\epsilon_1 = \sum_j \frac{x_{j1}}{x_1} \epsilon_{j1}$, which is the weighted average price elasticity of all modern varieties. This elasticity is higher (in absolute value) than the price elasticity of all varieties ϵ – which governs the efficiency cost of taxation in a world without an informal sector – as long as the varieties are substitutes. Intuitively, as the prices of the modern varieties increase households substitute to traditional varieties, this increases the behavioral response to taxation: $\epsilon_1 < \epsilon$ (see Appendix C for a formal proof). Efficiency considerations thus imply that taking into account the informal sector decreases the optimal level of consumption taxes. The redistributive benefit of the existence of the informal sector outweighs its efficiency cost when IECs are very steep and/or households' willingness to substitute between the formal and informal varieties is low. In this case the optimal level of taxation is higher when we take into account the informal sector.

When the government can set different tax rates on products j the optimal rates τ_{j1}^* are given by:

$$\tau_{j1}^* = \frac{\int_i (\bar{g} - g^i)\phi^i \frac{s_{j1}^i}{s_{j1}} di}{-\epsilon_{j1}\bar{g}} \tag{8}$$

In the presence of an informal sector it is optimal to subsidize food relative to non-food ($\tau_{F1}^* < \tau_{NF1}^*$) if the *formal* food Engel curve is downward-sloping. We have shown above that food Engel curves are downward-sloping and that IECs are downward-sloping. If the food only IECs are very steep then the formal food Engel curves can be flat, or even upward-sloping. In this case it will no longer be optimal to subsidize food relative to non-food.

4.4 Optimal taxation over the development path

Having established how the presence of an informal sector impacts optimal tax policy, we now turn to the how optimal tax policy should change over the development path. We define economic development as an increase in households' income which takes the form of a uniform proportional change, so that the incomes of the rich and the poor increase by the same proportional amount. This assumes away potential changes in

inequality over the development path, and allows us to focus on the implications of the shape of Engel curves for optimal taxation.

Consider first a world without an informal sector. In this world, economic development does not impact the optimal uniform consumption rate, but does change the optimal product-specific rates through the shape of the Engel curves. The effect of a proportional increase in all households' incomes on the optimal rate on product j, τ_j^*, is given by:

$$\frac{\partial \tau_j^*}{\tau_j^*} = \frac{\int_i (\bar{g} - g^i)\phi^i \frac{s_j^i}{s_j}(\frac{\partial s_i}{s_i} - \frac{\partial s}{s})di}{\int_i (\bar{g} - g^i)\phi^i \frac{s_j^i}{s_j} di} + \frac{\partial \epsilon_j}{\epsilon_j} \tag{9}$$

Assume for now that the efficiency cost of taxing product j doesn't change with development, so that the last term drops out. This expression states that the optimal rate on product j increases with development if poorer households budget shares increase less (or decrease more) than the average budget share. In other words, taxing product j redistributes more with economic development (so that the optimal rate on product j increases) when the poors' share of the total consumption of the good decreases over the development path.

Applying this expression to the question of the optimal subsidy and food relative to non-food products and assuming a log-linear functional form on Engel curves enables us to reach a clear prediction regarding how this optimal relatival subsidy changes over the development path. With a log-linear food Engel curve the level changes in food budget shares with economic development are all equal to the slope of the Engel curve β_F: $\partial \bar{s}_F = \partial s_F^i = \beta_F, \forall i$. Food Engel curves are downward-sloping so we also have $\beta_F < 0$ and $\bar{s}_F < s_F^i$ if $g^i > \bar{g}$. This implies that as economies develop the food budget share of the poor falls proportionally less than the average budget share, so the optimal rate on food falls. If, in addition, the food budget share is not too large with respect to the non-food budget share, we can show that the optimal relative subsidy ($\frac{\tau_F^*}{\tau_{NF}^*}$) is decreasing with economic development.

We finally turn to how optimal policy changes over the development path in the presence of an informal sector. The effect of a proportional increase in all households' incomes on the optimal uniform rate τ_1^* is given by:

$$\frac{\partial \tau_1^*}{\tau_1^*} = \frac{\int_i (\bar{g} - g^i)\phi^i \frac{s_1^i}{s_1}(\frac{\partial s_1^i}{s_1^i} - \frac{\partial s_1}{s_1})di}{\int_i (\bar{g} - g^i)\phi^i \frac{s_1^i}{s_1} di} + \frac{\partial \epsilon_1}{\epsilon_1} \tag{10}$$

The first term captures the change in the redistributive effect of taxing only the formal sector over the development path. Here again imposing a log-linear functional form on Informality Engel Curves helps reach a clear conclusion. Log-linearity of the IEC implies that $\partial \bar{s}_1 = \partial s_1^i = \beta_1$ where β_1 is equal to minus the slope of the IEC. As long as the IEC is downward sloping the first term is negative, which pushes the optimal rate downwards over the development path. Intuitively, the shape of the IEC implies that as the economy grows the share of formal products consumed by the poor increases, so that taxing only the formal sector redistributes less. This is true even if the slope of the IEC does not change over the development path. If in addition this slope becomes smaller as countries develop the optimal rate will fall faster.

The second term captures the change in the efficiency cost of taxing only the formal sector as the economy grows. As explained above the behavioral response to commodity taxes is higher in the presence of an informal sector than in a world in which all varieties are taxed, since households can substitute consumption towards informal varieties as taxes on formal varieties rise. In Appendix C we show that, under reasonable assumptions, the price elasticity of formal varieties ϵ_1 is decreasing (and the efficiency cost of taxation increasing) in the size of the informal sector. Given the large decrease in informal sector size over the course of development (see Figure 3) we therefore expect the efficiency cost of taxing only formal varieties to fall with economic development. Efficiency considerations thus push optimal tax rates upward with development, unlike equity considerations.

The change in commodity-specific rate over the development path in the presence of an informal sector is given by a similar expression, outlined in Appendix C. Turning to the question of the optimal relative subsidy on food products, we show that the optimal relative food subsidy will still increase with development as long as i) the IEC for non-food products isn't much steeper than the IEC for food products, and ii) the informal share of consumption doesn't shrink substantially faster with development for food products than for non-food products.

5 Results: Optimal Tax policy and Progressivity

In this section we combine our model and empirical results to calibrate optimal tax rates and assess the progressivity of the tax system under three policy scenarios. We first consider how the optimal level of consumption taxes changes over the development path in the presence of an informal sector. Second, we calibrate optimal rates on food and non-food products and show how the optimal relative subsidy on food is affected by

the existence of an informal sector and economic development. We finally assess the progressivity of the tax system in each country in our sample: we consider progressivity achieved by 'de-jure' optimal food rate differentiation, 'de-facto' exemption of the informal sector and the combination of these two redistribution mechanisms.

5.1 Calibrated government and household preferences

We calibrate the optimal commodity tax rates defined in expressions (5) to (6) using the observed expenditure shares described in sections 3.1 and 3.2 above. In each country we aggregate the data for each decile of the total expenditure distribution to compute budget shares. We calibrate government preferences so that the government places twice as much weight on income received by households in the poorest decile than on income received by households in the richest decile.[32] Our baseline results assume no differences in inequality across countries: we set the value of β_i equal to the average across countries for each decile. This is in line with our model's assumption and enables us to concentrate on the role played by the shape of Engel curves in driving differences in optimal tax policy over the development path. We consider results using the observed β_i in each country (using total expenditures to proxy for income) at the end of this section.

We set the price elasticity of demand ϵ_j to -1 for all products at baseline and present two set of calibrated optimal tax rates with different assumptions regarding the price elasticity of demand for formal varieties ϵ_{j1}. The first assumes that these price elasticities, and therefore the efficiency cost of taxation, do not change with the size of the informal sector: we set $\epsilon_{j1} = -1$ for all products. The only source of variation across countries in optimal rates therefore comes from household budget shares. Our second set of results allows the price elasticity of demand for the formal variety of product j to decrease with the share of informal consumption in total consumption of product j. This increases the efficiency cost of taxation when we assume only formal varieties can be taxed. We show in Appendix C that we can write the price elasticity of the formal variety of product j as a function of the price elasticity of product j, the informal consumption share for j, income elasticities of demand and the elasticity of substitution between formal and informal varieties. We use our data to calibrate informal consumption shares for each product, and calibrate income elasticities using our estimates of the slopes of the Engel curves - see Appendix D for more details. We assume an elasticity of substitution between formal and informal varieties of 2, within the range of estimates obtained by Faber and Fally (2017) and Atkin et al. (2018b)

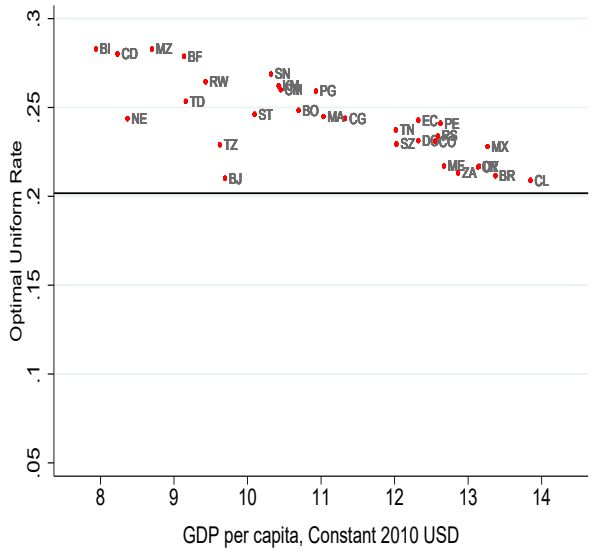
[32] Formally we set the marginal social welfare weight g_i equal to 2 for households in the bottom decile, decreasing in steps of 0.1, such that households in the top decile have a weight of 1.

5.2 Results: optimal tax policy over the development path

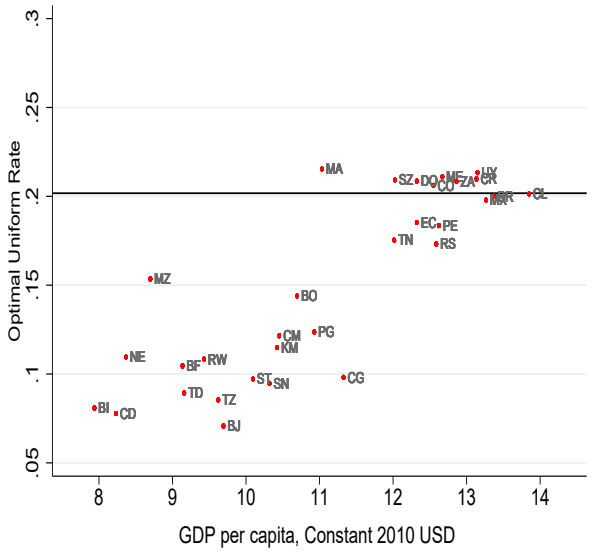
Figure 5 presents the optimal uniform tax rate for each country in our core sample, as a function of economic per capita GDP. The black line indicates the optimal uniform rate obtained under the assumption that there is no informal sector (both varieties can be taxed) – expression (5) above. This does not vary with development, as expected.[33] The red dots are the optimal uniform rates obtained when we assume informal varieties are untaxed – expression (7). In panel (a) we assume that the price elasticity of demand of formal varieties is the same in all countries. We see that optimal rates are decreasing with economic development, as predicted by the model: from 24% in the poorest countries in our sample where most household expenditure is informal, to 14% – the optimal rate in the absence of an informal sector – in Chile, where the informal sector is small and the IEC is nearly flat. Optimal rates look very different in panel (b) in which we allow the efficiency cost of taxation to increase with the size of the informal sector. We see that the efficiency effect dominates overall: as informal sectors shrink the efficiency cost of taxation falls and optimal rates are now increasing with economic development.

Figure 5: Optimal uniform rates over the development path

(a) No change in efficiency



(b) Efficiency change over development



Both panels plot the optimal uniform rate in each country as a function of the country's log per capita GDP. Optimal rates are obtained by calibrating expression XX. In panel (a) we assume that price elasticities do not change with economic development, in panel (b) we allow price elasticities to change over the course of development, following a specification detailed in appendix C.

[33] Allowing for different levels of inequality introduces variations in the optimal rate even in the absence of an informal sector, see Appendix Figure A.9.

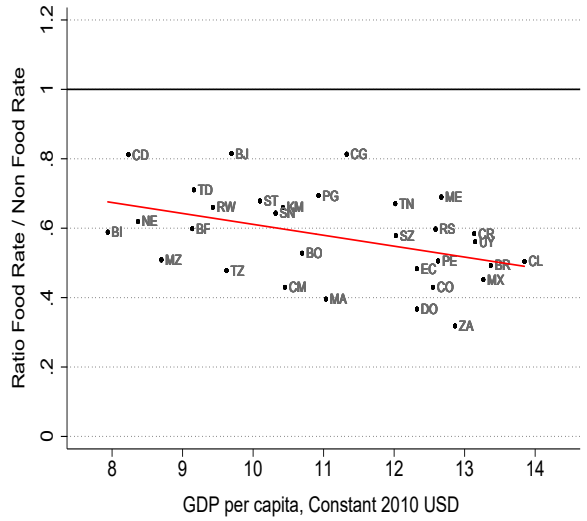
We then look at the optimal tax rates on food and non-food products, a commodity tax schedule used throughout the world with the objective of introducing progressivity to consumption taxes. Figure 6, panel (a), shows the ratio of the optimal food rate to the optimal non-food rate obtained under the assumption that both the modern and traditional varieties can be taxed (the counterfactual world without an informal sector – see expression (6) above). A ratio equal to one means that the optimal policy is to tax all products uniformly. We find a ratio of less than one in all countries, indicating that subsidizing food relative to non-food products is always optimal in the absence of an informal sector – as expected given the downward-sloping food Engel curves described above. On average across all countries the optimal rate on food products is slightly less than half the optimal rate on non-food products. The ratios are higher in poorer countries and hence the optimal rate on food is higher in those countries. This is due to the approximate linearity of food Engel curves with respect to log expenditures which makes subsidizing food products less attractive in poorer countries.[34]

Figure 6, panels (b) and (c) present ratios of optimal rates on food relative to non-food products under the more realistic assumption that informal varieties cannot be taxed. In panel (b) we assume that price elasticities are fixed across countries, so that differences in optimal rates only reflect equity considerations. We see that the existence of an informal sector increases these ratios in most countries, and hence increases the optimal relative rate on food. This is particularly true in low-income countries where poorer households consume most of their food in the informal sector and food IECs are even steeper: we see that the ratios are higher than 1 in some countries, indicating that the optimal rate on food is (slightly) higher than that on non-food products. Intuitively, this is because *formal* food Engel curves are not downward-sloping in the poorest countries, in which poorer households consume nearly all their food in the informal sector but still consume a non-trivial share of non-food products in the formal sector. Subsidizing food relative to non-food therefore entails subsidizing richer households' consumption more, so subsidizing food relative to non-food is simply not optimal from an equity perspective. Optimal rates in panel (c) take into account efficiency considerations by allowing price elasticities of both food and non-food formal varieties to vary with the size of the informal sector amongst food and non-food consumption. This affects our results in two ways. First, the optimal relative food subsidy falls in most countries. This is because the informal share of consumption is typically higher for food products than for non-food products. This

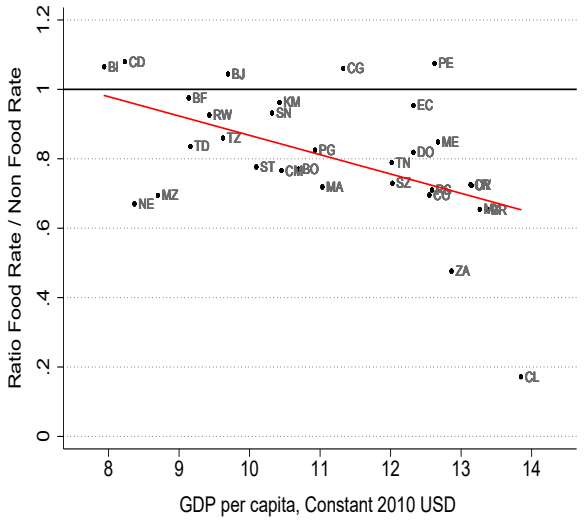
[34]This result would not hold for very high shares of food in total expenditures. In simulations we find that for food shares of 60% and above the relative food subsidy decreases with development. In practice only the very poorest countries in the world have food shares above 60%, as shown in Figure 4 panel (a)

Figure 6: Optimal Relative Rates on Food and Non-Food Products

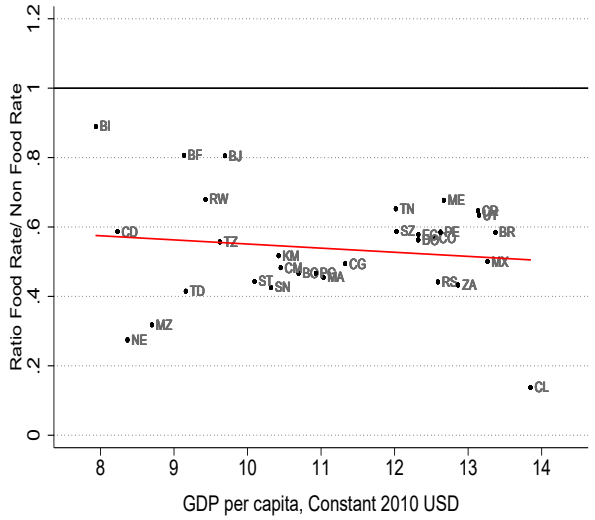
(a) All varieties taxed



(b) Only modern varieties taxed



(c) Only modern varieties taxed, efficiency change over development



All panels plot the ratio of the optimal rate on food products relative to the optimal rate on non-food products. A value equal to 1 indicates that both optimal rates are set at the same level, a lower value indicates that it is optimal to subsidize food products relative to non-food products. In panel (a) we assume that all varieties can be taxed, in panels (b) and (c) we assume that only modern varieties can be taxed. In panels (a) and (b) price elasticities do not change with economic development and in panel (c) we allow price elasticities to change over the course of development, following a specification detailed in appendix C.

increases the efficiency cost of taxing food relative to non-food products, and therefore lowers the optimal relative tax rate on food. Second, the optimal relative food rate still falls with economic development but significantly less than in panels (a) or (b). To understand this note that the informal share of consumption falls faster with economic

development for food than for non-food products, leading to a decrease in the relative efficiency cost of taxing food over the development path. Overall, we find that the optimal relative food subsidy increases with economic development, but less so once we take efficiency considerations into account.

Appendix Figures A.9 and A.10 present optimal tax results when we allow for different levels of inequality in each country. Once inequality is introduced, optimal uniform rates differ across countries even when both varieties are assumed taxed (more inequality implies a higher tax rate, see expression (5) above), but our main takeaway is unchanged.

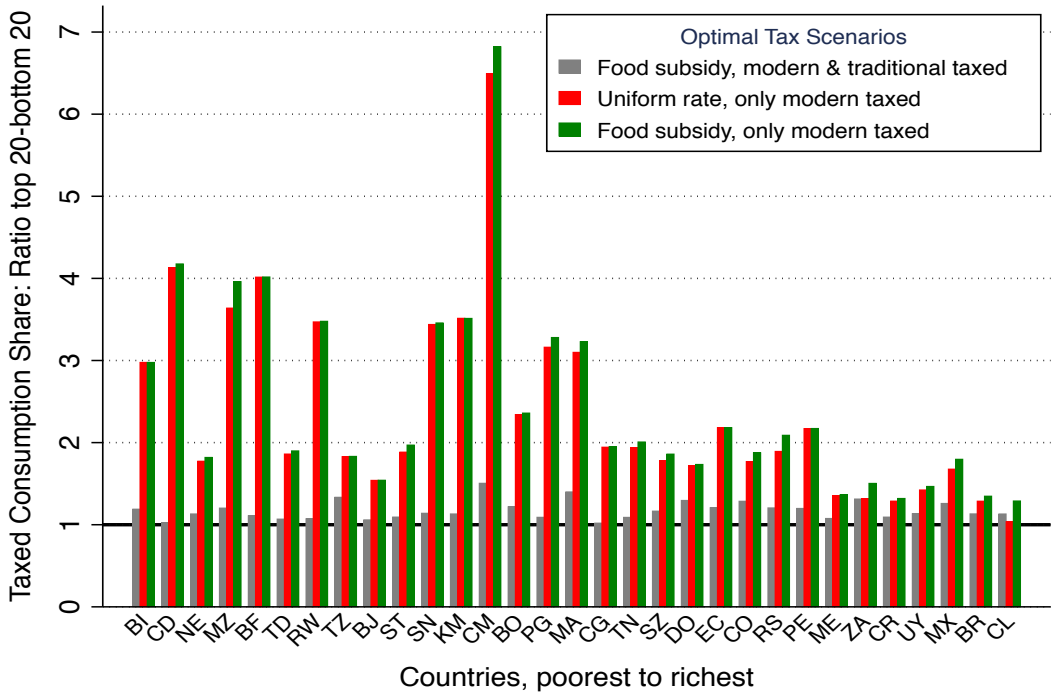
5.3 Progressivity of Optimal Consumption Taxes Across Countries

Finally, we turn to the distributional implications of our results. We use the calibrated tax rates and the observed consumption shares to compute effective tax rates by household income level and characterize the progressivity achieved under different optimal tax schedules. In Figure 7 we show the effective tax rates faced by the top 20% richest households relative to those faced by the 20% poorest in each core country sample. The first scenario, depicted in grey, considers the progressivity of policies that optimally differentiate rates on food and non-food products: on average this policy implies that the poorest households' effective tax rate is only 25% lower than that of the richest. This confirms the theoretical result in Sah (1983b) who argues that optimally differentiating commodity tax rates is unlikely to lead to much redistribution across households.

The second scenario, depicted in red, shows the progressivity achieved with a uniform rate once the informal sector is taken into account, the novel channel of this paper. We find that the effective tax rates paid by the top 20% are now 2.5 times higher than those paid by the bottom 20% on average. Progressivity is highest amongst low and lower-middle income countries, where the top 20% pay a 3 times higher effective tax rate than that of the bottom 20%. Taking into account distributional consumption in the informal sector leads to findings which differ from the consensus in the existing literature: we find that consumption taxes are an effective redistribution instrument when informal sectors are large and IECs downward-sloping.

The third and final scenario, depicted in green, shows the progressivity achieved once the informal sector is taken into account and we use the optimal rates on food and non-food products in Figure 6, Panel b). By construction this policy has to be more progressive than simply setting a uniform rate (the policy depicted in red). We find however, that rate differentiation only marginally improves progressivity: The effective tax rate ratio of the top 20% to the bottom 20% increase from 2.5 to 2.6 on average. This marginal increase is smallest in the poorest countries, in which optimally differentiat-

Figure 7: Progressivity of Consumption Taxes



This figure shows the ratio of effective tax rates on consumption faced by the top 20% richest households to the bottom 20%, for each country in our core sample. The grey bars shows the ratios in the counterfactual scenario where there is no informal sector and food products are optimally tax differentiated. The red bars shows the progressivity achieved when considering the existence of an informal sector but a uniform tax rate. Finally the green bars consider the existence of an informal sector and in addition the optimal food subsidy.

ing tax rates achieves almost no extra progressivity. Results are very similar when we calibrate optimal tax rates and compute the effective tax rates by decile in each country using our two alternative formality assignment methods. Appendix Figures A.13 and A.14 present results by country for our robust and probabilistic scenarios; the average ratio of effective tax rates falls slightly, to respectively 2.5 and 2.3.

5.4 Discussion and Extensions

Saving rates increasing with household income

So far we have not considered one of the standard arguments for the regressivity of consumption taxes, which is that savings rate are larger for the poor than for the rich. We note that this argument is accurate in a static model, but that if savings become future consumption, without changing the ordering of households' consumption deciles, then this argument is muted. Nonetheless, we try to address it here by modelling plausible distributional savings rate and showing their impact on progressivity.

Distributional savings rate are notoriously difficult to obtain, as they require accurate measures of both consumption and income at the household level. As previously mentioned, income is often poorly measured in low-income countries' household surveys, and the literature has been cautious in measuring savings from such data. A few countries collect instead dedicated consumer finance surveys, which aim to directly measure savings. Such data from the US indicates a range of savings from 0% for the bottom quintile to 12% for the top quintile.[35] We plug these distributional savings rates into our model, which makes the taxable base (consumption) smaller than income for rich households while it stays the same for the poorest households. This implies that optimal consumption taxes, either with an informal sector or optimally differentiated, become worse at redistributing. Hence both optimal uniform rates and rate differentiation are reduced, and the metric of comparison for redistribution becomes the share of income taxed, which also lowers measured progressivity (instead of the share of consumption taxed). Figure A.11 assesses the progressivity of consumption taxes when including distributional savings, equivalent to figure 7. Redistribution falls on average by 8% when considering savings: in the scenario with an optimal uniform rate and an informal sector (in green) the effective tax rate on the top to the bottom quintile falls from 2.6 to 2.4.

Pass-through of Taxes in the Informal Sector

Our analysis thus far ignores the possibility that informal firms might pay some taxes by purchasing intermediate inputs from formal firms which charge commodity taxes on their products (in a VAT system), or through general-equilibrium competition effects. In practice, we know that formal firms partially withhold taxes from their informal clients if the consumption tax takes the form of a Value-Added-Tax, as is the case in most countries (Keen, 2008). Appendix C shows that, when informal firms purchase some of their inputs from formal suppliers, the pass-through of taxes to informal retail prices is no longer zero but equal to the share of formal input costs in the informal retailers' total input costs, which reduces the progressivity of consumption taxes. There is evidence of substantial market segmentation between firms with different tax status in supply chains (Gadenne et al., 2019), even amongst formal firms. In work in progress we find that a reform in Mexico which increased the consumption tax rate had no significant effect on prices in the informal sector - a pass-through rate close to zero. This suggests our assumption of zero pass-through to informal prices may be reasonable.

Nevertheless, we use the 2013 Mexican firm census to document how the fact that

[35] Data based on the 1985 US Consumer Financial Survey.

some informal firms purchase inputs on which consumption taxes are paid affects our results. Among the 1.9 million retail firms, 85% report not paying any VAT on their sales - we classify those as informal. Amongst those, 8% report paying VAT on inputs, which conditionally on paying, is paid on just under 40% of their intermediate inputs. These firms represent 25% of informal sales, so this implies an aggregate formal input share in the informal sector of 10%. To take this into consideration we run our analysis under the assumption that tax pass-through is 10% in informal stores instead of 0.[36] Results are presented in Figure A.12: allowing for tax pass-through on informal prices decreases the redistribution achieved by a uniform tax rate thanks to the informal sector by 20-25%. Simulations suggest that the existence of an informal sector no longer makes commodity taxes more progressive than optimally differentiating rates on food versus non-food in the absence of an informal sector for a tax pass-through of 50% in informal stores.[37]

6 Conclusion

In this paper we study how consumption patterns vary with income both within and across countries and derive its implications for the design and redistributive potential of consumption taxes. We consider two key channels for redistribution: the usual channel of lower rating food products and a novel channel of informal consumption. We find that the informal expenditure share is decreasing in income in our core sample of 31 countries - a relation we denote as the Informality Engel Curve (IEC); we document that in all countries, the IECs remain robustly estimated within narrow market and product groups. We derive expressions for optimal commodity taxes in the presence of informal consumption, and calibrate optimal taxes for all countries in our sample, using the observed consumption patterns. We find that the informal sector makes optimal consumption taxes progressive as long as informal sector pass-though is limited, as appears to be the case empirically. This result runs counter to the consensus view that consumption taxes are distributionally neutral at best. In contrast, we find that the ubiquitous policy of optimal rate differentiation of food versus non-food products only marginally increases progressivity, and is particularly ineffective in low income countries.

What are the policy implications of our findings? Our results based on IEC do not imply that enforcement efforts to reduce the size of the informal sector should be

[36] We apply this pass-through rate to all expenditures in the informal sector including non-market consumption, which is likely to be unaffected. This should therefore provide an upper bound.

[37] Formally, the ratio of the effective rate paid by the top 20% to that paid by the bottom 20% is on average the same when rates are optimally differentiated in the absence of an informal sector (green lines) as when a uniform rate is set and we assume a 50% tax pass-through to the informal sector (red lines).

abandoned; rather, our findings caution that benefits from reducing the informal sector should be weighted against the adverse distributional costs. Our findings have sharper implications for the use of exemptions and reduced rates on necessities and food items in particular. In low income countries, we find that rate-differentiation has a very limited redistributive impact once informal consumption is considered. In these countries, where rate-differentiation can potentially generate large firm-distortions, there appears to be no equity motive for maintaining this redistributive tool. In practice, removing subsidies on food and other necessities is often met with fierce political resistance.[38] To be implementable, the removal of subsidies should therefore be combined with an increase in redistributive transfer programs (Hanna and Olken (2018)) and investments to develop more progressive income taxes.

References

Ahmad, E. and N. Stern (1984): "The theory of reform and Indian indirect taxes," *Journal of Public economics*, 25, 259–298.

Allingham, M. G. and A. Sandmo (1972): "Income tax evasion: a theoretical analysis," *Journal of Public Economics*, 1, 323–338.

Almås, I. (2012): "International Income Inequality: Measuring PPP bias by estimating Engel curves for food," *American Economic Review*, 102, 1093–1117.

Alvaredo, F., L. Chancel, T. Piketty, E. Saez, and G. Zucman (2018): *World inequality report 2018*, Belknap Press.

Anker, R. et al. (2011): "Engel's law around the world 150 years later," *Political Economy Research Institute Working Paper No*, 247.

Atkin, D., B. Faber, T. Fally, and M. Gonzalez-Navarro (2018a): "A New Engel on the Gains from Trade," Tech. rep., Working paper.

Atkin, D., B. Faber, and M. Gonzalez-Navarro (2018b): "Retail globalization and household welfare: Evidence from mexico," *Journal of Political Economy*, 126, 1–73.

Banks, J., R. Blundell, and A. Lewbel (1997): "Quadratic Engel curves and consumer demand," *Review of Economics and statistics*, 79, 527–539.

[38]The current civil unrest in Ecuador was triggered by a government decision to reduce fuel subsidies on October 2, 2019 (media-link). Removal of subsidies has in recent years similarly sparked social protest in Nigeria (media-link) and Iran (media-link).

Best, M., A. Brockmeyer, H. Kleven, J. Spinnewijn, and M. Waseem (2015): "Production vs Revenue Efficiency With Limited Tax Capacity: Theory and Evidence From Pakistan," *Journal of Political Economy*, 123(6), 1311–1355.

Bick, A., N. Fuchs-Schündeln, and D. Lagakos (2018): "How do hours worked vary with income? Cross-country evidence and implications," *American Economic Review*, 108, 170–99.

Boadway, R. and M. Sato (2009): "Optimal tax design and enforcement with an informal sector," *American Economic Journal: Economic Policy*, 1, 1–27.

Bronnenberg, B. J. and P. B. Ellickson (2015): "Adolescence and the path to maturity in global retail," *Journal of Economic Perspectives*, 29, 113–34.

Burgess, R. and N. Stern (1993): "Taxation and development," *Journal of economic literature*, 31, 762–830.

Cremer, H. and F. Gahvari (1993): "Tax evasion and optimal commodity taxation," *Journal of Public Economics*, 50, 261–275.

De Paula, A. and J. A. Scheinkman (2010): "Value-added taxes, chain effects, and informality," *American Economic Journal: Macroeconomics*, 2, 195–221.

Deaton, A. (1997a): *The analysis of household surveys: a microeconometric approach to development policy*, The World Bank.

——— (1997b): *The analysis of Household surveys: a microeconometric approach to development Policy*, John Hopkins University Press.

Deaton, A. and C. Paxson (1998): "Economies of scale, household size, and the demand for food," *Journal of political economy*, 106, 897–930.

DeSoto, H. (1989): *The other path*, Harper & Row New York.

Diamond, P. A. (1975): "A many-person Ramsey tax rule," *Journal of Public Economics*, 4, 335–342.

Donovan, K., J. Lu, T. Schoellman, et al. (2018): "Labor Market Flows and Development," in *2018 Meeting Papers*, Society for Economic Dynamics, vol. 976.

Enste, D. H. and F. Schneider (2000): "Shadow Economies: Size, Causes, and Consequences," *Journal of Economic Literature*, 38, 77–114.

Faber, B. and T. Fally (2017): "Firm heterogeneity in consumption baskets: Evidence from home and store scanner data," .

Feldman, N. E. and J. Slemrod (2007): "Estimating tax noncompliance with evidence from unaudited tax returns," *The Economic Journal*, 117, 327–352.

Gadenne, L. (2019): "Can Rationing Increase Welfare? Theory and an Application to India's Ration Shop System," Tech. rep., CEPR Working Paper 13080.

Gadenne, L., T. Nandi, and R. Rathelot (2019): "Taxation and Supplier Networks: Evidence from India," Tech. rep., Mimeo, University of Warwick.

Gerard, F. and G. Gonzaga (2016): "Informal labor and the efficiency cost of social programs: Evidence from the brazilian unemployment insurance program," Tech. rep., National Bureau of Economic Research.

Gordon, R. and W. Li (2009): "Tax structures in developing countries: Many puzzles and a possible explanation," *Journal of Public Economics*, 93, 855 – 866.

Hanna, R. and B. Olken (2018): "Universal Basic Incomes versus Targeted Transfers: Anti-Poverty Programs in Developing Countries," *Journal of Economic Perspectives*, 32, 201–226.

Harris, T., D. Phillips, R. Warwick, M. Goldman, J. Jellema, K. Goraus, and G. Inchauste (2018): "Redistribution via VAT and cash transfers: an assessment in four low and middle income countries," Tech. rep., Institute for Fiscal Studies Working Paper W18/11.

Hsieh, C.-T. and P. J. Klenow (2014): "The Life Cycle of Plants in India and Mexico," *The Quarterly Journal of Economics*, 129, 1035–1084.

Hsieh, C.-T. and B. A. Olken (2014): "The Missing "Missing Middle"," *Journal of Economic Perspectives*, 28, 89–108.

Huang, J. and J. Rios (2016): "Optimal tax mix with income tax non-compliance," *Journal of Public Economics*, 144, 52–63.

Jenkins, G. P., H. P. Jenkins, and C. Y. Kuo (2006): "Is the value added tax naturally progressive?" *Available at SSRN 897677*.

Jensen, A. (2019): "Employment Structure and the Rise of the Modern Tax system," Tech. rep., National Bureau of Economic Research.

Keen, M. (2008): "VAT, tariffs, and withholding: Border taxes and informality in developing countries," *Journal of Public Economics*, 92, 1892–1906.

Keen, M. and J. Mintz (2004): "The optimal threshold for a value-added tax," *Journal of Public Economics*, 88, 559 – 576.

Kleven, H. J., C. T. Kreiner, and E. Saez (2016): "Why can modern governments tax so much? an agency model of firms as fiscal intermediaries," *Economica*, 51.

Kumler, T., E. Verhoogen, and J. A. Frías (2015): "Enlisting Employees in Improving Payroll-Tax Compliance: Evidence from Mexico," NBER Working Papers 19385.

La Porta, R. and A. Shleifer (2014): "Informality and development," *Journal of Economic Perspectives*, 28, 109–26.

Lagakos, D. (2016): "Explaining cross-country productivity differences in retail trade," *Journal of Political Economy*, 124, 579–620.

Lustig, N. (2018): *Commitment to Equity Handbook: Estimating the Impact of Fiscal Policy on Inequality and Poverty*, Brookings Institution Press.

Morrow, P., M. Smart, and A. Swistak (2019): "VAT Compliance, Trade, and Institutions," CESifo Working Paper Series 7780, CESifo Group Munich.

Muñoz, S. and S. S.-W. Cho (2003): "Social Impact of a Tax Reform; The Case of Ethiopia," IMF Working Papers 03/232, International Monetary Fund.

Naritomi, J. (2018): "Consumers as tax auditors," *CEPR Discussion Paper*.

Pissarides, C. A. and G. Weber (1989): "An expenditure-based estimate of Britain's black economy," *Journal of Public Economics*, 39, 17 – 32.

Pritchett, L. and M. Spivack (2013): "Estimating Income/Expenditure Differences Across Populations: New Fun with Old Engel's Law," *Center for Global Development Working Paper*.

Ray, R. (1986): "Sensitivity of 'optimal'commodity tax rates to alternative demand functional forms: An econometric case study of India," *Journal of Public Economics*, 31, 253–268.

Saez, E. and S. Santcheva (2016): "Generalized Social Marginal Welfare Weights for Optimal Tax Theory," *American Economic Review*, 106, 24–45.

Sah, R. K. (1983a): "How much redistribution is possible through commodity taxes?" *Journal of Public Economics*, 20, 89–101.

——— (1983b): "How much redistribution is possible through commodity taxes?" *Journal of Public Economics*, 20, 89–101.

Srinivasan, P. (1989): "Redistributive impact of 'optimal'commodity taxes: Evidence from Indian data," *Economics Letters*, 30, 385–388.

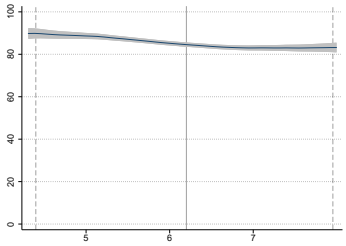
Ulyssea, G. (2018): "Firms, informality, and development: Theory and evidence from Brazil," *American Economic Review*, 108, 2015–47.

Warren, N. (2008): "A Review of Studies on the Distributional Impact of Consumption Taxes in OECD Countries," Tech. rep., OECD Social, Employment and Migration Working Papers No. 64.

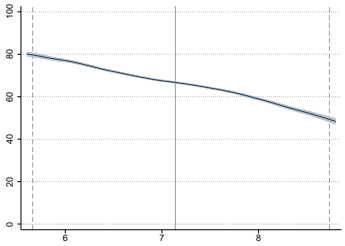
Working, H. (1943): "Statistical laws of family expenditure," *Journal of the American Statistical Association*, 38, 43–56.

A Additional Figures and Tables

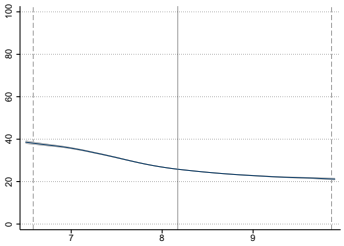
Figure A.1: Informality Engel Curves



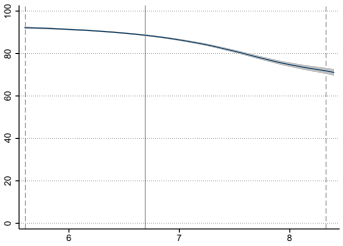
(a) Benin



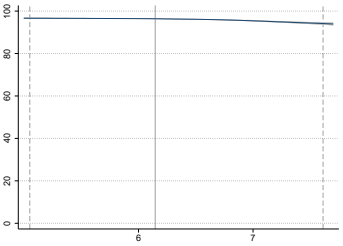
(b) Bolivia



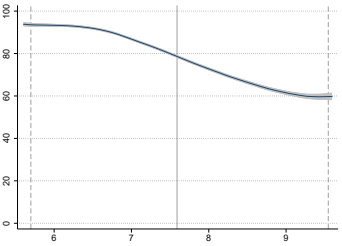
(c) Brazil



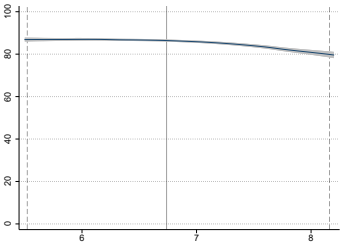
(d) Burkina Faso



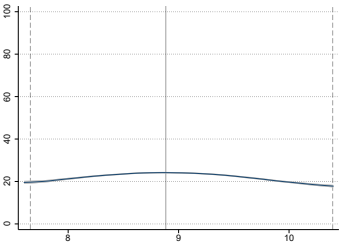
(e) Burundi



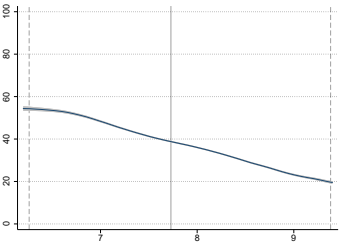
(f) Cameroon



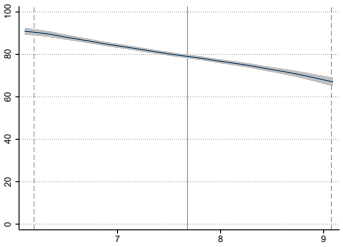
(g) Chad



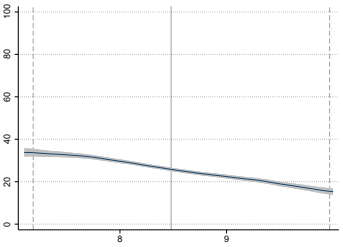
(h) Chile



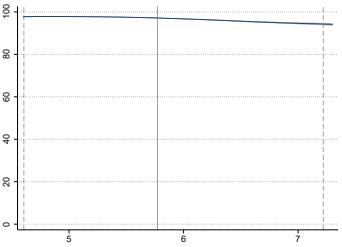
(i) Columbia



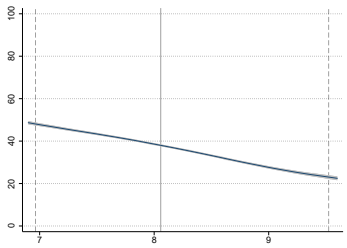
(j) Comoros



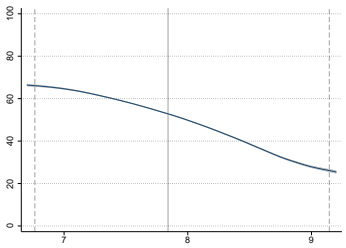
(k) Costa Rica



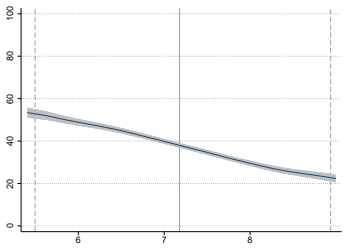
(l) Dem. Rep. of Congo



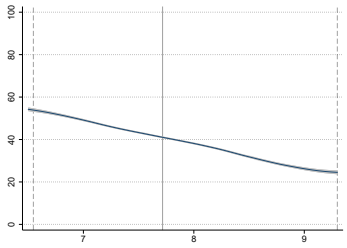
(m) Dominican Rep.



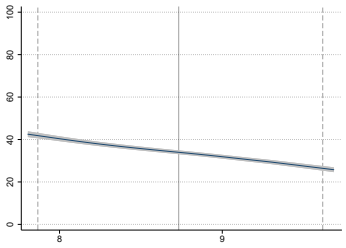
(n) Ecuador



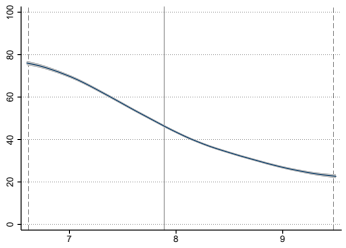
(o) Eswatini



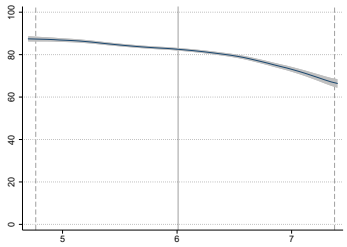
(p) Mexico



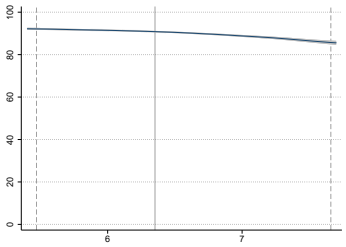
(q) Montenegro



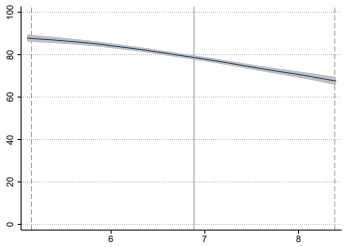
(r) Morocco



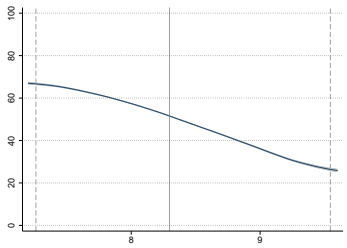
(s) Mozambique



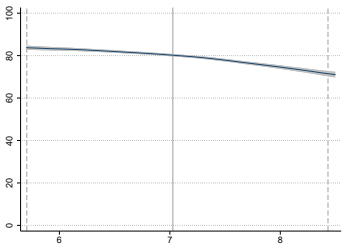
(t) Niger



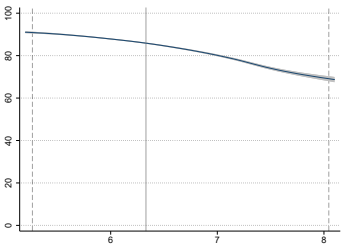
(u) Papua New Guinea



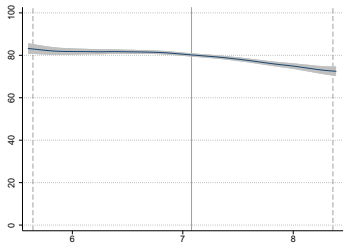
(v) Peru



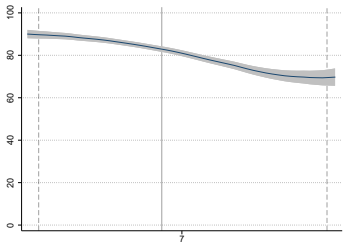
(w) Rep. of Congo



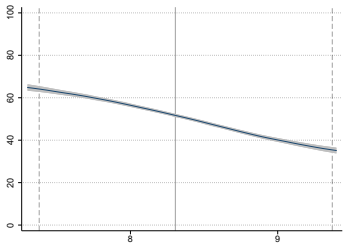
(x) Rwanda



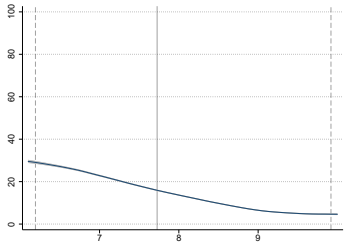
(y) Sao Tome



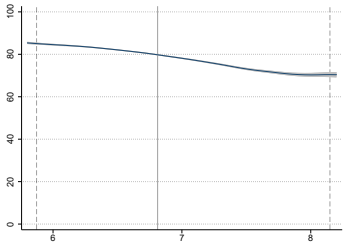
(z) Senegal



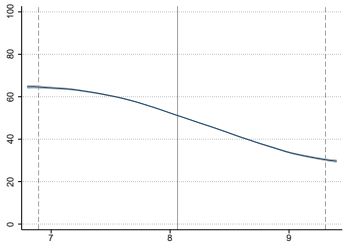
(aa) Serbia



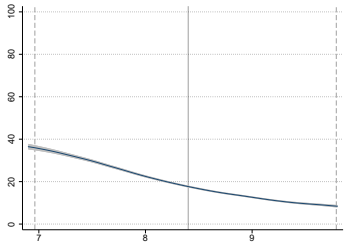
(ab) South Africa



(ac) Tanzania



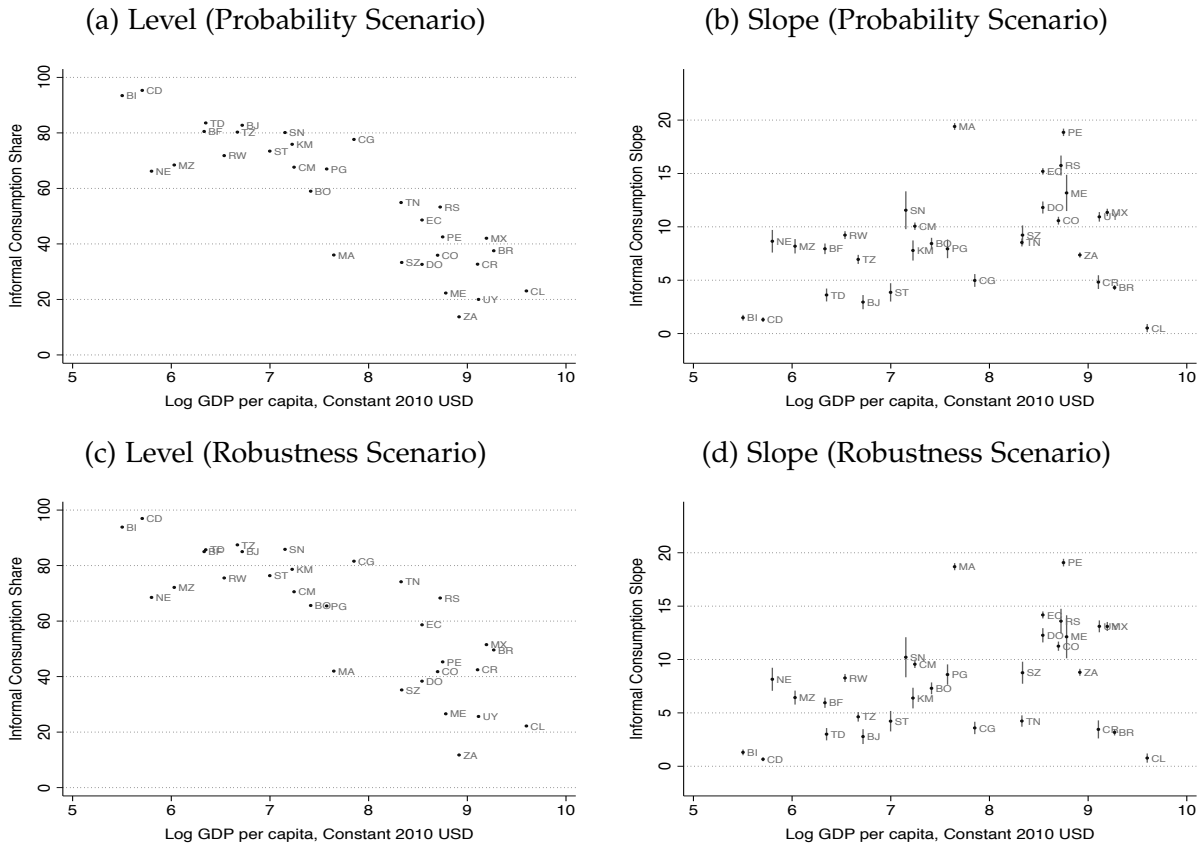
(ad) Tunisia



(ae) Uruguay

Local polynomial fit of the informality Engel curves for DR Congo and Mexico as presented in section 3.1. The informality Engel curves are defined as the share of informal consumption over the log per person expenditure of households. The shaded area around the polynomial fit corresponds to the 95% confidence interval. The solid grey line corresponds to the median of each country's expenditure distribution, while the dotted lines correspond to the 5th and 95th percentiles.

Figure A.2: Alternative Assignment Scenarios for Informal Consumption



These panel show the equivalent of figure 3 for two alternative choices of assignments of store types to formality. The Probability scenario uses the observed probability of VAT registration by store type in the Mexican census as the formality probability of each store across countries (see Figure 1). The robustness scenario differs from the central scenario by assigning specialized stores to the informal sector, in addition to maintaining corner stores, non brick and mortar and self-production in the informal sector. Panel (a) and (c) show informal consumption as a share of total consumption, on log per capita GDP. Panel (b) and (d) show the (negative) of the slope of informal consumption with respect to total household expenditure, on per capita GDP. The bars correspond to the 95% confidence interval of the slope coefficient. GDP per capita is in constant 2010 USD, and transformed in log base 2 (Source: World Bank WDI)

Figure A.3: Average Share of Unspecified Category by COICOP

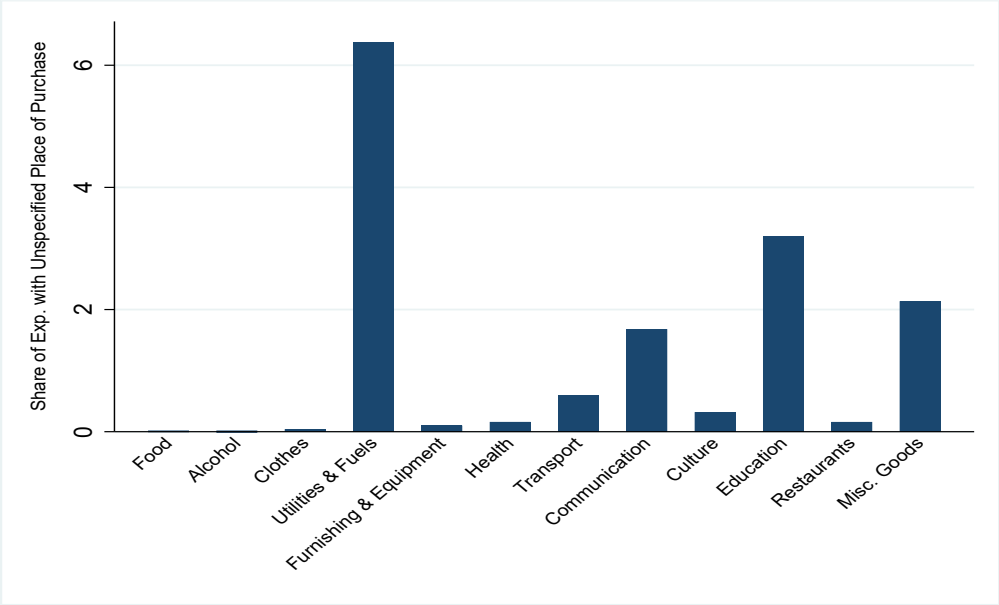
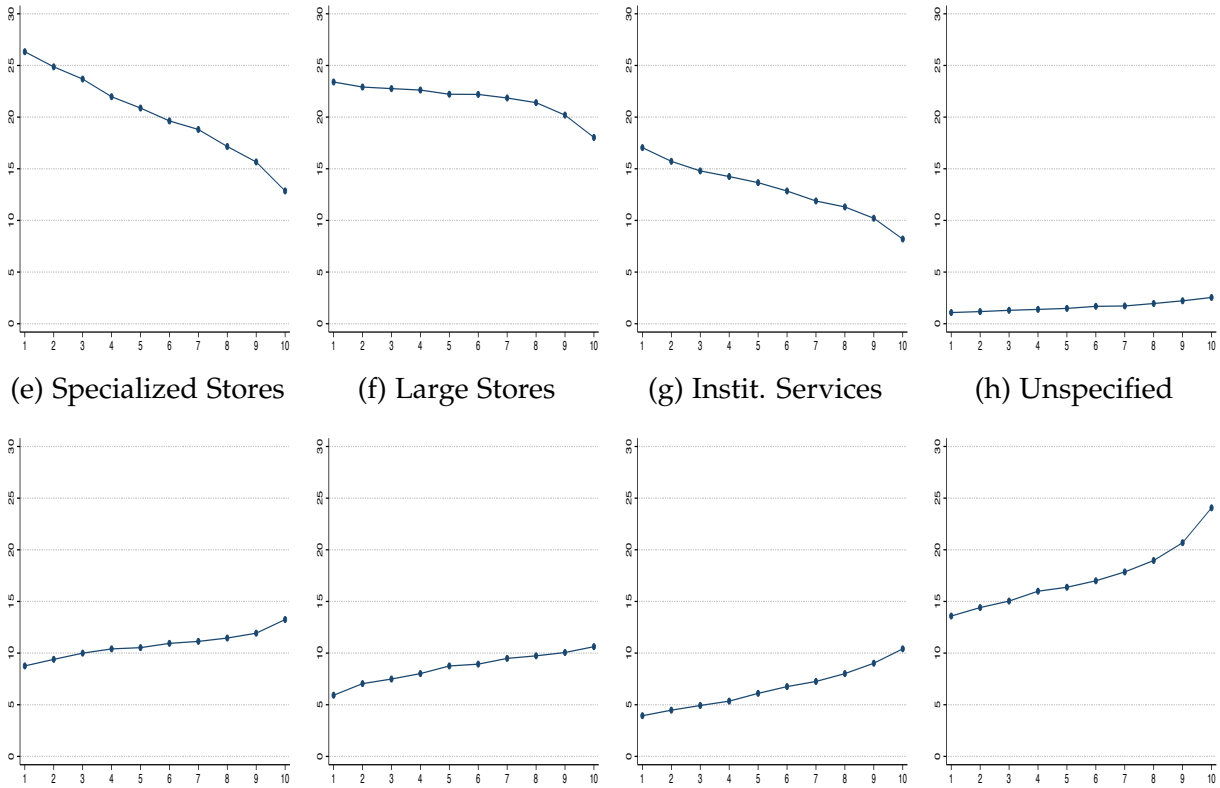


Figure A.4: Average Consumption of Each Decile by Place of Purchase

(a) Non Market (b) Non Brick&Mortar (c) Corner Stores (d) Indiv. Providers

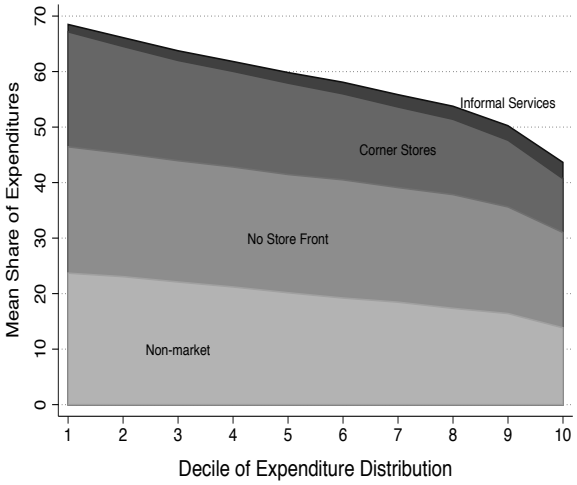


(e) Specialized Stores (f) Large Stores (g) Instit. Services (h) Unspecified

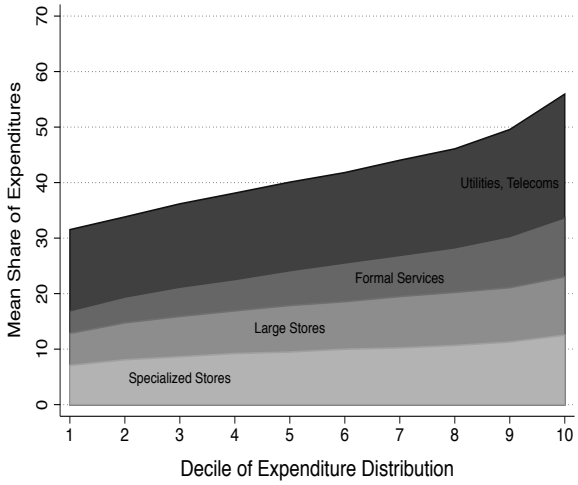
xx

Figure A.5: Average Consumption of Each Decile by Place of Purchase

(a) Informal Places of Purchase



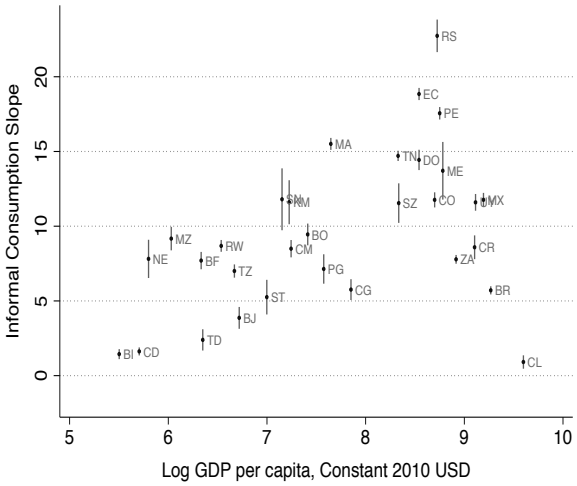
(b) Formal Places of Purchase



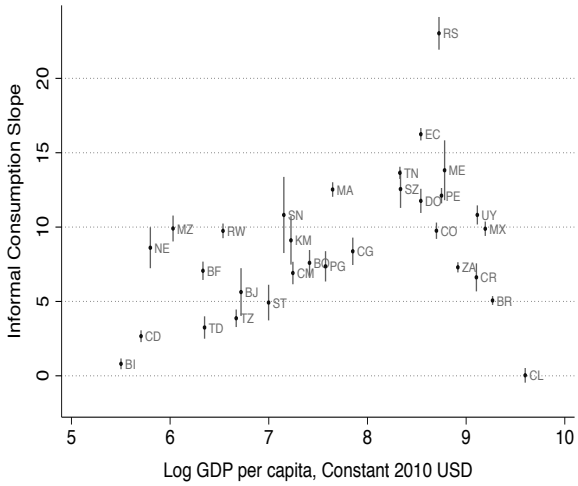
This figure shows the average consumption of each decile across countries by type of retailer, following the retailer taxonomy described in section 2.2. Panel (a) shows the places of purchase classified as informal and Panel (b) shows the places of purchase classified as formal in the central scenario of the paper.

Figure A.6: IEC Slopes Controlling for Geography

(a) Control: Rural Location



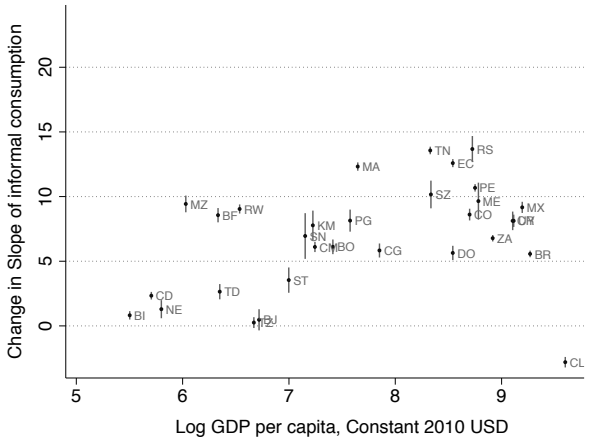
(b) Control: Survey Block



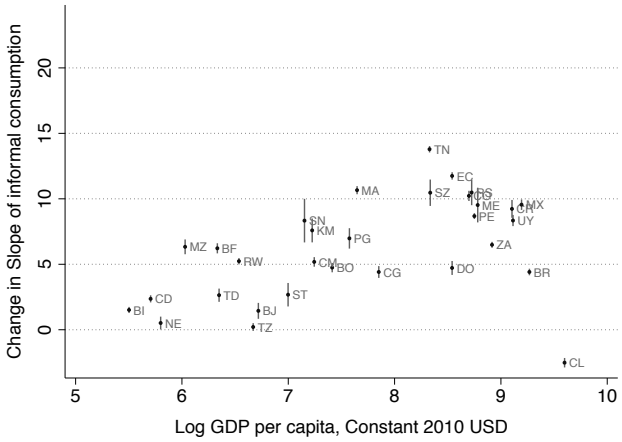
This figure shows the Informality Engel curves' slopes across countries when controlling for rural vs urban in panel (a) and for survey enumeration blocks in panel (b).

Figure A.7: IEC Slopes Controlling for Product Composition

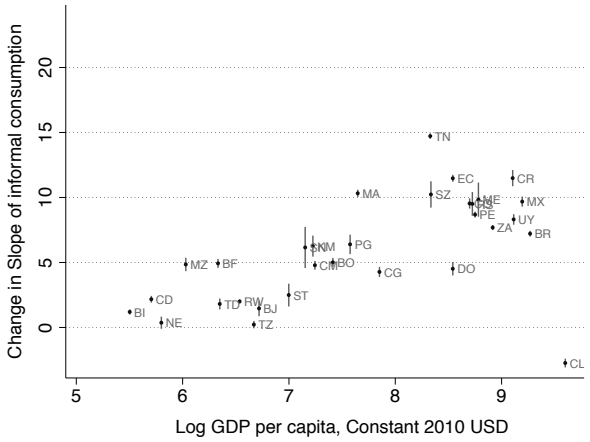
(a) Control: Food Products



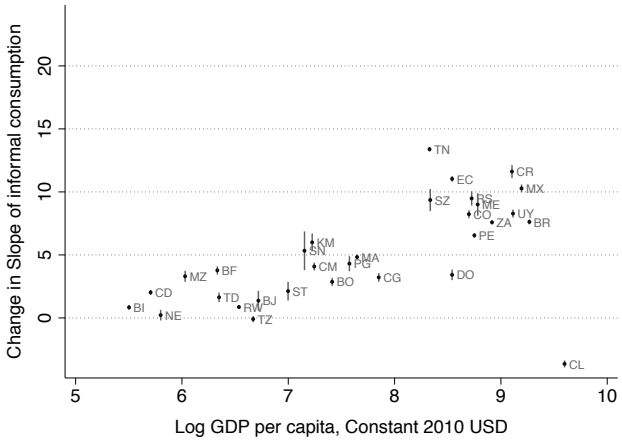
(b) Control: COICOP2



(c) Control: COICOP3



(d) Control: COICOP4

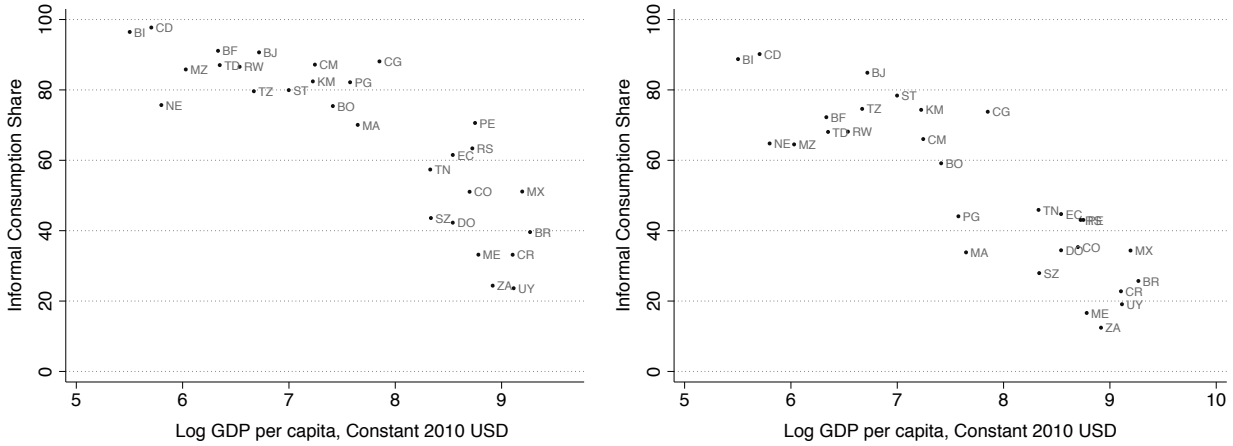


This figure shows the Informality Engel curves' slopes across countries when controlling for increasingly narrow products (and household controls). Panel (a) only controls for food products, panel (b) controls for the 12 COICOP2 good categories, panel (c) controls for the 53 COICOP3 categories and panel (d) controls for the 130 COICOP4 categories.

Figure A.8: Rural vs Urban Informal Consumption

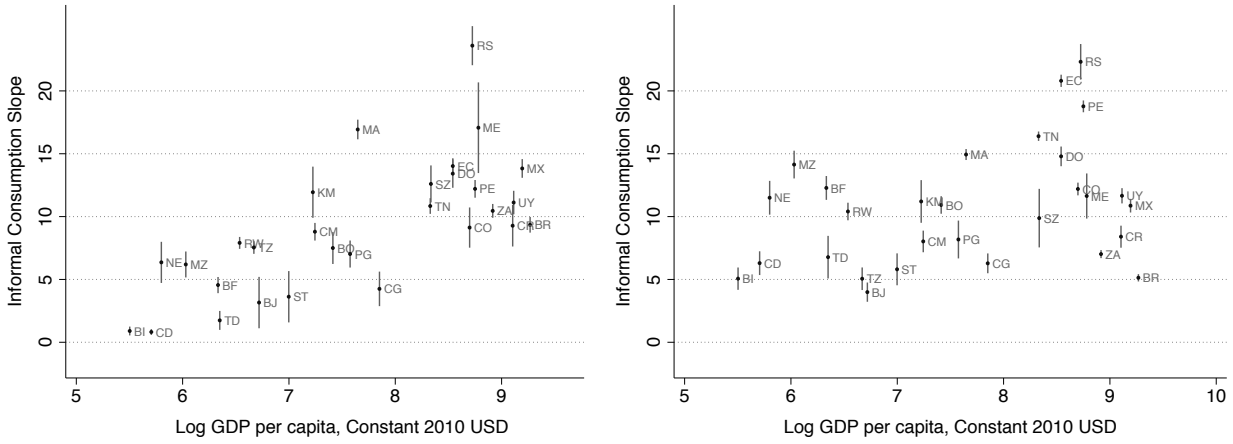
(a) Rural: Informality Level

(b) Urban: Informality Level



(c) Rural: IEC slope

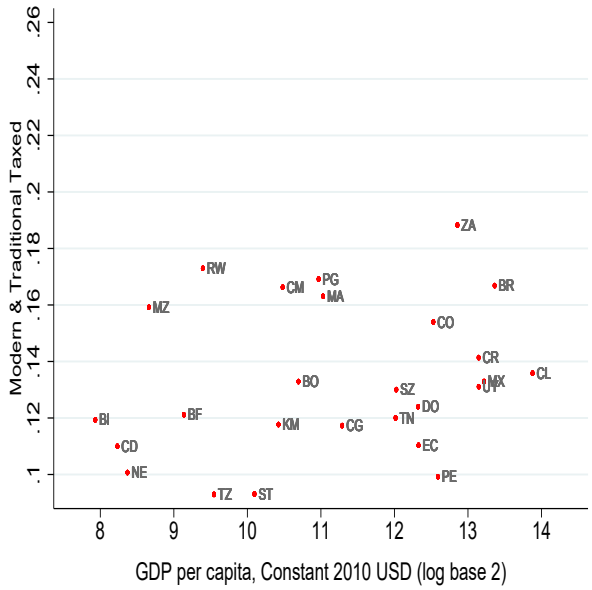
(d) Urban: IEC slope



This figure shows separately the informality levels and IEC slopes for rural and urban population within each country. We note that this figure only contains 29 countries instead of 31, since the expenditure surveys in Chile and Senegal concern urban population only.

Figure A.9: Optimal uniform rates over development with inequalities

(a) Modern & Traditional Taxed



(b) Only Modern Taxed

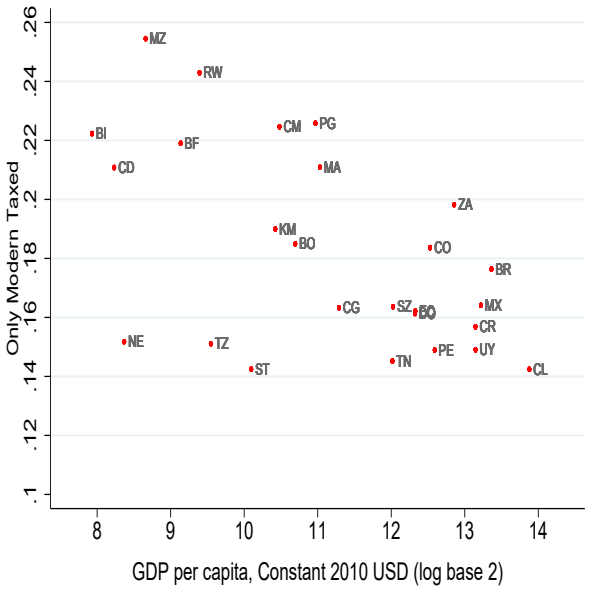
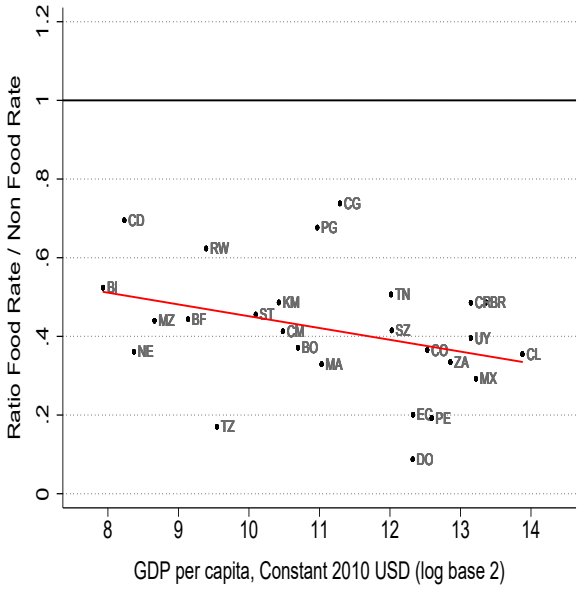


Figure A.10: Optimal Rates on Food and Non-Food, with inequalities

(a) Modern & Traditional Taxed



(b) Only Modern Taxed

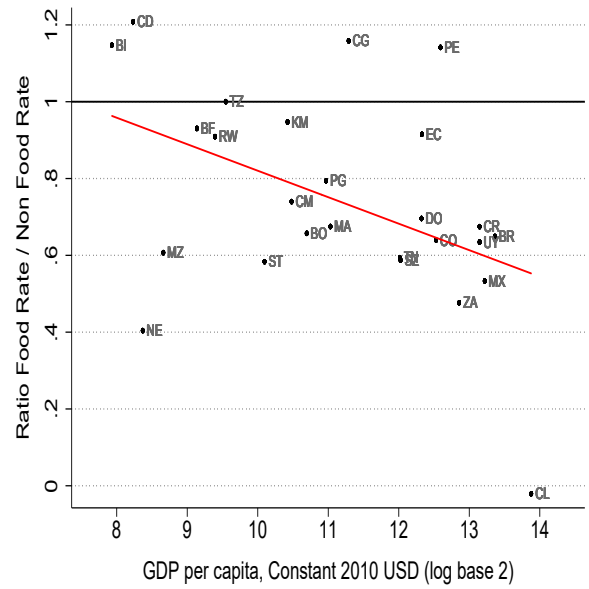


Figure A.11: Consumption Tax Progressivity with Savings

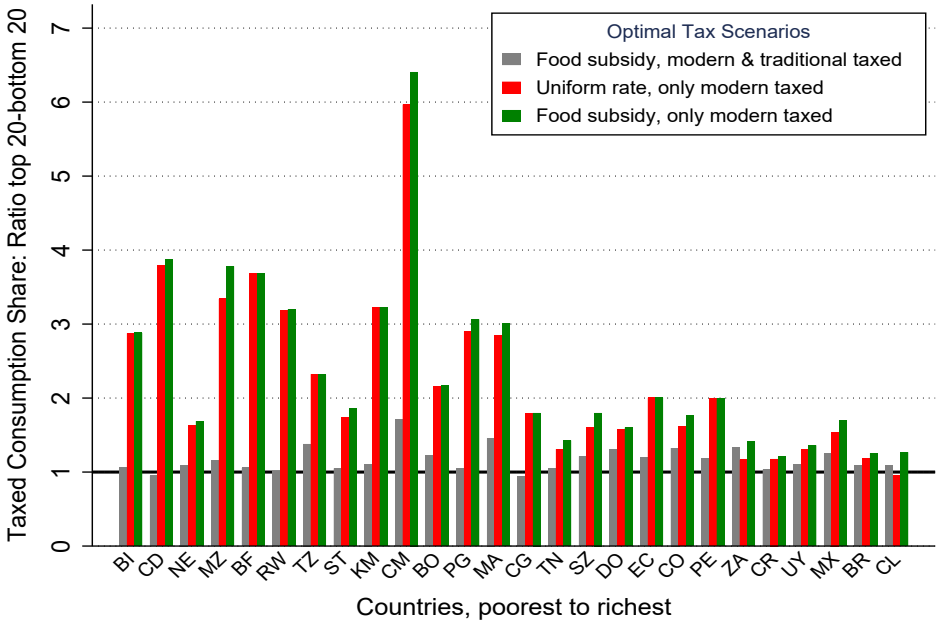


Figure A.12: Consumption Tax Prog., 10% Informal Sector Pass-through

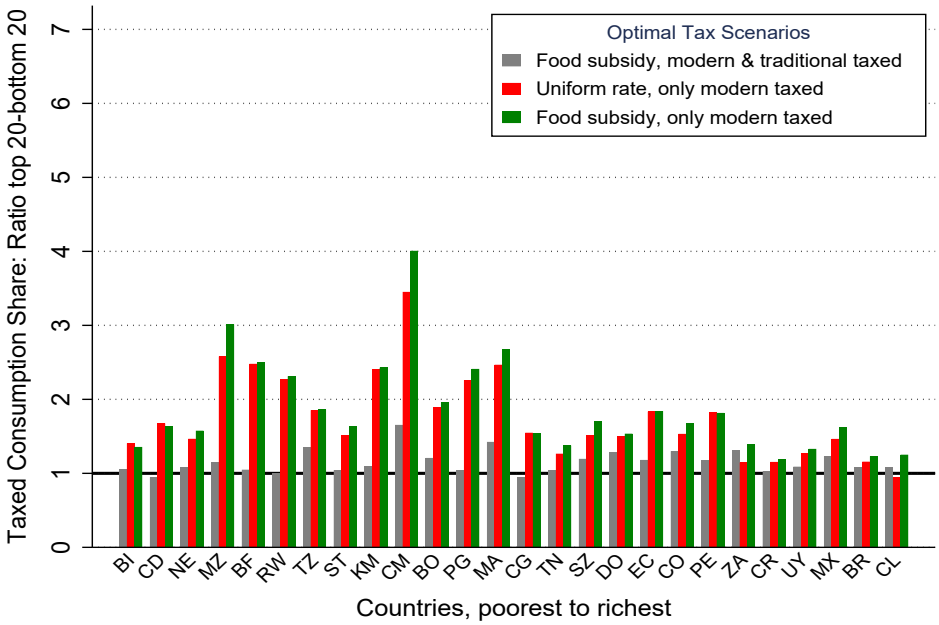


Figure A.13: Consumption Tax Progressivity, Robust Scenario

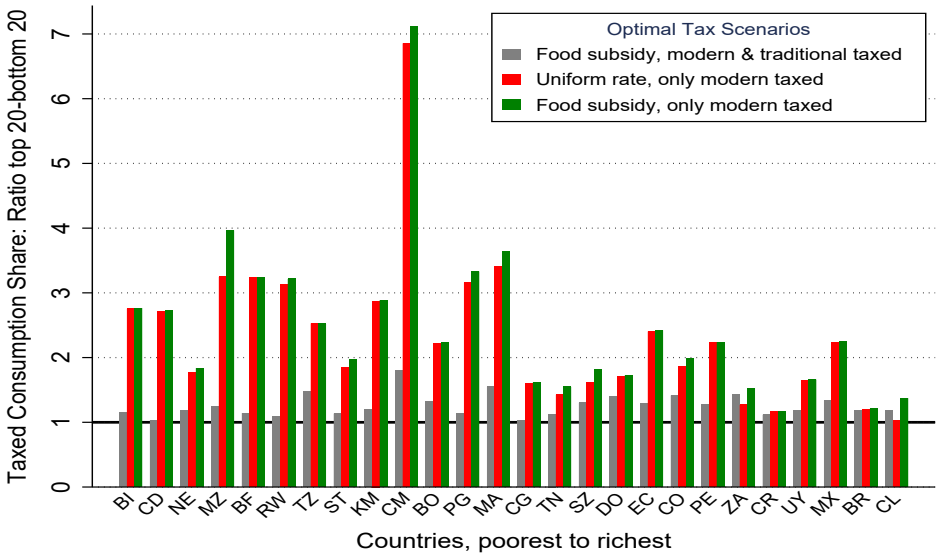


Figure A.14: Consumption Tax Progressivity, Probabilistic Scenario

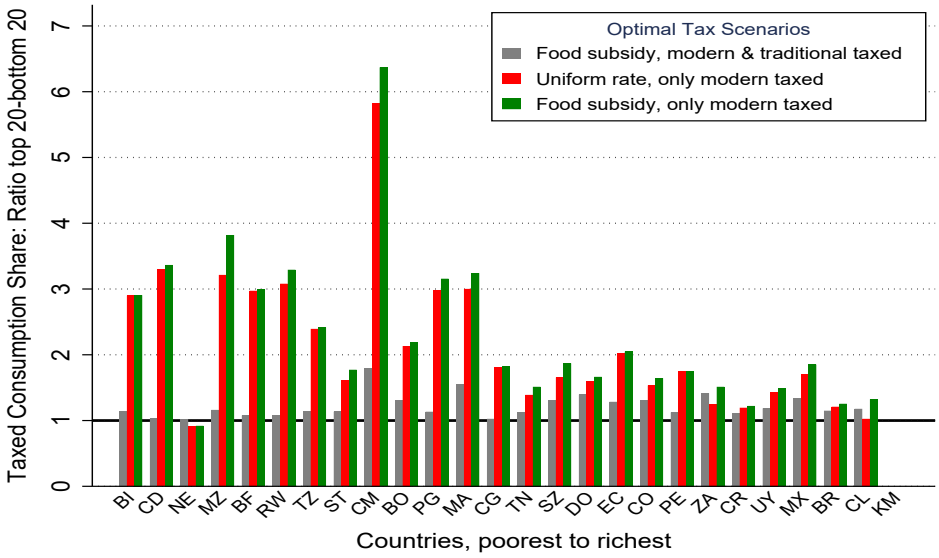


Table A1: IEC Slopes by Country

Country	Main		Geography			Product Codes			All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Benin	3.40	3.91	0.47	1.44	1.46	1.37	2.24	3.87	5.63
	(0.37)	(0.38)	(0.42)	(0.31)	(0.30)	(0.40)	(0.59)	(0.37)	(0.82)
Bolivia	10.28	11.99	6.11	4.74	5.01	2.86	2.68	9.45	7.59
	(0.29)	(0.33)	(0.29)	(0.18)	(0.18)	(0.16)	(0.25)	(0.37)	(0.44)
Brazil	6.19	6.64	5.57	4.41	7.21	7.62	6.23	5.71	5.06
	(0.13)	(0.14)	(0.13)	(0.13)	(0.11)	(0.10)	(0.11)	(0.14)	(0.15)
Burkina Faso	9.84	10.61	8.56	6.22	4.93	3.77	2.45	7.70	7.06
	(0.30)	(0.32)	(0.28)	(0.20)	(0.17)	(0.17)	(0.19)	(0.30)	(0.31)
Burundi	1.89	2.34	0.82	1.51	1.19	0.84	0.33	1.44	0.80
	(0.16)	(0.17)	(0.17)	(0.12)	(0.10)	(0.10)	(0.12)	(0.17)	(0.18)
Cameroon	10.58	11.16	6.10	5.18	4.78	4.08	2.54	8.50	6.91
	(0.19)	(0.21)	(0.20)	(0.19)	(0.17)	(0.16)	(0.25)	(0.30)	(0.39)
Chad	4.23	4.61	2.65	2.64	1.81	1.63	0.96	2.39	3.25
	(0.34)	(0.37)	(0.30)	(0.26)	(0.21)	(0.19)	(0.24)	(0.36)	(0.38)
Chile	0.97	0.91	-2.81	-2.52	-2.74	-3.65	-3.24	0.91	0.03
	(0.21)	(0.23)	(0.21)	(0.18)	(0.17)	(0.14)	(0.16)	(0.23)	(0.25)
Colombia	12.38	12.72	8.61	10.22	9.54	8.23	7.01	11.76	9.76
	(0.22)	(0.25)	(0.23)	(0.20)	(0.20)	(0.16)	(0.18)	(0.26)	(0.28)
Comoros	10.00	12.18	7.77	7.58	6.26	6.00	4.45	11.61	9.11
	(0.58)	(0.71)	(0.58)	(0.47)	(0.41)	(0.36)	(0.56)	(0.75)	(0.83)
Congo, Dem. Rep.	2.04	2.76	2.34	2.36	2.17	2.02	1.53	1.62	2.66
	(0.15)	(0.17)	(0.15)	(0.14)	(0.13)	(0.10)	(0.16)	(0.14)	(0.20)
Congo, Rep.	6.38	7.50	5.83	4.41	4.27	3.21	3.10	5.76	8.37
	(0.34)	(0.36)	(0.28)	(0.23)	(0.20)	(0.18)	(0.27)	(0.36)	(0.47)
Costa Rica	7.88	9.63	8.12	9.23	11.49	11.61	9.47	8.59	6.62
	(0.36)	(0.39)	(0.37)	(0.35)	(0.32)	(0.27)	(0.33)	(0.41)	(0.48)
Dominican Republic	14.34	14.85	5.64	4.72	4.51	3.42	1.96	14.44	11.76
	(0.30)	(0.35)	(0.28)	(0.27)	(0.26)	(0.22)	(0.24)	(0.35)	(0.41)
Ecuador	18.94	20.60	12.59	11.75	11.47	11.03	8.55	18.85	16.24
	(0.17)	(0.19)	(0.16)	(0.14)	(0.14)	(0.11)	(0.13)	(0.21)	(0.21)
Eswatini	11.64	12.37	10.16	10.46	10.23	9.36	9.43	11.55	12.56
	(0.51)	(0.62)	(0.55)	(0.51)	(0.52)	(0.45)	(0.47)	(0.67)	(0.65)
Mexico	12.39	13.95	9.17	9.55	9.68	10.28	6.82	11.77	9.89
	(0.20)	(0.23)	(0.22)	(0.20)	(0.20)	(0.16)	(0.18)	(0.24)	(0.25)
Montenegro	15.56	17.09	9.65	9.52	9.83	9.00	7.35	13.71	13.82
	(0.85)	(0.94)	(0.73)	(0.68)	(0.67)	(0.45)	(0.50)	(0.98)	(1.03)
Morocco	20.93	22.32	12.32	10.66	10.32	4.83	1.55	15.51	12.53
	(0.17)	(0.18)	(0.16)	(0.16)	(0.14)	(0.10)	(0.14)	(0.20)	(0.25)
Mozambique	9.93	11.57	9.43	6.33	4.84	3.30	2.41	9.17	9.91
	(0.38)	(0.40)	(0.33)	(0.29)	(0.26)	(0.22)	(0.27)	(0.40)	(0.44)
Niger	8.18	9.40	1.29	0.51	0.37	0.22	0.58	7.81	8.61
	(0.55)	(0.56)	(0.36)	(0.25)	(0.24)	(0.22)	(0.24)	(0.65)	(0.70)
Papua New Guinea	8.59	9.34	8.14	6.98	6.39	4.30	3.13	7.14	7.36
	(0.49)	(0.49)	(0.43)	(0.40)	(0.38)	(0.30)	(0.32)	(0.50)	(0.52)
Peru	20.10	22.38	10.67	8.68	8.68	6.55	4.23	17.56	12.11
	(0.18)	(0.19)	(0.15)	(0.11)	(0.11)	(0.09)	(0.11)	(0.21)	(0.26)
Rwanda	9.90	10.61	9.04	5.23	2.00	0.86	0.07	8.68	9.75
	(0.19)	(0.20)	(0.18)	(0.12)	(0.08)	(0.07)	(0.08)	(0.20)	(0.25)
Sao Tome and Principe	4.67	5.22	3.54	2.67	2.49	2.12	2.13	5.25	4.92
	(0.50)	(0.59)	(0.50)	(0.46)	(0.45)	(0.38)	(0.40)	(0.59)	(0.61)
Senegal	14.72	11.80	6.95	8.33	6.15	5.33	5.10	11.80	10.82
	(0.98)	(1.06)	(0.90)	(0.85)	(0.81)	(0.78)	(1.00)	(1.06)	(1.31)
Serbia	20.91	24.24	13.67	10.48	9.50	9.48	8.47	22.74	23.03
	(0.58)	(0.58)	(0.51)	(0.49)	(0.47)	(0.29)	(0.29)	(0.56)	(0.56)
South Africa	7.50	8.69	6.78	6.49	7.69	7.59	6.92	7.78	7.29
	(0.12)	(0.14)	(0.13)	(0.12)	(0.10)	(0.08)	(0.11)	(0.15)	(0.18)
Tanzania	8.31	7.30	0.26	0.21	0.22	-0.10	-0.40	7.00	3.87
	(0.22)	(0.22)	(0.21)	(0.15)	(0.14)	(0.13)	(0.17)	(0.23)	(0.30)
Tunisia	15.42	15.48	13.56	13.79	14.72	13.39	11.78	14.71	13.65
	(0.15)	(0.17)	(0.15)	(0.13)	(0.11)	(0.09)	(0.12)	(0.17)	(0.21)
Uruguay	11.52	11.67	8.13	8.34	8.31	8.28	7.38	11.60	10.82
	(0.25)	(0.28)	(0.24)	(0.22)	(0.21)	(0.16)	(0.17)	(0.29)	(0.33)
All Countries (Mean)	10.3	11.2	9.6	8.8	6.9	6.3	6.1	5.2	4.2
Household Characteristics		X	X	X	X	X	X	X	X
Urban/Rural			X						
Survey Blocks				X					X
Food Products					X			X	
COICOP 2-dig						X			
COICOP 3-dig							X		
COICOP 4-dig								X	X

This table shows the average slope of the Informal Engel curve across countries for different specifications. The slopes are estimated from: $Share\ Informal_i = \beta.ln(expenditure\ pc)_i + \Gamma X_i + \varepsilon_i$, where the dependent variable is the informal expenditure share and the explanatory variable is the log expenditure pp. Controls include household characteristics (household size, age, gender, and education of head), geographic indicators (urban/rural and survey enumeration blocks), and product codes for food compared to the rest and at the 2nd, 3rd and 4th level of the United Nation's COICOP classification.

Table A2: IEC Slopes: Probabilistic Formality Assignment

Specification:	**Main**		**Geography**		**Product Codes**				**All**
Avg. of 31 Countries	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Slope	8.9	9.5	8.0	7.2	5.5	4.9	4.8	4.1	3.3
Confidence Interval	[8.3,9.5]	[8.8,10.1]	[7.3,8.7]	[6.4,8.0]	[4.9,6.0]	[4.4,5.3]	[4.3,5.2]	[3.6,4.4]	[2.7,3.6]
# of p-values < 0.05	31	31	31	30	29	30	30	30	28
R^2 adjusted	0.18	0.20	0.25	0.41	0.46	0.56	0.60	0.60	0.63
Household Characteristics		X	X	X	X	X	X	X	X
Urban/Rural			X						
Survey Blocks				X					X
Food Products					X				
COICOP 2-dig						X			
COICOP 3-dig							X		
COICOP 4-dig								X	X

This table shows the average slope of the Informal Engel curve across countries for different specifications under the probabilsitic assignment of palces of purchase to formality. The Probabilistic scenario uses the observed probability of VAT registration by store type in the Mexican census as the formality probability of each store across countries (see Figure 1). The slopes are estimated from: $Share\ Informal_i = \beta.ln(expenditure\ pc)_i + \Gamma X_i + \varepsilon_i$. The dependent variable is informal expenditure share and the main explanatory variable is log expenditure per capita. Controls include household characteristics (household size, age, gender, education of head), location indicators (urban/rural, survey enumeration blocks), and product codes for food vs all other purchases, 2nd, 3rd and 4th level of the COICOP classification.

Table A3: IEC Slopes: Robustness Formality Assignment

Specification:	**Main**		**Geography**		**Product Codes**				**All**
Avg. of 31 Countries	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Slope	8.5	9.0	7.6	6.9	5.0	4.4	4.5	3.9	3.3
Confidence Interval	[7.9,9.2]	[8.3,9.7]	[6.9,8.3]	[6.0,7.7]	[4.3,5.6]	[3.8,4.9]	[4.0,4.9]	[3.4,4.2]	[2.6,3.6]
# of p-values < 0.05	31	31	31	30	28	30	30	26	26
R^2 adjusted	0.16	0.18	0.22	0.38	0.43	0.54	0.59	0.60	0.63
Household Characteristics		X	X	X	X	X	X	X	X
Urban/Rural			X						
Survey Blocks				X					X
Food Products					X				
COICOP 2-dig						X			
COICOP 3-dig							X		
COICOP 4-dig								X	X

This table shows the average slope of the Informal Engel curve across countries for different specifications under the robustness scenario assignment of places of purchase to formality. The robustness scenario differs from the central scenario by assigning specialized stores to the informal sector, in addition to maintaining corner stores, non brick and mortar and self-production in the informal sector. The slopes are estimated from: $Share\ Informal_i = \beta.ln(expenditure\ pc)_i + \Gamma X_i + \varepsilon_i$. The dependent variable is informal expenditure share and the main explanatory variable is log expenditure per capita. Controls include household characteristics (household size, age, gender, education of head), location indicators (urban/rural, survey enumeration blocks), and product codes for food vs all other purchases, 2nd, 3rd and 4th level of the COICOP classification.

Table A4: Main Reason for Choosing Place of Purchase

	Morocco			**Dem. Rep. Congo**			**Rep. of Congo**			**Burundi**		
Reason	Total	Formal	Informal	Total	Formal	Informal	Total	Formal	Informal	Total	Formal	Informal
Access	56%	54%	56%	27%	24%	28%	35%	41%	34%	50%	41%	50%
Price	24%	7.3%	27%	35%	24%	36%	32%	20%	36%	28%	22%	29%
Quality	7.8%	18%	6.2%	19%	43%	18%	16%	28%	14%	8.5%	32%	6.9%
PoP Attributes	7.8%	9.6%	7.5%	10%	4.6%	10%	8.8%	5.7%	9.7%	4.2%	1.0%	4.4%
Other	4.6%	11%	3.7%	8.8%	4.2%	7.9%	8.3%	5.7%	7.5%	9.4%	4.9%	9.7%

'Access" is defined as a combination of proximity and necessity and "Attributes of PoP" is defined as a combination of homogeneity of products, offering of credit, and quality of reception.

B Data Appendix

B.1 Core Sample

Inclusion Criteria

Our dataset consists of 31 nationally representative household expenditure surveys. We use surveys which satisfy the following four criteria:

1. The household expenditure survey is nationally representative and dates from the 21st century.

2. The expenditure module(s) in the survey is structured as an open consumption diary, rather than a pre-fill diaries for a limited set of products.

3. The expenditure survey includes a variable for the place of purchase: there is information on where each consumption item was purchased. The place of purchase choices are detailed enough for us to apply our taxonomy of store types, as further outlined in section B.2.

4. The place of purchase variable rarely contains missing values, particularly for food, clothing and household goods product categories. Particularly, the place of purchase is never unspecified for more than a third of the observations and represent less than 7% of total expenditures in each COICOP categories as shown in Figure A.3.

Data Sources and Coverage

We obtained the data principally from two sources: (i) the World Bank Microdata Library and (ii) National Statistical Agencies. Our first step for accessing data was to search the restricted-access World Bank Microdata Library for household Income and Expenditure, Living Standards, and Budget Surveys to see in which countries criteria (1)-(4) above appeared to be satisfied. The datasets which satisfied these criteria varied in their ease of access: for some countries, the micro data were accessible for download on the World Bank platform, others were licensed and required applications through the World Bank, which would in turn sometimes contact the country's national statistical agency for approval. If a survey was listed without its micro through the World Bank platform, we reached directly the country's' statistical agency. Most countries for which we requested the data sent us the micro data, but in a few cases we could not obtain data which included the place of purchase variables. The countries which ultimately satisfied

the criteria for inclusion span four regions of the world, concentrated in Sub-Saharan Africa and Latin America and the Caribbean, as detailed in Table B1. Unfortunately we were not able to include countries in Asia, since the question on the place of purchase was almost always missing from their budget surveys.

Table B1: Regional Survey Representation

Region	**# Countries**	**Pop. of Surveyed Countries** (Millions)	**Total Pop.** (Millions)	**Proportion of pop.**
Sub-Saharan Africa	16	379	1078	35%
Middle East & North Africa	2	48	449	11%
Europe & Central Asia	2	9	918	1%
Latin America & Carribean	10	489	641	76%
East Asia & Pacific	1	9	2328	0.4%

While some surveys appeared from their questionnaire to satisfy our criteria, we ultimately could not include them, either because of issues with data access, or because when we looked more closely at the data one of our criteria was violated. For completeness, Table B2 further details countries that were considered, but could not be included/

Table B2: Discarded Household Expenditure Surveys

Country	**Survey**	**Year**	**Reason not Included**
Armenia	Integrated Living Conditions Survey	2016	PoP often missing
Bosnia & Herzegovina	Household Budget Survey	2007	PoP asked as a purchasing habit
El Salvador	Encuesta de Hogares de Propositos Multiples	2010	PoP often missing; limited product categories
The Gambia	Integrated Household Survey	2003	No Data access to PoP
Gerogia	Integrated Household Survey	2018	Limited product categories
Ghana	Living Standards Survey	2006	No Data access to PoP
Guatemala	Encuesta Nacional sobre Condiciones de Vida	2000	PoP often missing; limited product categories
Mauritius	Household Budget Survey	2012	PoP asked as a purchasing habit
Namibia	Namibia Household Income and Expenditure Survey	2015	limited product categories
Nicaragua	Encuesta Nacional de Hogares sobre Medicion de Nivel	2014	PoP asked as a purchasing habit; limited product categories
Tajikistan	Household Budget Survey	2016	limited product categories
Turkey	Household Income and Consumption Survey	2009	No Data access to PoP

Table B3 lists the 31 countries which we could include, with summary statistics and the structure of each survey. Any slight deviation from our inclusion criteria is outlined in the last column.

Table B3: Household Expenditure Surveys

Country name	Survey	Year	Source	# HH	# items/HH	Exp/HH	Urban	HH Size	# PoP	# Modules	Product Code	Comments
						Cst. 2010 USD						
Benin	EMICOV	2015	World Bank	19902	31.8	321	48.2%	3.9	12	22	COICOP	
Bolivia	ECH	2004	Stat. Office	9149	49.4	495	60.7%	4.2	24	3	COICOP	
Brazil	POF	2009	Stat. Office	56049	48	3892	84.4%	3.3	33	8	Country-specific	
Burkina Faso	EICVM	2009	Stat. Office	8404	161.6	563	29.3%	6.7	45	1	COICOP	
Burundi	ECVM	2014	World Bank	6681	90.2	242	9.0%	4.8	13	23	COICOP	
Cameroon	ECAM	2014	World Bank	10303	95.8	1889	44.5%	4.6	17	1	COICOP	
Chad	ECOSIT	2003	World Bank	6747	92	356	10.9%	5.9	17	18	Country-specific	
Chile	EPF	2017	Stat. Office	15237	129.2	6872	100.0%	3.3	22	1	COICOP	No self-production, Only urban
Colombia	ENIG	2007	Stat. Office	42733	79.6	1850	82.4%	3.8	24	5	COICOP	
Comoros	EDMC	2013	Stat. Office	3139	83.5	1809	49.1%	5	12	19	COICOP	
Congo DRC	E123	2005	World Bank	12098	106.9	198	16.0%	5.3	13	1	COICOP	
Congo Rep	ECOM	2005	World Bank	5002	84.8	641	63.8%	5.1	17	1	COICOP	
Costa Rica	ENIGH	2014	Stat. Office	5705	67.5	5256	73.2%	3.4	41	1	COICOP	
Dominican Rep	ENIGH	2007	Stat. Office	8363	89.1	2396	67.6%	3.7	88	3	COICOP	
Ecuador	ENIGHUR	2012	World Bank	39617	88.6	1923	68.0%	3.9	75	7	COICOP	
Eswatini	HIES	2010	World Bank	3167	43.9	1283	37.4%	4.5	13	2	COICOP	
Mexico	ENIGH	2014	Stat. Office	19459	57.4	2272	64.5%	3.8	19	1	COICOP	
Montenegro	HBS	2009	World Bank	1223	134.1	3764	62.7%	3	7	3	COICOP	Cant separate categories 3 & 4
Morocco	ENCDM	2001	World Bank	14243	87.5	1679	61.6%	5.9	47	17	COICOP	
Mozambique	IOF	2009	World Bank	10809	48.7	363	28.9%	4.7	6	6	COICOP	
Niger	ENCBM	2007	World Bank	4000	153.3	325	17.2%	6.4	15	6	COICOP	
Papua NG	HIES	2010	World Bank	3811	111.2	1002	11.3%	5.1	6	1	COICOP	
Peru	ENAHO	2017	Stat. Office	43545	78.5	2609	76.8%	3.9	41	8	Country-specific	
Rwanda	EICV	2014	World Bank	14419	53.6	417	17.1%	4.6	11	8	COICOP	Pre-filled items
SaoTome	IOF	2010	World Bank	3145	105.9	705	68.1%	3.8	21	3	COICOP	
Senegal	EDMC	2008	World Bank	1443	517.8	640	100.0%	7.7	41	1	COICOP	Only urban
Serbia	HBS	2015	World Bank	6531	106	1888	61.9%	2.8	9	2	COICOP	
South Africa	IES	2011	U. of Cape Town	25325	44.2	3557	67.3%	3.8	6	1	COICOP	Cant separate categories 3 & 4
Tanzania	HBS	2012	World Bank	10186	317.8	478	21.9%	5	13	2	COICOP	Cant separate categories 3 & 4
Tunisia	ENBCNV	2010	Stat. Office	11281	139.1	1732	67.6%	4.3	9	1	Country-specific	Cant separate categories 3 & 4
Uruguay	ENIGH	2005	Stat. Office	7042	77.5	2855	84.9%	3	39	1	COICOP	

Consumption Module Structure

Expenditure surveys do not have a fully homogeneous structure across countries. Table B3 presents information on their structure and we provide a summary below:

- **Number and frequency of modules**
  The number of consumption modules ranges from 1 to 17 across countries in the sample. All surveys have a module which is a diary of consumption over some short to medium period of time and some countries complement these with recall modules for more infrequent purchases. For example, Costa Rica has a single consumption module, while Morocco has 17, with modules specialized by frequency and products. Surveys with multiple modules typically asked for consumption linked to the frequency of expenditures (e.g. weekly diary, quarterly recall).

- **Durables**
  Durable items, which are not purchased frequently are included whenever available, but their inclusion is more probable in surveys which have recall modules.

- **Self-production**
  Self production is included as a "place of purchase" for all countries but Chile where it was not available. In some countries, it was pre-coded as an option for the place of purchase, while in other cases we added it as a place of purchase based on other variables, such as "mode of acquisition," which had "purchased or "self/home production." Self-production values are typically asked as value if you had purchased(or sold) this item at a market place.

- **Product codes**
  All surveys have product codes for each consumption item, which typically follow the United Nations Classification of Individual Consumption According to Purpose (COICOP) or which we could matheced to the COICOP with a cross-walk. For a few countries the COICOP classification was not available and we could not find a product crosswalk. We used the nationally-specific product classification scheme for these countries (Brazil, Chad, Peru and Tunisia).

B.2 Categories of Places of Purchase

Our core sample of 31 surveys has by construction a place of purchase for each household purchase. Evidently, the names of places of purchases (PoP) available to respondents differ across countries. However, the places of purchases can be classified into broad categories which are approximately equivalent across countries. We detail below the taxonomy used in this paper, which separates the consumption of goods into five broad categories of places of purchase, and services into four broad categories.

- **Goods**

(1) Non-market consumption (i.e. Self-production)

(2) Market consumption, no store front (i.e. markets, street stalls)

(3) Market consumption, corner and convenience shops

(4) Market consumption, specialized shops (i.e. brand stores, bakeries)

(5) Market consumption, large stores (i.e. supermarkets, malls)

- **Services**

(6) Services provided by institutions (i.e. bank, hospital, university)

(7) Service provided by individuals (i.e. maid services, gardening)

(8) Entertainment (i.e. restaurants, hotels)

(9) Informal Entertainment (i.e. food truck)

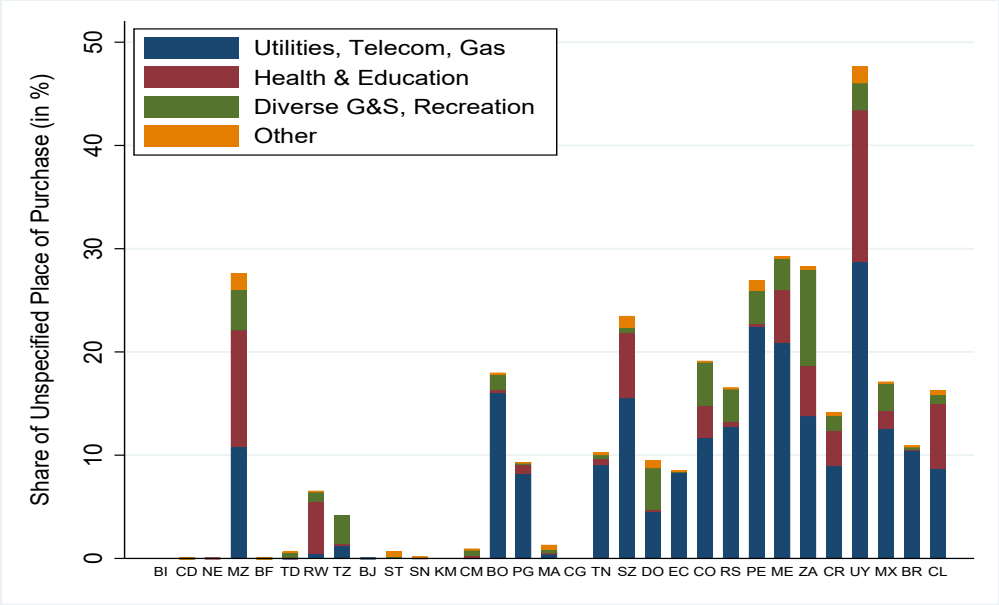
- **Unspecified**

(99) N.A/other (i.e. other, not applicable, unspecified)

The majority of countries have places of purchase for each of the five good categories. In some countries one of these categories is missing, all such cases are reported in the last column of table B3. Four countries in particular do not distinguish between specialized stores (category 4) and corner/convenience stores (category 3). For these countries we imputed at the decile level the relative shares of categories 3 and 4, based on countries with income up to 50% smaller or larger.

For services, it is more frequent that some of categories are missing. In particular some countries do not have a detailed list of institutions as potential places of purchases for services. These are typically also countries in which the share of expenditures with

Figure B.1: Average Share of Unspecified Category by COICOP



an 'unspecified' place of purchase is relatively large. Indeed when looking at what types of products compose the unspecified category, over half are utilities, while the remaining is principally education and health spending.

Finally, we assign the remaining places of purchase that are harder to categorize (as purchases over the internet or from abroad) to category (6) "services provided by institutions". While this might not be accurate, we note that these PoP typically represent a very small share of total expenditure.

The countries-specific assignment of places of purchase to the broad categories presented above is detailed in Table B4, for each PoP representing more than 0.5% of total purchases. The table also reports the share of expenditures purchased from each category, including the unspecified category.

Finally, we note that, to the best of our knowledge, the only other project which constructs a common taxonomy of places of purchases across countries is the International Price Comparison (ICP) project, which builds purchasing power parity indexes. The ICP provides a store type classifier for marketed consumption which is used by individual countries to obtain price quotes from a variety of retailer types. Our classification mirrors that of the ICP.

B.3 Global Consumption Database

A limitation of our core sample of 31 countries is that it tends to be geographically clustered and in particular does not contain countries in Asia. To obtain an idea on whether our results might be relevant for all low and middle income countries, we compare food expenditure with that of the Global Consumption Database (GCD). The GCD is the most comprehensive data source on consumer spending patterns in developing countries to date, by assembling all available representative household surveys across countries. In particular it contains data on most countries in Asia. The dataset is curated by the World Bank, and aggregate consumption statistics and further details on sources and methodology are available at http://datatopics.worldbank.org/consumption/.

We obtained access to the Global Consumption Database microdata, in order to compare food expenditure in our core sample of 31 countries to the 79 low and middle income countries available in the GCD. From our sample of 31 countries, 21 countries overlap with the GCD and usually have the exact same survey as an original source. With this enhanced dataset, which represents 51% of the world population,[39] we measure food consumption as a share of total consumption and the slope of the food Engel curve. First we note that for the 21 overlapping countries we find Engel curve slopes within 5% of the GDC estimate. Second we compare food expenditure patterns in our core sample 31 countries to the 58 countries which only appears in the GCD (see Table B5 for a list of countries in each sample). We find remarkably similar food expenditure shares and food Engel curve slopes, as a function of development, which supports that our core sample informality measures could be informative to the entire population located in developing countries, with the caveat that informality patterns might still differ geographically.

[39] We exclude rich countries by design, but a few populous countries such as China, Egypt and Iran are not part of the GCD. This explains the lion's share of the missing population in the GCD sample.

Table B4: Country-Specific Places of Purchase

BENIN

Assigned	%	Original name	Classification
Formal	1.9	autre lieu d'achat formel sur le territoir	5 large stores
	0.7	achat dans un super marché	5 large stores
	0.4	achat dans un magasin ou un atelier formel	4 specialized shops
	0.6	achat dans un magasin ou un atelier formel	4 specialized shops
	2.8	achat au secteur public ou parapublic	6 institutions
	0.2	achat à l'étranger	6 institutions
Informal	28.3	achat au domicile du vendeur, dans une pet	3 corner shops
	22.8	achat sur un marché public	2 no store front
	26.1	achat chez un ambulant, ou poste fixe sur	2 no store front
	7.0	autre lieu d'achat informel (indépendant)	1 non-market
	0.6	cadeau recu	1 non-market
	8.6	bien ou service autoproduit	1 non-market
Unspec.	0.0	other	99 n.a./other

BOLIVIA

Formal	0.9	supermercado	5 large stores
	11.5	tienda especializada	4 specialized shops
	0.4	transporte	6 institutions
	0.0	instituciones financieras	6 institutions
	0.0	utilidades	6 institutions
	1.2	institución de salud	6 institutions
	0.5	comunicación	6 institutions
	0.0	institución religiosa	6 institutions
	3.5	instituto educativo	6 institutions
	1.5	hotel, bar, restaurante	8 entertainment
	0.1	internet	8 entertainment
Informal	14.2	tienda de conveniencia	3 corner shops
	0.2	taller	3 corner shops
	2.0	vendedor ambulante	2 no store front
	3.5	puesto/kiosco	2 no store front
	3.9	feria	2 no store front
	19.2	mercado	2 no store front
	1.5	auto consumo	1 non-market
	1.9	de un hogar / transferencia	1 non-market
	0.4	servicio individual	7 service from individual
	0.2	recreación	9 informal entertainment
	5.4	cantina	9 informal entertainment
Unspec.	28.0	other	99 n.a./other

BRAZIL

Formal	11.5	supermarket	5 large stores
	0.7	department store	5 large stores
	22.1	specialized shop	4 specialized shops
	6.5	vehicle	4 specialized shops
	4.0	pharmacy	4 specialized shops
	0.0	real estate agent	6 institutions
	0.1	non-profit	6 institutions
	0.7	bank	6 institutions
	0.6	internet	6 institutions
	0.2	public administration	6 institutions
	1.5	health institution	6 institutions
	2.5	education institution	6 institutions
	0.3	public health	6 institutions
	0.4	communication company	6 institutions
	0.5	lottery	6 institutions
	2.7	restaurant	8 entertainment
	0.1	hotel	8 entertainment
Informal	3.4	grocery store	3 corner shops
	0.8	small shop	2 no store front
	0.2	market	2 no store front
	1.2	fair	2 no store front
	1.4	small market	2 no store front
	1.4	street seller	2 no store front
	8.3	person	1 non-market
	0.2	own production	1 non-market
	0.3	from farm	1 non-market
	0.4	own production_from other household	1 non-market
	4.8	private service	7 service from individual
	2.1	bar-cafe	9 informal entertainment
	0.5	recreation events	9 informal entertainment
Unspec.	20.8	other	99 n.a./other

BURKINA FASO

Formal	0.6	magasin de gros a petits prix	5 large stores
	0.7	grands magasin	5 large stores
	0.3	supermarche	5 large stores
	0.4	service de soins personnels	4 specialized shops
	0.1	librairie, papeterie	4 specialized shops
	0.1	vente de piece de rechange pour moto	4 specialized shops
	0.2	boucherie	4 specialized shops
	0.5	quincallerie (petite taille)	4 specialized shops
	0.1	poissonnerie	4 specialized shops
	0.1	pressing, blanchisserie	4 specialized shops
	0.0	boutique de station service	4 specialized shops
	1.3	atelier, service reparation	4 specialized shops
	0.0	vendeur vehicules concessionaire	4 specialized shops
	0.0	boulangerie, patisserie	4 specialized shops
	1.7	station service (lubrifiants)	4 specialized shops
	1.0	pharmacie	4 specialized shops
	0.3	clinique, laboratoire medical prive	6 institutions
	1.0	clinique, laboratoire medical public	6 institutions
	0.7	ecole, lycee, universite publics	6 institutions
	0.2	etranger	6 institutions
	0.0	pointe vente internet	6 institutions
	1.2	ecole, lycees, universite privas	6 institutions
	0.6	cabine telephone privee	6 institutions
	2.0	telephone, eau, electricite	6 institutions
	0.1	cabine telephone publique	6 institutions
	0.3	autres services publics	6 institutions
	0.2	service de transport public	6 institutions
	0.0	service postal	6 institutions
	1.8	bar, cafe, restaurant, hotel	8 entertainment
Informal	14.3	boutique de quartier	3 corner shops
	0.1	petit marche boutique (ecole, universite)	3 corner shops
	0.3	mini marche, autre magasin non specialise	3 corner shops
	0.4	kiosque ou echoppe marche	2 no store front
	0.1	rue	2 no store front
	0.9	kiosque ou echoppe quartier	2 no store front
	0.0	foire, salon, festival	2 no store front
	40.5	marche	2 no store front
	1.0	marchant ambulants	2 no store front
	10.1	bien ou service autoproduit	1 non-market
	0.0	moulin	1 non-market
	12.4	menage	1 non-market
	0.5	cadeau recu en nature ou en espace	1 non-market
	2.1	autres service prives	7 service from individual
	1.4	service de transport prive	7 service from individual
Unspec.	0.3	other	99 n.a./other

BURUNDI

Assigned	%	Original name	Classification
Formal	3.6	Autre lieu d'achat formel	5 large stores
	0.1	Supermarche	5 large stores
	0.8	Magasin, atelier formel (societe) tenu	4 specialized shops
	0.4	Magasin ou atelier formel (societe)	4 specialized shops
	0.4	Hors lieu de residence ou a l'etranger	6 institutions
	2.0	Secteur public ou parapublic	6 institutions
Informal	26.8	Domicile du vendeur, petite boutique	3 corner shops
	32.9	Marche public	2 no store front
	3.9	Vendeur ambulant ou poste fixe sur voie	2 no store front
	0.9	Cadeau Recu	1 non-market
	14.3	Bien ou service autoproduit	1 non-market
	13.9	Autre lieu d'achat informel	1 non-market
Unspec.	0.0	Other	99 n.a./other

CAMEROON

Formal	1.0	Supermarche/Grand magasin	5 large stores
	3.8	Magasin specialistes	4 specialized shops
	2.9	Presetation de services publics	6 institutions
	7.5	Secteur transport	6 institutions
	2.1	Cliniques	6 institutions
	7.0	Hotels/bars/restaurants	8 entertainment
Informal	10.7	Epiceries/Boutiques/Echoppes	3 corner shops
	0.8	Vendeurs specialises hors magasins	2 no store front
	3.4	Kiosque de jeux et Call Box	2 no store front
	3.4	Vente ambulante	2 no store front
	26.4	Marches	2 no store front
	14.8	Don, cadeau recu	1 non-market
	1.9	Domicile de vendeur	1 non-market
	3.6	Auto production	1 non-market
	0.9	Dans la nature/forit/brousse	1 non-market
	2.5	Prestation de services individuels	7 service from individual
Unspec.	7.3	Other	99 n.a./other

CHAD

Formal	0.7	Supermarche	5 large stores
	5.7	Boutique	4 specialized shops
	1.1	Magasins	4 specialized shops
	0.6	Prestataire service santé public	6 institutions
	4.1	Autre prestataire de service privé	6 institutions
	0.7	Enseignement privé	6 institutions
	0.7	Transport privé	6 institutions
	0.0	Ailleurs dans le monde	6 institutions
	0.1	Pays africain hors zone franc	6 institutions
	0.6	Enseignement public	6 institutions
	0.9	Autre prestataire de service public	6 institutions
	0.6	Prestataire service santé privé	6 institutions
	0.1	Transport public	6 institutions
	0.1	Dans un pays de la zone franc	6 institutions
	1.6	Hôtel, Restaurant, ..	8 entertainment
Informal	0.5	Echoppe	3 corner shops
	2.6	Marchand ambulant	2 no store front
	25.8	Marché de quartier ou spécialisé	2 no store front
	30.1	Marché centraux	2 no store front
	1.2	Tablier	2 no store front
	17.3	Self-Consumption	1 non-market
Unspec.	4.9	Other	99 n.a./other

CHILE

Formal	0.1	multitiendas	5 large stores
	26.5	supermercados	5 large stores
	4.3	distribuidoras - mayoristas	5 large stores
	1.6	ferreterÍas y multiferreterÍas	4 specialized shops
	0.5	tienda especializada	4 specialized shops
	4.9	farmacias	4 specialized shops
	2.0	internet	6 institutions
	0.5	extranjero	6 institutions
	2.4	hospital pÚblico y consultorios	6 institutions
	5.4	clÍnicas	6 institutions
	1.0	restaurantes y bares	8 entertainment
Informal	13.3	almacÉn tradicional	3 corner shops
	3.3	ferias libres	2 no store front
	2.9	comercio ambulante	2 no store front
	0.9	vegas - mercados	2 no store front
Unspec.	30.5	other	99 n.a./other

COLOMBIA

Formal	0.3	Supermercado de cajas de compensación	5 large stores
	0.2	Centrales mayoristas de abastecimiento	5 large stores
	0.0	Cooperativas, fondos de empleados y comisa	5 large stores
	4.7	Supermercados de barrio	5 large stores
	10.1	Almacenes o supermercados de cadena y tien	5 large stores
	2.1	Plazas de mercado y galerías	5 large stores
	0.6	Hipermercados	5 large stores
	0.0	Cigarrerías, salsamentarias y delicatesse	4 specialized shops
	11.1	Establecimientos especializados en la vent	4 specialized shops
	1.7	Farmacias y droguerías	10 pharmacies
	1.0	Televentas y ventas por catálogo	6 institutions
	0.0	A través de Internet	6 institutions
	5.1	Restaurantes	8 entertainment
Informal	0.9	Graneros	3 corner shops
	13.5	Tiendas de barrio	3 corner shops
	0.1	Bodegas o fábricas	3 corner shops
	0.2	Misceláneas de barrio y cacharrerías	3 corner shops
	0.0	Ferias especializadas: artesanal, del hoga	2 no store front
	1.7	Vendedores ambulantes o ventas callejeras	2 no store front
	0.1	Mercados móviles	2 no store front
	0.3	Sanandresitos	2 no store front
	1.1	Persona particular	1 non-market
	0.9	Transfers, from household	1 non-market
	10.3	Self production	1 non-market
	1.0	Cafeterías y establecimientos de comidas	9 informal entertainment
Unspec.	32.9	Other	99 n.a./other

COMOROS

Formal	2.0	Achat dans un super marché	5 large stores
	16.9	Autre lieu d'achat formel	5 large stores
	4.9	Achat dans un magasin ou un atelier formel	4 specialized shops
	8.3	Achat dans un magasin ou un atelier formel	4 specialized shops
	6.1	Achat hors lieu de résidence ou à l'étr	6 institutions
	8.6	Achat au secteur public ou parapublic	6 institutions
Informal	19.1	Achat au domicile du vendeur, dans une pet	3 corner shops

CONGO DRC

	12.6	Achat sur un marché public	2 no store front
	5.5	Achat chez un ambulant, ou poste fixe sur	2 no store front
	4.2	Bien ou service autoproduit	1 non-market
	2.3	Cadeau reçu	1 non-market
	8.3	Autre lieu d'achat informel (indépendant)	1 non-market
Formal	0.2	Autre secteur formel	5 large stores
	0.5	Achat supermarche	5 large stores
	3.2	Achat magasin indo-pakistanais	4 specialized shops
	3.1	Achat secteur public	6 institutions
	0.0	Achat etranger	6 institutions
Informal	3.8	Achat magasin non indo-pakistanais	3 corner shops
	36.5	Achat marche public	2 no store front
	10.1	Achat Ambulant	2 no store front
	17.5	Bien ou service autoproduit	1 non-market
	5.8	Autre lieu informel	1 non-market
	17.9	Achat domicile	1 non-market
	1.4	Cadeau recu	1 non-market
Unspec.	0.1	Other	99 n.a./other

COSTA RICA

Assigned	%	Original name	Classification
Formal	1.2	tienda por departamentos	5 large stores
	17.1	supermercado	5 large stores
	0.3	centro comercial	5 large stores
	11.3	local especializado	4 specialized shops
	0.3	verduleria	4 specialized shops
	4.3	gasolinera y estación de servicio	4 specialized shops
	1.1	carniceria / pescadería	4 specialized shops
	1.0	salones de estética o belleza	4 specialized shops
	3.4	almacén de electrodomésticos y de tecnol	4 specialized shops
	0.4	veterinaria / tienda para animales	4 specialized shops
	3.4	tienda de ropa / zapatería / perfumería	4 specialized shops
	0.3	Ópticas / tiendas de artículos ortopédi	4 specialized shops
	0.0	centro diurno de ancianos	6 institutions
	0.2	centro recreativos / deportivos (estadios)	6 institutions
	1.9	laboratorio / clínica / centro médico	6 institutions
	0.0	otras instituciones del estado	6 institutions
	0.3	instituciones educativas privadas	6 institutions
	0.2	por internet (compras)	6 institutions
	0.0	hogar comunitario	6 institutions
	1.1	en el exterior	6 institutions
	0.2	instituciones educativas públicas	6 institutions
	0.2	otras instituciones sin fines de lucro	6 institutions
	0.0	organización religiosa	6 institutions
	0.1	comedor escolar y colegial	8 entertainment
	3.9	restaurante / soda / cafetería / helader	8 entertainment
	0.0	comedor universitario	8 entertainment
	1.7	comedor en lugar de trabajo	8 entertainment
Informal	6.2	pulpería o minisuper	3 corner shops
	0.1	bazar/pasamanería / pañalera	2 no store front
	2.4	vendedor ambulante o a domicilio	2 no store front
	0.5	feria del agricultor	2 no store front
	0.2	mercado municipal	2 no store front
	0.8	local de artículos usados	2 no store front
	9.1	recibido o comprado a otros hogares	1 non-market
	0.3	producción propia	1 non-market
	0.3	recibido gratuitamente de empresa privada	1 non-market
	0.8	retiro del negocio	1 non-market
	0.1	plaza de comidas	9 informal entertainment
	0.2	bar / discoteca	9 informal entertainment
Unspec.	25.1	other	99 n.a./other

DOMINICAN REPUBLIC

Formal	0.0	supercolmado	5 large stores
	3.8	tienda por departamentos	5 large stores
	3.6	supermercados	5 large stores
	0.1	muebleria	4 specialized shops
	0.0	tienda de compra de computadora	4 specialized shops
	0.7	tienda de respuestos de vehiculos	4 specialized shops
	1.1	taller de mecanica en general, desabulladu	4 specialized shops
	0.0	almacen de tejidos	4 specialized shops
	0.0	repadora de zapatos	4 specialized shops
	0.6	puesto de rifa de aguante y loteria electr	4 specialized shops
	1.2	tienda de ropa	4 specialized shops
	1.0	ferreterias	4 specialized shops
	0.8	carniceria	4 specialized shops
	1.2	tienda de electrodomesticos	4 specialized shops
	0.0	maderera	4 specialized shops
	0.1	pescaderia	4 specialized shops
	0.1	joyerias	4 specialized shops
	0.0	funerarias	4 specialized shops
	0.0	emp. pulidora y brilladora de pisos	4 specialized shops
	0.0	fabrica de puertas y ventanas	4 specialized shops
	0.1	taller de banesteria, carpinteria	4 specialized shops
	0.3	tienda de zapatos	4 specialized shops
	0.5	peluqueria	4 specialized shops
	0.0	sastreria	4 specialized shops
	0.1	panaderia	4 specialized shops
	0.0	modistas	4 specialized shops
	0.2	librerias	4 specialized shops
	0.0	reposteria	4 specialized shops
	0.0	gomero	4 specialized shops
	0.0	relojeria	4 specialized shops
	1.4	salon de belleza	4 specialized shops
	0.2	barberias	4 specialized shops
	2.3	farmacias	4 specialized shops
	0.0	superfarmacias	4 specialized shops
	0.0	consultorio médico	6 institutions
	0.0	optica	6 institutions
	0.3	importadora de vehiculos	6 institutions
	0.2	acueducto	6 institutions
	0.1	laboratorio de análisis clinico	6 institutions
	1.2	compañia de teléfonos	6 institutions
	0.2	instituto comercial	6 institutions
	0.2	boticas populares	6 institutions
	0.0	centro médico	6 institutions
	0.4	parada de guagua	6 institutions
	0.0	centro diagnostico especial	6 institutions
	0.0	comedores económicos	6 institutions
	1.9	envasadora de gas	6 institutions
	0.1	impuestos interno	6 institutions
	0.4	compañia de television por cable	6 institutions
	0.1	distribuidora de gas	6 institutions
	1.8	comedor popular	6 institutions
	2.4	clinica	6 institutions
	0.3	universidad	6 institutions
	1.3	hospitales	6 institutions
	1.7	corporación de electricidad	6 institutions
	0.1	consultorio dental	6 institutions
	0.0	ayuntamineto	6 institutions
	0.3	fuera del pais	6 institutions
	0.1	oficina de abogados	6 institutions
	3.5	estación de gasolina	6 institutions
	1.5	colegio	6 institutions
	0.3	agencia de viaje	8 entertainment
	0.2	hotel	8 entertainment
	0.1	cadena de comida rapida	8 entertainment
	0.1	fantasias	8 entertainment
	0.5	restaurante	8 entertainment
Informal	20.3	colmado	3 corner shops
	0.3	colmadones	3 corner shops
	0.0	minimarket	3 corner shops
	0.7	almacen de provisiones	3 corner shops
	0.0	puesto de venta de habichuelas blanditas	2 no store front
	0.6	picapollo	2 no store front
	0.5	puesto de pollo	2 no store front
	3.2	vendedora ambulante	2 no store front
	1.2	mercados	2 no store front
	0.2	ventorillo	2 no store front
	0.0	puesto de leche	2 no store front
	0.9	puestos de venta	2 no store front
	1.9	autoproduction	1 non-market
	0.0	hostal, motel, pensión	9 informal entertainment
	1.6	cafeteria	9 informal entertainment
	0.2	fondas de comida	9 informal entertainment
	0.0	heladerias	9 informal entertainment
	0.0	cine	9 informal entertainment
	0.1	bares	9 informal entertainment
Unspec.	29.5	other	99 n.a./other

ECUADOR

Assigned	%	Original name	Classification
Formal	0.0	fabricas y distribuidoras mayoristas	5 large stores
	4.0	supermercados de cadena	5 large stores
	1.2	hipermercados	5 large stores
	0.0	instrumentos musicales	4 specialized shops
	0.5	repuestos de automotores	4 specialized shops
	0.4	ferreterias	4 specialized shops
	0.2	productos textiles	4 specialized shops
	0.0	conceccionarios	4 specialized shops
	0.0	articulos deportivos	4 specialized shops
	0.1	mascotas y accesorios	4 specialized shops
	0.5	tercena/carnicera	4 specialized shops
	1.3	librerias y papelerias	4 specialized shops
	0.2	sastres, costureras, modistas	4 specialized shops
	0.0	fotocopiadoras	4 specialized shops
	0.0	inmoviliarias	4 specialized shops
	0.0	delicatessen	4 specialized shops
	1.4	otros sitios de compra especializados	4 specialized shops
	1.0	salas de belleza	4 specialized shops
	0.5	computadoras y accesorios	4 specialized shops
	0.3	agencias de turismo	4 specialized shops
	0.0	mecanica industrial	4 specialized shops
	1.1	gasolineras	4 specialized shops
	0.1	verdulera/frutera	4 specialized shops
	2.1	electrodomesticos y accesorios	4 specialized shops
	0.1	joyerias y relojerias	4 specialized shops
	4.1	ropa de todo tipo	4 specialized shops
	1.9	calzado de todo tipo	4 specialized shops
	1.2	panaderas	4 specialized shops
	0.1	lavanderias	4 specialized shops
	0.0	pescadera	4 specialized shops
	0.0	floristeria	4 specialized shops
	0.0	reparacion de calzado	4 specialized shops
	0.0	productos de atencion personal	4 specialized shops
	0.0	equipos medicos	4 specialized shops
	0.2	jugueteria	4 specialized shops
	1.2	mecanicas automotrices	4 specialized shops
	0.0	perfumerias	4 specialized shops
	0.8	muebles y enceres	4 specialized shops
	5.1	boticas y farmacias	4 specialized shops
	0.0	comisariatos de empresa	6 institutions
	2.2	establecimientos privados de salud	6 institutions
	0.0	organizaciones sociales	6 institutions
	4.7	establecimientos educativos	6 institutions
	0.1	establecimientos publicos de salud	6 institutions
	0.2	establecimientos financieros	6 institutions
	0.5	instituciones publicas	6 institutions
	0.1	servicio de telefonla publica o privada	6 institutions
	2.2	transporte de pasajeros	6 institutions
	0.1	centros de proteccion social	6 institutions
	1.2	venta por cat·logo o television	6 institutions
	0.6	aseguradoras	6 institutions
	0.4	laboratorios clinicos	6 institutions
	1.4	servicios profesionales (abogados, arqu)	6 institutions
	0.0	estacionamientos	6 institutions
	0.0	ventas por internet	6 institutions
	0.0	servicio de fletes, carga	6 institutions
	0.2	hoteles y otros serv. de alojamiento	8 entertainment
	0.1	cadena de restaurantes	8 entertainment
	0.4	centros, serv. de recreacion, estadios	8 entertainment
	2.3	restaurantes, salones	8 entertainment
Informal	12.8	tiendas de barrio	3 corner shops
	1.5	bodegas, distribuidores	3 corner shops
	10.4	mercados	2 no store front
	0.3	bazares	2 no store front
	1.1	ferias libres	2 no store front
	0.2	kioscos fijos	2 no store front
	2.0	vendedores ambulantes	2 no store front
	11.1	productos autoconsumo, autosuministro	1 non-market
	0.9	personas particulares	7 service from individual
	0.1	servicio de reparacion electrodomestico	7 service from individual
	0.3	servicio artesanal (plomero, albaoil)	7 service from individual
Unspec.	12.9	other	99 n.a./other

ESWATINI

Formal	27.5	supermarket	5 large stores
	1.4	butchery	4 specialized shops
	0.0	book store	4 specialized shops
	1.7	hardware store	4 specialized shops
	5.6	clothes/footwear/linen	4 specialized shops
Informal	5.8	grocery	3 corner shops
	0.5	spaza	3 corner shops
	4.0	street vendor	2 no store front
	1.9	market	2 no store front
	11.2	gifts/transfers	1 non-market
	6.9	self production	1 non-market
Unspec.	33.4	other	99 n.a./other

MEXICO

Formal	2.1	tiendas departamentales	5 large stores
	1.0	tiendas con membresia	5 large stores
	11.4	supermercados	5 large stores
	21.1	tiendas especificas del ramo	4 specialized shops
	0.1	internet	6 institutions
	0.0	lecheria liconsa	6 institutions
	0.5	compras fuera del pais	6 institutions
	0.7	diconsa	6 institutions
	2.4	restaurantes	8 entertainment
Informal	12.8	tiendas de abarrotes	3 corner shops
	0.6	tiendas de conveniencia	3 corner shops
	5.6	persona particular	2 no store front
	3.1	vendedores ambulantes	2 no store front
	2.0	tianguis o mercado sobre ruedas	2 no store front
	3.7	mercado	2 no store front
	1.3	auto producción	1 non-market
	0.3	cafeterias	9 informal entertainment
	2.6	loncherias, fondas, torterias , cocina	9 informal entertainment
	0.0	pulqueria, cantina o bar	9 informal entertainment
Unspec.	28.8	other	99 n.a./other

MONTENEGRO

Formal	0.1	abroad	5 large stores
	17.2	supermarket	5 large stores
	36.2	store	4 specialized shops
Informal	5.3	stall	2 no store front
	0.2	gifts received/transfers	1 non-market
	5.3	own production	1 non-market
Unspec.	35.8	other	99 n.a./other

MOROCCO

Assigned	%	Original name	Classification
Formal	0.4	Supermarket or hypermarket	5 large stores
	0.2	Herbalist	4 specialized shops
	0.0	Factory, workshop	4 specialized shops
	0.7	Gas stations (benzine)	4 specialized shops
	0.5	Small Bookshop, kiosk	4 specialized shops
	0.4	Pastry shop, bakery	4 specialized shops
	0.8	Modern clothing shop	4 specialized shops
	0.3	Modern shoe shop	4 specialized shops
	0.2	Big bookshop	4 specialized shops
	0.1	Artifacts shop	4 specialized shops
	3.1	Butcher or retail chicken seller	4 specialized shops
	0.1	Drugstore	4 specialized shops
	0.3	Retail fisher seller	4 specialized shops
	0.3	Furniture, durables shop	4 specialized shops
	1.2	Craftsman's shop (hairdresser, tailor)	4 specialized shops
	1.8	Pharmacy	10 pharmacies
	35.5	Public and semi-public agencies	6 institutions
	0.3	Public education institution	6 institutions
	0.4	Banks, financing institutions	6 institutions
	1.5	Regular transportation means (bus, train,	6 institutions
	1.7	Public administration	6 institutions
	0.7	Public baths, shower, swimming pool	6 institutions
	4.1	Private education institution	6 institutions
	0.2	Medical care in public institution	6 institutions
	0.1	Abroad	6 institutions
	1.3	Medical care in private institution	6 institutions
	0.2	Cafe, standing restaurant	8 entertainment
	0.0	Standing hotel	8 entertainment
Informal	0.3	Greengrocer's	3 corner shops
	1.9	Grocer's	3 corner shops
	9.7	Neighborhood or village grocer	3 corner shops
	0.1	Traditional meetings or exhibitions	2 no store front
	0.1	Spare parts shop	2 no store front
	2.6	Neighborhood market	2 no store front
	0.0	Old, second-hand shop	2 no store front
	0.1	Wholesale market	2 no store front
	1.3	Itinerant merchant selling on sidewalks	2 no store front
	0.1	Poultry, meat, and fish wholesale market	2 no store front
	10.7	Weekly market	2 no store front
	0.5	City market or central market	2 no store front
	0.2	Seller's residence	1 non-market
	3.6	Self production/consumption	1 non-market
	0.1	Farm, arboretum, forest	1 non-market
	0.2	Irregular transportation means (carts, etc	7 service from individual
	0.0	Non-standing hotel	9 informal entertainment
	0.7	Cafe, non-standing restaurant	9 informal entertainment
Unspec.	11.2	Other	99 n.a./other

MOZAMBIQUE

Formal	8.8	loja	4 specialized shops
Informal	18.6	mercado informal	2 no store front
	12.2	mercado	2 no store front
	31.5	auto produção	1 non-market
	0.3	transferências	1 non-market
Unspec.	28.5	Other	99 n.a./other

NIGER

Formal	0.1	Supermarche/Grand magasin	5 large stores
	0.1	Magasin specialistes	4 specialized shops
	0.0	Etranger	6 institutions
	2.6	Clinique, laboratoire, ecole	6 institutions
	6.7	Prestation services publiques	6 institutions
	6.4	Secteur transport	6 institutions
	1.5	Hotel, bar restaurant	8 entertainment
Informal	27.1	Epicerie, boutique	3 corner shops
	13.0	Marche	2 no store front
	2.9	Vente ambulante	2 no store front
	1.6	Cadeau recu	1 non-market
	15.2	Auto production	1 non-market
	4.7	Prestation service individuels	7 service from individual
Unspec.	18.1	Other	99 n.a./other

PAPUA NEW GUINEA

Formal	34.5	Supermarket	5 large stores
Informal	9.4	Small shop, canteen, tuck shop	3 corner shops
	10.5	Local market	2 no store front
	3.8	Street vendor	2 no store front
	14.2	Home production	1 non-market
	10.2	Gift	1 non-market
Unspec.	17.6	Other	99 n.a./other

PERU

Formal	3.4	Supermercado	5 large stores
	1.1	Bodega (por mayor)	5 large stores
	0.8	Panadería	4 specialized shops
	0.6	Peluquería	4 specialized shops
	0.9	Librería	4 specialized shops
	0.4	Ferretería	4 specialized shops
	0.5	Tienda especializada al por mayor	4 specialized shops
	5.8	Tienda especializada al por menor	4 specialized shops
	3.6	Farmacia	10 pharmacies
	0.0	ONG	6 institutions
	0.0	Botiquín Comunal	6 institutions
	0.1	Banco de la Nación	6 institutions
	0.2	Municipalidad	6 institutions
	0.2	Establecimientos MINSA	6 institutions
	3.4	Empresas de Transporte formales	6 institutions
	1.6	Centro de estudios	6 institutions
	1.3	Grifos de empresas	6 institutions
	0.0	Teléfono público	6 institutions
	0.0	Hospital de las FF.AA. y Policía Nacional	6 institutions
	0.2	Club / Asociación	6 institutions
	0.0	CLAS	6 institutions
	0.1	Consultorio particular	6 institutions
	0.5	Talleres formales	6 institutions
	0.0	Establecimientos ESSALUD	6 institutions
	0.6	Clínica particular	6 institutions
	0.4	Óptica	6 institutions
	1.0	Restaurantes y/ó bares	8 entertainment
Informal	14.6	Bodega (por menor)	3 corner shops
	24.0	Mercado (por menor)	2 no store front
	0.2	Bazar	2 no store front
	2.7	Feria	2 no store front
	0.1	Talleres informales	2 no store front
	0.1	Grifos informales	2 no store front
	3.3	Mercado (por mayor)	2 no store front
	0.2	Camioneta, Camión	2 no store front
	0.1	Kiosco	2 no store front
	5.0	Ambulante	2 no store front
	0.2	Self consumption	1 non-market
	0.5	Empresas de Transporte informales	7 service from individual
	0.0	Casa de huesero, Curandero	7 service from individual
Unspec.	22.3	Other	99 n.a./other

CONGO REPUBLIC

Formal	1.0	Grands magasins	5 large stores
	0.3	Supermarches	5 large stores
	7.0	Autres commerces modernes	4 specialized shops
	3.9	Secteur transports	6 institutions
	0.2	Etranger	6 institutions
	2.5	Cliniques, laboratoires médicaux et écol	6 institutions
	5.8	Prestataires de services publics	6 institutions
	3.9	Hotels, restaurants, bars, cafes	8 entertainment
Informal	3.4	Epiceries modernes	3 corner shops
	8.4	Echoppes sur marches et sur bord de route	2 no store front
	6.2	Marchands ambulants	2 no store front
	42.8	Marches	2 no store front
	3.9	Ménages	1 non-market
	4.5	Produit autoconsommes	1 non-market
	0.5	Cadeau recu	1 non-market
	5.5	Prestataires de services individuels	7 service from individual
Unspec.	0.0	Other	99 n.a./other

RWANDA

Assigned	%	Original name	Classification
Formal	0.6	Supermarket/big shop	5 large stores
	4.6	Specialized shop	4 specialized shops
	2.4	Bar/restaurant	8 entertainment
Informal	13.5	Small shop/boutique	3 corner shops
	1.7	Individual	2 no store front
	0.8	Mobile seller	2 no store front
	12.5	Market	2 no store front
	11.5	Self production	1 non-market
	26.5	From a household	1 non-market
	13.1	Service provider	7 service from individual
Unspec.	12.7	Other	99 n.a./other

SAO TOME

Formal	5.2	Lojas modernas	5 large stores
	0.1	Supermercado	5 large stores
	5.4	Grandes Lojas	5 large stores
	1.3	Outros comercios modernos	4 specialized shops
	0.5	Clinicas laboratorios medicos Hospitais	6 institutions
	0.8	Sector de transportes	6 institutions
	4.3	Prestates de servicios publicos	6 institutions
	0.1	Estrangeiros	6 institutions
	0.9	Hotels, restaurantes, bares, cafes	8 entertainment
Informal	33.9	Quiosque / Quitanda	3 corner shops
	24.0	Mercado	2 no store front
	7.8	Vendedor Ambulante	2 no store front
	0.5	Prendas Recebidas	1 non-market
	0.9	Campo, mato	1 non-market
	1.9	Auto Consumo	1 non-market
	0.6	Autoabastecimento	1 non-market
	3.7	Prestates de servicios individuais	7 service from individual
	1.6	Candongueiro	7 service from individual
Unspec.	6.4	Other	99 n.a./other

SENEGAL

Formal	0.2	Magasins de gros ou à prix réduits	5 large stores
	0.1	Supermarché	5 large stores
	0.2	Grand magasin	5 large stores
	0.1	Boutique de station service	4 specialized shops
	0.1	Quincaillerie (petite taille)	4 specialized shops
	0.0	Pressing, blanchisserie et assimilés	4 specialized shops
	0.2	Boucherie	4 specialized shops
	2.0	Station service (carburants, lubrifiants,e	4 specialized shops
	0.0	Vendeur de véhicules,concessionnaire	4 specialized shops
	0.0	Poissonnerie	4 specialized shops
	0.0	Librairie, papeterie	4 specialized shops
	1.8	Boulangerie, pâtisserie	4 specialized shops
	0.0	Service de soins personnels	4 specialized shops
	0.1	Atelier et service de réparation	4 specialized shops
	0.2	Pharmacies	10 pharmacies
	0.3	Ecole, lycée, université privés	6 institutions
	0.0	Points de vente sur Internet	6 institutions
	0.1	Clinique, laboratoire médical, hôpital p	6 institutions
	0.0	Etranger	6 institutions
	0.2	Cabine téléphonique privée	6 institutions
	1.2	Service de transport public	6 institutions
	0.4	Société de téléphonie et de distributi	6 institutions
	0.0	Service postal	6 institutions
	0.0	Ecole, lycée, université publics	6 institutions
	0.1	Clinique, laboratoire médical, hôpital p	6 institutions
	0.0	Cabine téléphonique publique	6 institutions
	0.0	Autres services publics	6 institutions
	11.6	Bar, café, restaurant, hôtel	8 entertainment
Informal	0.1	Mini-marchés et autres magasin non-spéci	3 corner shops
	40.3	Boutique de quartier	3 corner shops
	0.3	Kiosque ou échoppe au marché	2 no store front
	0.6	Marchand Ambulant	2 no store front
	10.8	Kiosque ou échoppe au quartier	2 no store front
	16.3	Marchés	2 no store front
	3.5	Cadeau reçu en nature	1 non-market
	1.6	Bien ou service autoproduit	1 non-market
	1.2	Autres services privés	7 service from individual
	5.1	Service de transport privé	7 service from individual
Unspec.	1.4	Other	99 n.a./other

SERBIE

Formal	8.9	Hypermarket	5 large stores
	23.8	Specialized shop	4 specialized shops
	2.9	Discounted shop	4 specialized shops
Informal	29.6	Minimarket	3 corner shops
	4.8	Market/open	2 no store front
	1.8	Gray economy	2 no store front
	5.3	Own production/Own business	1 non-market
	2.2	Gifts/received transfers	1 non-market
Unspec.	20.7	Other	99 n.a./other

SOUTH AFRICA

Formal	38.6	Chain store	5 large stores
	11.2	Other retailer	4 specialized shops
	0.3	Internet	6 institutions
Informal	0.9	Street trading	2 no store front
	2.7	Other	2 no store front
	0.6	From a household	1 non-market
Unspec.	45.7	Other	99 n.a./other

TANZANIA

Formal	0.5	Duka kubwa(Department stores)	5 large stores
	0.1	Supermarket	5 large stores
	37.3	Shop	4 specialized shops
	0.1	Aid from different institutions	6 institutions
Informal	2.4	Street vendor	2 no store front
	22.8	Market	2 no store front
	4.7	Other household	1 non-market
	0.5	Gift from other household	1 non-market
	0.0	From the bush	1 non-market
	16.0	Produced by household	1 non-market
	1.8	Gift or free	1 non-market
Unspec.	13.7	Other	99 n.a./other

TUNISIA

Formal	1.3	Hyper, supermarche	5 large stores
	67.8	Boutique privee	4 specialized shops
Informal	1.2	Point de vente marche	2 no store front
	4.5	Ambulant	2 no store front
	0.1	Foire	2 no store front
	1.6	Cadeau	1 non-market
	1.3	Auto production	1 non-market
Unspec.	22.2	Other	99 n.a./other

URUGUAY

Formal	11.7	autoservicio, cadena de supermercados	5 large stores
	0.1	fabrica, mayorista	5 large stores
	1.0	shopping o galeria	5 large stores
	0.4	papeleria, libreria	4 specialized shops
	0.1	jugueteria	4 specialized shops
	0.0	casa de musica	4 specialized shops
	0.4	muebleria	4 specialized shops
	0.5	barraca, ferreteria, vidrieria	4 specialized shops
	1.3	casa de electrodomesticos, telefonos	4 specialized shops
	0.7	verduleria, puesto, fruteria	4 specialized shops
	0.9	zapateria, marroquineria, talabarteria	4 specialized shops
	2.3	merceria, tienda	4 specialized shops
	1.5	panaderia, confiteria	4 specialized shops
	0.0	estacion de servicio	4 specialized shops
	0.2	rotiseria	4 specialized shops
	0.1	casa de computacion	4 specialized shops
	0.3	veterinaria	4 specialized shops
	2.6	carniceria, polleria, pescaderia	4 specialized shops
	0.2	fabrica de pastas	4 specialized shops
	0.7	farmacia, perfumeria, panalera	4 specialized shops
	0.8	fuera del pais	6 institutions
	0.0	internet	6 institutions
	0.5	cantina, trabajo, colegio	8 entertainment
	0.9	restaurante, parrillada	8 entertainment
Informal	7.8	almacen	3 corner shops
	0.1	almacen de ramos generales	3 corner shops
	1.1	vendedor ambulante, puesto callejero, carr	2 no store front
	0.7	quiosco, salon	2 no store front
	0.1	expoferia	2 no store front
	1.5	feria vecinal	2 no store front
	0.5	distribuidor o repartidor a domicilio	1 non-market
	0.1	agropecuaria, semilleria	1 non-market
	0.0	casa de articulos usados	1 non-market
	0.1	cibercafe	9 informal entertainment
	0.2	comida rapida, plaza de comidas	9 informal entertainment
	0.1	heladeria	9 informal entertainment
	0.8	bar, pizzeria	9 informal entertainment
	0.2	boliches nocturnos	9 informal entertainment
Unspec.	59.6	other	99 n.a./other

Table B5: GCD and IEC countries

GCD Countries	Not in GCD
Afghanistan	Brazil
Armenia	Chad
Bangladesh	Costa Rica
Belarus	Dominican Republic
Benin	Ecuador
Bhutan	Comoros
Bolivia	Papua New Guinea
Bosnia	Uruguay
Bulgaria	Tunisie
Burkina Faso	
Burundi	
Cabo Verde	
Cambodia	
Cameroon	
Colombia	
Congo DRC	
Congo Rep	
El Salvador	
Eswatini	
Ethiopia	
Fiji	
Gabon	
Gambia	
Ghana	
Guatemala	
Guinea	
Honduras	
India	
Indonesia	
Iraq	
Jamaica	
Kazakhstan	
Kenya	
Kyrgyzstan	
Laos	
Latvia	
Lesotho	
Liberia	
Lithuania	
Madagascar	
Malawi	
Maldives	
Mali	
Mauritania	
Mauritius	
Mexico	
Moldova	
Mongolia	
Montenegro	
Morocco	
Mozambique	
Namibia	
Nepal	
Nicaragua	
Niger	
Nigeria	
Pakistan	
Peru	
Philippines	
Republic of North Macedonia	
Romania	
Rwanda	
Sao Tome	
Senegal	
Serbia	
Sierra Leone	
South Africa	
Sri lanka	
Tajikistan	
Tanzania	
Thailand	
Timor-Leste	
Togo	
Turkey	
Uganda	
Ukraine	
Vietnam	
Zambia	

C Theory Appendix

C.1 Supply-side assumptions

This subsection shows that our assumptions regarding the pass-through of taxes to prices in the formal and informal sector can be modelled as a equilibrium responses of firms with a simple supply-side model.

Each variety $j1$ is produced by a firm that pays taxes (a formal firm), and each variety $j0$ by a firm that does not pay taxes (an informal firm). All firms produce using only labor L with the following production function $x_{jl} = \phi_{jl} L_{jl}, \forall l = 0, 1$, labor is paid a fixed wage w. Firms maximize their profit $\pi_{jl} = q_{jl} x_{jl} - w x_{jl} / \phi_{jl}$ where q_{jl} are the endogenous producer prices, which then determine consumer prices $p_{j1} = q_{j1}(1 + t_j)$ if the firm is formal, $p_{j0} = q_{j0}$ if the firm is informal.

We assume firms compete under monopolistic competition, which implies that firms maximize profit π_{jl} whilst taking into account the demand function $x_{jl}(p_{jl})$ they face. Writing ϵ_{jl} the price elasticity of demand for variety jl and taking the first-order-condition with respect to q_{jl} we obtain:

$$q_{jl} = \frac{\epsilon_{jl}}{\epsilon_{jl} - 1} \frac{w}{\phi_{jl}} \tag{11}$$

This implies the following expression for consumer prices:

$$p_{j1} = (1 + t_j) \frac{\epsilon_{j1}}{\epsilon_{j1} - 1} \frac{w}{\phi_{j1}} \tag{12}$$

and

$$p_{j1} = \frac{\epsilon_{j0}}{\epsilon_{j0} - 1} \frac{w}{\phi_{j0}} \tag{13}$$

This in turn implies a pass-through of one to prices in the formal sector, zero to prices in the informal sector.

C.2 Supply-chain considerations

To consider how our pass-through assumptions are affected by allowing informal retailers to buy from formal suppliers, consider an extension to the above model in which downstream firms produce varieties jl using inputs produced by upstream firms k. Upstream firms produce using only labor $x_k = L_k$. Downstream firms' production function is given by:

$$x_{jl} = \Big(\sum_k \alpha_{jlk} x_{jlk}^{\frac{\rho-1}{\rho}} \Big)^{\frac{\rho}{\rho-1}} \tag{14}$$

where x_{jlk} is the quantity of inputs k used by the downstream firm producing variety jl, and ρ the constant elasticity of substitution in production.

The consumer price of variety jl can now be written as:

$$p_{jl} = (1 + t_j f_{jl}) \frac{P_{jl}}{\phi_{jl}} \frac{\epsilon_{jl}}{\epsilon_{jl} - 1} \tag{15}$$

where $f_{jl} = 1$ if the firm producing jl is formal, zero otherwise, and P_{jl} is its input cost index. P_{jl} is obtained by cost minimization and equal to:

$$P_{jl} = \Big(\sum_k \alpha_{jlk}^{\rho} p_{jlk}^{1-\rho} \Big)^{1/(1-\rho)} \tag{16}$$

Here p_{jlk} is equal to the net of tax price paid for the product k by the firm producing variety jl. We assume the consumption tax is a Value-Added-Tax, so that if both firms k and jl are informal no tax is paid, if firm k is informal no tax is paid, and only if firm k is formal and firm jl informal the tax is paid on the transaction between them. Formally:

$$p_{jlk} = (1 + t_k f_k (1 - f_{jl})) w \frac{\rho}{\rho - 1} \tag{17}$$

Combining expressions (15), (16) and (17), we can write the pass-through of taxes to the price of formal and informal downstream firms. The pass-through of taxes to the price of formal downstream firms ($f_{jl} = 1$) is still equal to 1:

$$\frac{\partial p_{j1}}{\partial t_j} \frac{1 + t_j}{p_{j1}} = 1 \tag{18}$$

The pass-through of taxes to the price of informal downstream firms ($f_{jl} = 0$) can be written as:

$$\frac{\partial p_{j0}}{\partial t_j} \frac{1 + t_j}{p_{j0}} = s_{j0F} \tag{19}$$

where s_{j0F} is the share of formal inputs in firm $j0$'s total production costs:

$$s_{j0F} = \sum_k f_k \alpha_{j0k}^{\rho} P_{j0}^{\rho-1} p_{j0k}^{1-\rho} \tag{20}$$

C.3 Price elasticities over the development path

In this subsection we explain how we obtain the result that the price elasticities of a formal variety ϵ_{j1} is decreasing in the price of the informal sector.We drop the subscript j in what follows for simplicity. Writing total demand for the product x as the sum of demand for the formal and the informal varieties of that product ($x = x_1 + x_0$) we can decompose the price elasticity of demand for the product x as the weighted sum of own- and cross- price elasticities of demand of both varieties:

$$\epsilon = \epsilon_1 \alpha + \epsilon_0(1-\alpha) + \epsilon_{1,0}\alpha + \epsilon_{0,1}(1-\alpha) \tag{21}$$

where ϵ_{uv} is the elasticity of demand for variety u with respect to the price of variety v and $\alpha = x_1/x$ is the share of formal consumption in total consumption of the good.

Using Slutsky's equation and Slutsky symmetry to introduce the compensated cross-price elasticity $\epsilon^C_{0,1}$, assuming $\epsilon^C_0 = \epsilon^C_1$ and re-organizing we obtain the following expression:

$$\epsilon_1 = \epsilon - \epsilon^C_{0,1}(1-\alpha)(1+p_0/p_1) + \eta_0(1-\alpha)s \tag{22}$$

where η is the income elasticity of demand for the product and we have used the fact that $\eta_1\alpha + \eta_0(1-\alpha) = \eta$. We see that ϵ_1 is increasing in the formal consumption share α as long as η_0 (the income elasticity of the informal variety) is not too large relative to the compensated cross-price elasticity of demand between the two varieties.

D Calibration Appendix

D.1 Calibration of price elasticities

This section describes how we calibrate the price elasticity of demand for formal varieties ϵ_1. Recall expression (22) above:

$$\epsilon_1 = \epsilon - \epsilon^C_{0,1}(1-\alpha)(1+p_0/p_1) + \eta_0(1-\alpha)s \tag{23}$$

There are no estimates of the cross-price elasticity of demand between formal and informal varieties of different products so we use estimates of the elasticity of substitution in consumption across stores of different types in the litterature. The cross-price elasticity is related to this elasticity of substitution σ in a CES utility function by the expression $\epsilon^C_{0,1} = \sigma s_0$. Faber and Fally (2017) estimate an elasticity of substitution between large and small stores in the US of 2.2, Atkin et al. (2018b) estimate the elasticity of substitution

between foreign and domestic supermarkets and find estimates in the 2-4 range. We expect the elasticity of substitution between formal and informal stores to be somewhat lower than the parameters estimated in those papers, and use a value of 2.

To obtain an estimate of the income elasticity of demand for the informal variety we use our estimates of the slopes of the Informality Engel Curves β_0 and the expression $\eta_0 = 1 + \frac{\beta_0}{s_0}$. We set $\epsilon = -1$, in line with how we calibrate optimal tax rates in scenarios without an informal sector. We calibrate $1 - \alpha$ using the country-level share of informal consumption in total consumption and the budget shares s and s_0 using the country-level budget shares.

When calibrating the price elasticity of demand for all formal varieties ϵ_1 we use the total informal consumption share, total formal budget shares (where by definition $s = 1$) and the slope of the overall Informality Engel Curves described in the main body of the paper. When calibrating the price elasticity of demand for formal varieties of food and non-food products we the informal share of food or non-food products on consumption and the product-specific budget shares, and estimate the slopes of the food or non-food specific Informality Engel Curves for each country to calibrate income elasticities.