Informality, Consumption Taxes and Redistribution

Pierre Bachas, Lucie Gadenne & Anders Jensen*

December 2019

Abstract

Are consumption taxes progressive in developing countries? We assemble expenditure surveys from 30 countries at different levels of development to study the redistributive potential of taxation. 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. Using place of purchase information 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 and calibrate it to our data. We find that optimal uniform rates decrease with development, while relative tax subsidies on food products increase with development. Overall, the informal sector makes consumption taxes progressive: on average households in the richest quintile face an effective tax rate that is 2.5 times higher than that of the poorest quintile. Optimally differentiating rates on food versus non-food products only marginally increases progressivity, especially in the poorest countries.

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 has grown or remained high over the past 25 years (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 therefore raise revenue primarily through indirect consumption taxes, which are considered regressive (Warren, 2008), and use blunt instruments to introduce progressivity, such as reduced tax rate on necessities – typically food products. The potential negative distributional impact of consumption taxes features prominently in policy debates on tax policy's role to limit inequality in developing countries (Lustig, 2018).

In this paper we assemble a micro-dataset of expenditure surveys from 115 countries and combine it with theory to study the redistributive capacity of consumption taxes in low and middle income countries. First, we develop a new method to measure informal (untaxed) consumption from expenditure diaries available in a core set of 30 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. Calibrating the model to our data, we find that optimal uniform rates fall over the development path, while the optimal subsidy on food relative to non-food products increases. Intuitively, the downward-sloping IECs imply that taxing consumption is progressive; this progressivity falls as the informal sector shrinks with development. In poor countries moreover poorer households consume most of their food from the informal sector, weakening the redistributive potential of food subsidies.

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 30 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. We find that the aggregate share of informal consumption decreases with development, as expected, from 90% in the lowest income countries to 20% in the richest upper-middle income countries. We also find a negative correlation between informal consumption and log total expenditure *within* countries: IECs are always decreasing, and approximately linear with respect to log household expenditure. The informal consumption share decreases by 7 percentage points on average when household expenditure doubles. We 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 slightly more than half 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 find evidence suggesting 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 our uniquely large sample of 115 countries. Consistent with an extensive literature, we confirm that the food Engel curves are close to linear with respect to log household expenditure; the food budget share decreases by 8 percentage points on average when household expenditure doubles. This exercise also allows us to compare our core sample of 30 countries to this larger sample, which is representative of 60% of the world population: food consumption patterns look similar in our core countries and in other countries of similar development levels. This suggests that our results regarding Informality Engel Curves may also generalize to a larger sample.

Our second contribution is to characterize the implications of these 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 allow for changes in consumption patterns over the development path. The presence of an informal sector increases the efficiency cost of taxation, because households can substitute to informal varieties when taxes increase,

but makes consumption taxes progressive when IECs are downward-sloping. Calibrating the model using data for each country in our core sample yields two main findings. First, we find that optimal consumption tax rates fall with development because of the informal sector. This result is due to the (approximate) log-linearity of the IECs, which implies that as 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, turning to the 'de jure' rate differentiation policy in place in most countries, we find that the optimal level of subsidy on food relative to non-food products is lowest in poorer countries. This is true even if we assume all 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.[2] Taking into account informal consumption re-enforces this result, because poorer households consume most of their food in the informal sector in low-income countries. For the poorest countries in our sample our results suggest that taxing food less than non-food products is simply not optimal.

We finally quantify the progressivity of different tax scenarios in each country in our core sample. We find that setting a uniform rate on consumption is progressive once the informal sector is taken into account: the effective tax rates paid by the top 20% are 2.5 times higher than those paid by the bottom 20% on average. Progressivity is highest amongst low and lower-middle income countries, where on average the top 20% pay 3 times higher effective rates than the bottom 20% (and up to 6 times higher). 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. 'De jure' progressivity attempts on the other hand only have a limited redistributive potential.

Our baseline results rely on several assumptions that 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 un-

[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.

[2] This statement is conditional on food budget shares not being too large, a condition met in all countries in our sample.

ambiguously informal (non-market consumption) or formal (large supermarkets). Second, we allow for the efficiency cost of taxation to decrease with development as the informal sector shrinks. Our main take-away is unchanged when we calibrate this for plausible parameter values. Third, we use total expenditures to proxy for household income because of data limitations. Allowing for savings rates to increase with income decreases the redistributive effect of consumption taxes, as expected, but only slightly so.[3] Finally, 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 can redistribute. 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;

[3]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.

Boadway and Sato, 2009) which mainly focuses on the efficiency properties of taxation.[4] 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.[5] 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,

[4]One exception is Kopczuk (2001) who considers optimal income tax policy under heterogeneous evasion behavior and shows that avoidance schemes can redistribute when used mostly by the needy.

[5]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).

2014), at the worker-level (Gerard and Gonzaga, 2016; Jensen, 2019), or both (Ulyssea, 2018). Our approach complements theirs by considering the 'demand side' of informality through consumers' shopping behavior in formal and informal retailers.[6] These studies 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 that shed light on the redistributive potential of consumption taxes: a 'core sample' and a 'large sample'. We detail below the method we use to proxy for informal consumption at the household level, and the core sample of 30 countries for which the data required to apply the method is available. We then detail our 'large sample' of 115 countries to document product-specific Engel curves, focusing on the food Engel curve.

2.1 Core Sample

Our core dataset combines nationally representative household expenditure surveys from a wide range of countries. We use surveys which satisfy three criteria. First, they are nationally representative (and from the 21st century): this enables us to study the progressivity of different tax scenarios using representative samples. Second, they use open consumption diaries rather than pre-filled diaries which only document selected products: this we the data contains information on all household expenditure, including durables. 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.[7]

We obtain survey from 30 countries which satisfy these criteria. Our results currently use 26 countries and our core dataset contains information for nearly 400,000 households.[8] Table 1 lists alphabetically the countries in the core data, with survey names, years, number of households, and average number of expenditure items reported by

[6] 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).

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

[8] Analysis of the last 4 countries is on-going.

households. Countries in the sample are principally located in Latin America and Sub-Saharan Africa, with the exceptions of Morocco, Papua New Guinea, and Tunisia. The paucity of Asian countries is due to their household expenditure surveys not reporting places of purchase.[9] We cover however a wide range of development levels, from Burundi (GDP per capita of 244 USD in PPP) to Chile (15,000 USD). We show in section 3.2 that our core sample is representative of all developing countries along the dimensions we observe in our large sample.

Appendix B details the data sources used and the surveys considered for inclusion, and explains the reasons for discarding specific surveys. Table B1 details the structure of the surveys' expenditure modules for each country in our 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 source. We first discuss how we allocate each place of purchase reported in each survey to a '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, barter), (2) non brick-and-mortar stores (e.g. street stalls, public markets), (3) corner and convenience stores, (4) specialized stores (e.g. pharmacies, 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, the vast majority of these pertain to expenditure on utilities and telecoms.[10]

We assign each of these categories to either the formal or informal sector. We classify

[9] Survey design appears strongly correlated across countries within regions, showing the influence of regional development partners and/or historical ties across statistical administrations. For example, surveys from francophone countries in sub-Saharan Africa more frequently feature the place of purchase question than anglophone countries.

[10] We exclude expenditure on housing from our analysis because of lack data. Whilst most surveys ask questions about rent expenditure, few document imputed rents, making a comprehensive analysis of housing expenditure unfortunately unfeasible.

Table 1: Core Sample: Household Expenditure Surveys

Country	Code	Survey	Year	GDP per capita	Sample size	Nb items
Bolivia	BO	ECH	2004	1657.5	9149	49.42
Brazil	BR	POF	2009	10538.8	56049	47.99
Burkina Faso	BF	EICVM	2009	562.8	8404	72.02
Burundi	BI	ECVM	2014	244.1	6681	89.76
Cameroon	CM	ECAM	2014	1428.2	10303	95.76
Chile	CL	EPF	2017	15059.5	15237	129.19
Colombia	CO	ENIG	2007	5910.3	42733	79.65
Comoros	KM	EDMC	2013	1373.4	3139	83.80
Congo (DRC)	CD	E123	2005	300.8	12098	106.93
Congo (Rep. of)	CG	ECOM	2005	2503.3	5002	84.76
Costa Rica	CR	ENIGH	2014	9065	5705	67.48
Dominican Rep.	DO	ENIGH	2007	5121.1	8363	89.07
Ecuador	EC	ENIGHUR	2012	5140.3	39617	88.65
Eswatini	SW	HIES	2010	4168.5	3167	43.9
Mexico	MX	ENIGH	2014	9536.6	19459	57.36
Morocco	MA	ENCDM	2001	2091.2	14243	87.5
Mozambique	MZ	IOF	2009	404.6	10809	48.75
Niger	NE	ENCBM	2007	330.6	3980	61.05
Papua New Guinea	PG	HIES	2010	2004.8	3811	111.17
Peru	PE	ENAHO	2017	6172.7	43530	56.43
Rwanda	RW	EICV	2014	672.6	14419	89.53
Sao Tome e Principe	ST	IOF	2010	1094.7	3145	105.89
South Africa	ZA	IES	2011	7416.7	25325	44.2
Tanzania	TZ	HBS	2012	747.7	10168	185.44
Tunisia	TN	ENBCNV	2010	4142	11281	139.12
Uruguay	UY	ENIGH	2005	9068.2	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.

a category as belonging to the formal sector if there are reasons to believe 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 firms do not remit taxes (de facto informality) (see Keen and Mintz, 2004); we do not distinguish between these two forms of informality.[11]

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

[11]Note that the prevalence of de jure informality may itself be endogenous to the level of de facto informality: the legal decision to exempt small firms from commodity taxes is often in itself motivated by the fact that enforcement (to the authorities) and compliance (to the firms) costs are large for those firms relative to their potential tax liabilities.

is also strong. In particular 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. Kumler et al. (2015) find that compliance is indeed increasing in firm size in the context of Mexico. We also expect larger stores to have more customers, which makes evasion harder to sustain: Naritomi (2018) shows that if customers have either intrinsic motives or are financially rewarded to report evasion, the likelihood of evasion will decrease with the size of the customer base. More generally Lagakos (2016) follows Bronnenberg and Ellickson (2015) in categorizing 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 argues that traditional retailers are mostly informal.[12]

We use information on VAT registration by firm size, available in the retail firm census of four core sample countries (Cameroon, Mexico, Peru and Rwanda) to check that the correlation between formality and firm size is indeed positive. Figure 1 Panel (a) shows the share of formal firms as a function of log employment in each country.[13] 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 one core country (Mexico) the census classifies retailers in categories which are similar to our 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 see that non-brick-and-mortar stores and convenience stores are small and rarely formal, whereas nearly all large stores are formal.

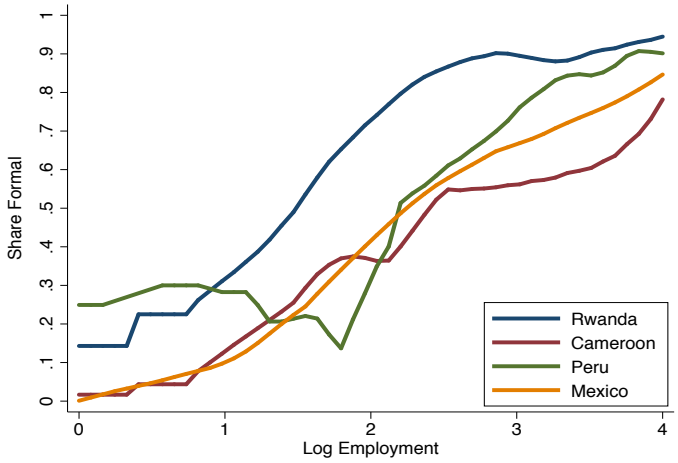
Our baseline formality assignment follows that in Lagakos (2016) and assigns categories (1) to (3) to the informal sector, and categories (4) and (5) to the formal sector. Goods in category (1) – non-market consumption – are by definition not purchased in markets and therefore untaxed. Categories (2) and (3) (non brick and mortar stores and corner stores) are likely very small and mostly informal, whereas category (5) (large

[12]Related, 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 taxing capital is easier than taxing labor in these countries. This also implies higher compliance rates amongst larger retailers.

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

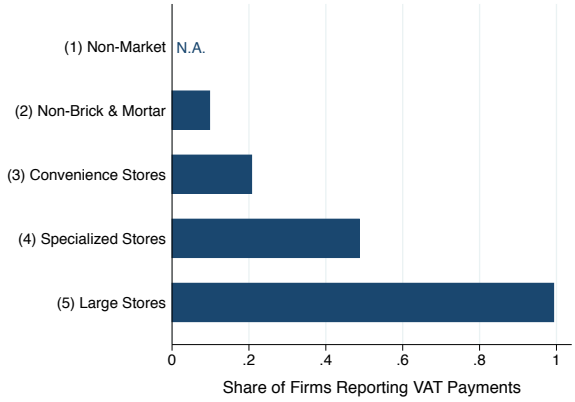
Figure 1: Employment Size, Formality & Store Types

(a) # Employees on Formality in Retail Censuses



(b) # Employees by Store in Mexico

(c) % Paying VAT by Store in Mexico



Source: Firm Censuses in Cameroon (Recensement General des Enterprises 2010), Mexico (Censo Economico 2013), Peru (Censo Nacional Economico 2008), Rwanda (Establishment Census 2014). Only firms in the retail category. Panel (a) plots the share of formal firms on log employment. Formality is defined as paying the VAT (Mexico, Peru) or being registered with the tax authority (Cameroon, Rwanda). Panel (b) and (c) are based only on the Mexican census which contains a retailer classification mirroring the expenditure surveys. Panel (b) shows log median employment by retailer type. Panel (c) shows the share of firms reporting VAT payments by retailer type.

stores) consists mostly of supermarkets which are unlikely to fly below the radar of the tax authorities. For services, we assume that institutions are more likely to pay taxes than individual providers, 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 those are utilities typically provided by large institutions which cannot evade taxes. Appendix B 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.[14]

We use two alternative formality assignment rules to examine the robustness of our results. First, we consider an alternative classification 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 robustness scenario is probabilistic: to take into account the fact that some small retailers pay taxes, we assign to each category a probability that it is formal, using the formality rates observed in the Mexican retail census.[15]

Our formality assignment rule thus enables us to measure, for each household in 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).[16]

2.3 Larger Sample for Food Engel Curves

To study product-level Engel curves we use nationally representative surveys with extensive consumption diaries; such surveys are available for most countries in the world. We use the World Bank's Global Consumption Database (GCD), the most comprehensive database on consumption patterns in developing countries: it compiles household expenditure suveys for 105 countries and harmonizes product categories across all surveys at the 2-digit COICOP product level.[17] Merging this dataset with our core sample we obtain our large sample with data for 115 countries. This dataset represents 60% of the world population.[18]

We use this large sample to characterize product-specific Engel curves, focusing on

[14] A tiny share of expenditure is classified as purchased online, which we assign to the formal sector.

[15] 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.

[16] We convert survey-specific product categories to COICOP categories. This can be done at the 2-digit level in all countries, and for most countries at lower levels. The three exceptions are Brazil, Peru and Tunisia, for these countries we use survey-specific product categories at the 3 and 4 digit levels.

[17] For more information on the database, please see http://datatopics.worldbank.org/consumption/. Most surveys date from the 2007-2010 period.

[18] Populous countries which are not captured in the GCD include China, Egypt and Iran.

food products. 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 Evidence on Informality and Food Engel Curves

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 on our large sample of countries.

3.1 Informality Engel Curves

3.1.1 Results

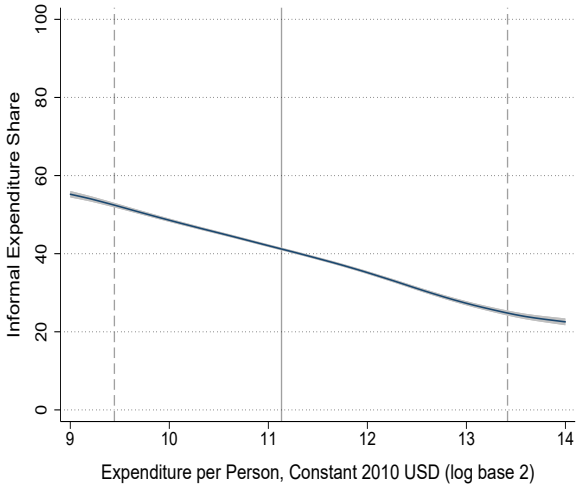
Our main object of interest is the within-country relationship between the informal expenditure share and household total expenditure, which we call the Informality Engel Curve (IEC). We follow the literature on Engel curves in developing countries and consider the relationship between expenditure shares 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 per capita expenditure per person in 2010 constant USD, measured in log base 2, such that a one unit increase on the horizontal axis corresponds to a doubling of per person expenditure. 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

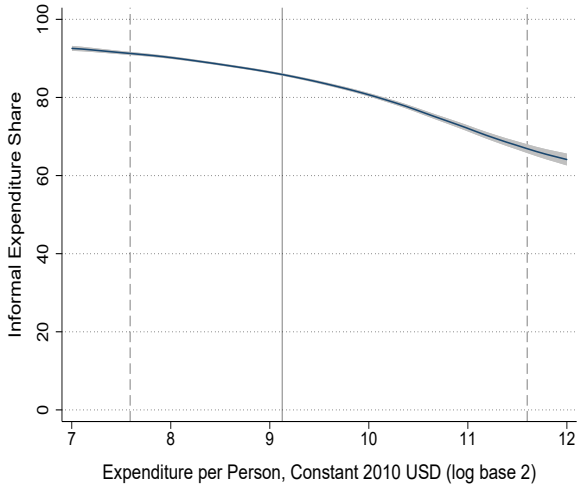
approximately linear with respect to log household expenditure. In Mexico, the informal expenditure share drops over the income distribution, from 60% to 20%. In Rwanda the IEC is slightly less steep: the informal share falls from over 90% for the poorest households to nearly 60% for the richest. Appendix Figure A.1 plots the relationship between informal shares and household expenditure in the same way for each country in our sample. We see that IECs are always downward sloping and approximately linear with respect to log expenditure.

Figure 2: Informality Engel Curves: Examples

(a) Mexico



(b) Rwanda



Local polynomial fit of the Informality Engel Curves in Mexico and Rwanda. Expenditure on the horizontal axis is measured in log base 2, such that a one unit increase on the horizontal axis corresponds to a doubling of household's expenditure. 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.

To summarize the information contained in each IEC and compare across countries we focus on two key moments: the average informal expenditure shares and the slope of the IECs. To estimate the slope we impose a linear relationship with respect to log income (base 2).[19] We obtain the slope from running the following regression separately in each country:

$$Share\ Informal_i = \beta ln(expenditure_i) + \varepsilon_i \quad (1)$$

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

[19] While the existing literature has found that product-specific Engel curves can be non-linear (Banks et al., 1997; Atkin et al., 2018a), in our setting visual inspection suggests that linearity is a reasonable approximation in most countries.

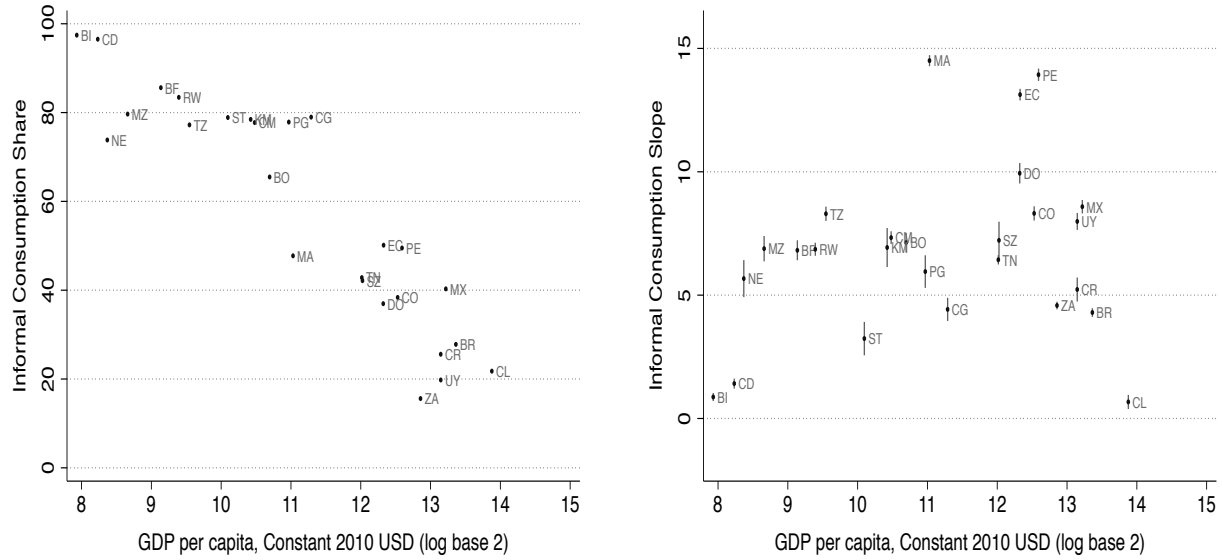
Results are presented in Figure 3. Panel (a) plots the average share of informal expenditure as a function of the country's per capita GDP (in constant 2010 USD). The average informal share is 58%, and we see a large negative correlation with the level of economic development: the average informal expenditures is around 80% in many Sub Saharan Africa countries, compared to only 20% in upper middle income countries such as Chile. This negative relationship between economic development and aggregate size of the informal sector is consistent with results in the existing literature which attempts to proxy for the size of the informal sector using very different methodologies (see in particular Enste and Schneider, 2000; La Porta and Shleifer, 2014; Morrow et al., 2019). These similarities between existing proxies for the informal sector and our consumption-based approach suggests 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 significantly further than existing approaches and consider the correlation between informal expenditure share and income at the household level for each country in our core sample. Panel (b) plots minus the country-specific estimated IEC slopes (the $-\beta$ coefficient in equation 1 above) as a function of economic development, the bars correspond to the 95% confidence intervals. In all countries the IEC slopes are negative, and significantly different from zero. The range of slopes varies from less than 1 in Chile and Burundi, and up to 15 in Morocco. The average slope across all countries is 6.8: on average the informal expenditure share is reduced by 6.8 percentage points when household expenditure doubles.

How do the above results vary based on the assignment of places of purchase to formality status? We present results for our two alternative scenarios in Appendix Figure A.2. Using alternative assignment rules changes the informal consumption averages, as expected, but country-specific estimates of the IEC slope are similar to those from our baseline assignment rule. Our results are therefore robust to changing the assumptions we make to reflect the underlying uncertainty in the formality status of different place of purchase categories. Appendix Figure A.3 explains why this is, by showing the average consumption in each decile by place of purchase on average across all countries. We see that slopes with respect to income are particularly steep for categories for which the formality assignment is more certain such as non-market purchases and large stores.

Figure 3: Informal Consumption Across Countries

(a) Average: Informal Consumption Share (b) Slope: Engel Curve of Informality



Panel (a) plots country-level informal consumption as a share of total consumption as a function of GDP per capita. Panel (b) shows the (inverse) slope of informal consumption with total per person household expenditure, on per capita GDP. The bars correspond to the 95% confidence interval of the slope coefficient. The slope measures the drop in informal consumption for a doubling of households' expenditure, within country. GDP per capita is in constant 2010 USD, and transformed in log base 2 (Source: World Bank WDI)

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. If they can be explained by household observable 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 en-

ables governments to obtain a more progressive tax system than they would otherwise.[20]

Observable household characteristics

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

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

where we gradually introduce household characteristics X_i as controls.

Table A2 presents our results. We show the across countries average of minus the coefficient 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: age, education, level and gender of the household head, and household size.[21] We find no evidence that these characteristics explain the correlation between informal consumption and income, if anything the IEC slopes increase slightly.

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.[22] 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 urban locations explains less than 15% of the slope.[23] Controlling for more detailed household location decreases the correlation further, suggesting that differences in market access by income

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

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

[22]The survey block is most granular location information in all surveys. 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).

[23]The informal budget share is 64% in rural areas and 51% in urban areas, see Appendix Figure A.4.

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)
Slope	6.8	7.4	6.3	5.5	4.1	4.0	3.4	3.0
Confidence Interval	[6.4,7.2]	[7.0,7.8]	[5.8,6.7]	[5.0,6.0]	[3.8,4.4]	[3.7,4.3]	[3.1,3.6]	[2.6,3.3]
# of p-values < 0.05	26	26	26	25	25	25	24	25
R^2 adjusted	0.20	0.22	0.27	0.42	0.51	0.54	0.54	0.58
Household Characteristics		X	X	X	X	X	X	X
Urban/Rural			X					
Census Blocks				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 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, census blocks), and product codes at the 2nd, 3rd and 4th level of the COICOP classification.

level do explain part of the differences in informal expenditure share.

In columns 5 to 7 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 effects and compute an average product-level estimate of β for each country.[24] We gradually consider variations within smaller product types: we use the COICOP 2-digit level classification in column 5, the 3-digit level classification in column 6 and the most granular 4-digit level classification in column 7. We find that different preferences across products do explain part of the correlation: controlling for broad product categories alone explains 44% of the slope. Controlling for narrow product categories slightly reduces the slope further, to on average of 3.4.[25]

Finally, column 8 shows the average IEC slope when all controls are included. The average IEC slope once all controls are included is 3, and it remains precisely estimated and statistically significant in all but two countries in our core sample. Overall, observ-

[24]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.

[25]Figure A.5 displays visually these results for each country by showing the change in slopes when controlling for increasingly narrow product groups.

able household characteristics explain nearly 60% of the correlation between informal expenditure shares and household income.

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.

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 A3 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.[26] 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 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.[27]

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).[28] 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 cannot be due to the latter not paying taxes.

[26] 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.

[27] Note that access also seems to play a major role: between a third and a half of respondents report access as their main reason for choosing a store. This is consistent with the idea that many households cannot invest in costly durables, such as cars, which may widen the set of accessible stores (Lagakos, 2016).

[28] 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.

Formally, in each country with available data we estimate the following specification:

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

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.

Table 3: Unit Values Across Places of Purchase

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Country	**Bolivia**	**Brazil**	**Burundi**	**DRC**	**Costa Rica**	**Dom.R.**	**Ecuador**	**Morocco**	**Mexico**	**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 cell. To increase statistical precision, in every country we exclude cells with 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. We winsorize unit values at 5% and 95% within country.

Results are reported in Table 3. 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 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.[29]

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.

3.2 Food Engel Curves

Most governments set reduced consumption tax rates on food products in an attempt to introduce progressivity to their consumption tax system.[30] 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 115 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 also in our core sample are presented in red. 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

[29]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.

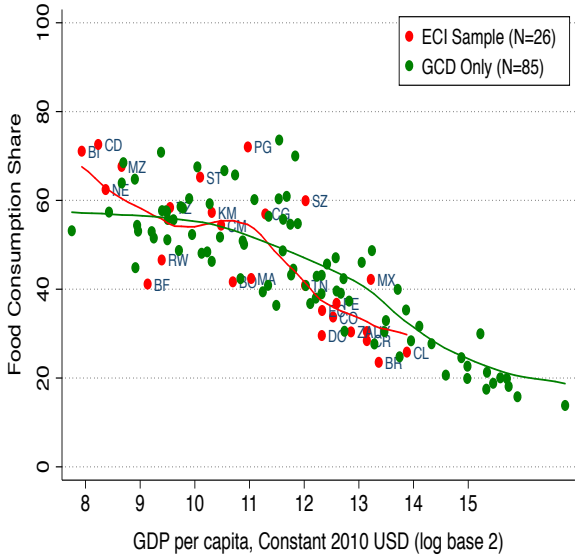
[30]Some countries apply reduced rates of exempt all food items, while other countries target 'basic' food items. In this paper, we consider a tax policy which reduces the rate on all food items. Targeting more narrow items may improve targeting but also allows firms to more easily misreport liable items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

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.

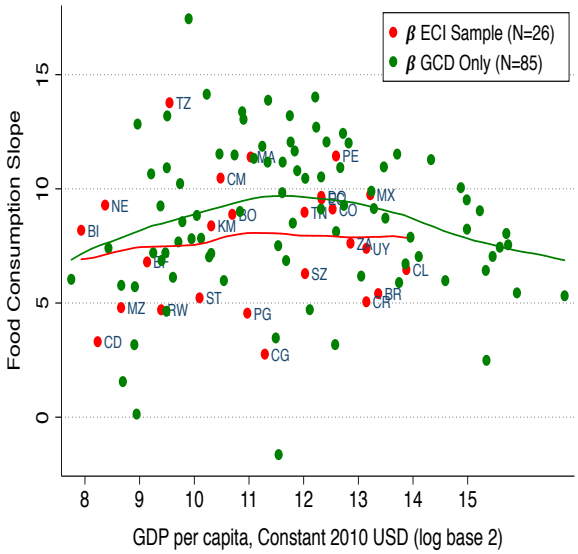
Not surprisingly, the food share of total expenditure falls with economic development (Panel A). In addition food Engel curves are as expected downward-sloping in all countries in our large sample. There is no clear relationship between slope magnitude and economic development, in line with the existing literature. We note that countries in our core sample resemble other countries at similar levels of development with respect to both the aggregate food expenditure share and the slopes of the food Engel curves. This is reassuring, as it suggests that these countries are also likely to be broadly 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 on our smaller core sample, but this evidence indicates that our conclusions may be relevant more generally.

Figure 4: Food Consumption Across Countries

(a) Average: Food Consumption Share



(b) Slope: Food Engel Curve



Source: Global Consumption Database from the World Bank and coutntry expenditure surveys for the IEC sample. Panel (a) shows country-average food consumption as a share of total consumption, plotted against log per capita GDP. 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)

4 Optimal Consumption Tax Policy over the Development Path

This section considers what the shape of the Engel curves described above imply for commodity tax policy. To do so we introduce the possibility that some goods cannot be taxed because they are produced in the 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.[31] 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.

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, and we consider below how assuming that the traditional sector cannot be taxed changes optimal tax policy. We assume that the two varieties $j0$ and $j1$ of product j are substitutes: to simplify the exposition we make two additional assumptions: we assume no cross-price elasticity between the different j products and consider a case in which there are only two types of households, poor P and rich R, with $y^R > y^P$.[32]

Producer prices q_{j0}, q_{j1} are exogenous. Consumer prices are given by $p_{jv} = q_{jv}(1 + t_{jv})$. If variety jv is produced in the informal sector, then $t_{jv} = 0$ and hence $p_{jv} = q_{jv}$. 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.

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

[32] These assumptions are relaxed in Appendix C.

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).[33] We assume throughout that $g^P > g^R$ and that $\mu = \bar{g}$, the latter simplifies expressions and corresponds to a government that has no preference for taxation unless it enables redistribution.

4.2 Optimal taxation when there is no informal sector

We consider first a set-up in which both types of varieties are taxed, equivalent to that in Diamond (1975). Assume that the government uses a uniform tax rate, $t_j = t, \forall j$. Welfare maximization yields the following expression for the optimal uniform tax rate $\tau^* = \frac{t^*}{1-t^*}s$:

$$\tau^* = \frac{g^P - g^R}{-\bar{g}\epsilon}(1 - \phi), \tag{5}$$

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 $\phi = \frac{y^P}{\bar{y}}$ is decreasing with inequality, with $\bar{y}$ average income.

The optimal tax rate increases with inequality, the strength of the government's redistributive preferences ($g^P - g^R$) and decreases in the overall price elasticity of demand. When the government can set different tax rates on products j the optimal commodity rates τ^*_j are given by:

$$\tau^*_j = \frac{g^P - g^R}{-\bar{g}\epsilon_j}(1 - \frac{s^P_j}{\bar{s}_j}\phi) \tag{6}$$

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

where $\bar{s}_j$ is product j's average budget share, and s_j^P the budget share spent by the poor. The optimal rate on product j is decreasing in $s_j^P/\bar{s}_j$, so the optimal tax schedule sets lower rates on products that are more consumed by the poor. 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 traditional varieties $j0$ cannot be taxed. The optimal uniform consumption rate becomes:

$$\tau_1^* = \frac{g^P - g^R}{-\bar{g}\epsilon_1}(1 - \frac{s_{j1}^P}{\bar{s}_{j1}}\phi) \tag{7}$$

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 the poor spend a lower budget share on the formal sector than the average ($\frac{s_{j1}^P}{\bar{s}_{j1}} < 1$). We have seen above that Informality Engel Curves are downward sloping in all countries, this implies directly that taxing only formal varieties shifts the burden of taxation towards richer households compared to taxing all varieties. Equity considerations therefore imply that the optimal level of consumption taxes is higher 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. 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.[34]

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

$$\tau_{j1}^* = \frac{g^P - g^R}{-\bar{g}\epsilon_{j1}}(1 - \frac{s_{j1}^P}{\bar{s}_{j1}}\phi) \tag{8}$$

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 food Engel curves are downward-sloping and that IECs are downward-sloping. If the food-only Informality Engel Curves are very steep the formal food Engel curves could be flat, or even upward-sloping. In this case it would 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 characterization of optimal tax policy over the course of economic development. 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:

$$\frac{\partial \tau_j^*}{\tau_j^*} = \frac{s_j^P \phi}{\bar{s}_j - s_j^P}(\frac{\partial \bar{s}_j}{\bar{s}_j} - \frac{\partial s_j^P}{s_j^P}) + \frac{\partial \epsilon_j}{\epsilon_j} \tag{9}$$

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 the poor's budget share s_j^P increases less (or

[34]Formally we can write $\epsilon_j = \alpha_j \epsilon_{j1} + (1 - \alpha_j)\epsilon_{j0} + \alpha_j \epsilon_{j1,j0} + (1 - \alpha_j)\epsilon_{j0,j1}$ where $\epsilon_{j0,j1}$ and $\epsilon_{j1,j0}$ are (uncompensated) cross-price elasticities and $\alpha_j = x_{j1}/x_j$ Assuming the varieties are strong substitutes, such that uncompensated cross-price elasticities are negative, and $\epsilon_1 \approx \epsilon_0$, we have $\epsilon_j > \epsilon_{j1}$.

decreases more) than the average budget share $\bar{s}_j$. Taxing product j redistributes more with economic development when the poor's 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 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: $\partial \bar{s}_F = \partial s_F^P = \beta_F$. Food Engel curves are downward-sloping so we also have $\beta_F < 0$ and $\bar{s}_F < s_F^P$. 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 (see Appendix).

We finally turn to how optimal policy changes over the development path in the presence of an informal sector. Here again imposing a log-linear functional forms on Informality Engel Curves helps reach a clear conclusion. The effect of a proportional increase in all households' incomes on the optimal uniform rate τ_1^* is given by:

$$\frac{\partial \tau_1^*}{\tau_1^*} = \frac{s_1^P \phi}{\bar{s}_1 - s_1^P}\left(\frac{\partial \bar{s}_1}{\bar{s}_1} - \frac{\partial s_1^P}{s_1^P}\right) + \frac{\partial \epsilon_1}{\epsilon_1} \tag{10}$$

The first term captures the change in the redistributive effect of taxing only the formal sector over the development path. The log-linearity of the IECs implies that $\partial \bar{s}_1 = \partial s_{p1} = \beta_1$ where β_1 is equal to minus the slope of the IEC. As long as the IEC is downward sloping (which implies $\beta_1 > 0$ and $s_1^P < s_1$) 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. Empirically little is known about the size of substitution

between formal and informal varieties. One possibility is that the substitution elasticity falls with economic development, as the size of the informal sector shrinks. Another possibility is that the substitution elasticity is largest in countries with equal sized formal and informal sectors (middle income) where many households are marginal between consuming formal and informal varieties for a wide range of products. Regardless, as long as the change in efficiency cost is small relative to the slope of the IEC and the increase in the size of the formal sector the first effect will dominate and optimal rates will fall over the development path.

The change in commodity-specific rate over the development path in the presence of an informal sector is given by a similar expression:

$$\frac{\partial \tau_{j1}^*}{\tau_{j1}^*} = \frac{s_{j1}^P \phi}{s_{j1} - s_{j1}^P}\left(\frac{\partial s_{j1}}{s_{j1}} - \frac{\partial s_{j1}^P}{s_{j1}^P}\right) + \frac{\partial \epsilon_{j1}}{\epsilon_{j1}} \tag{11}$$

The relative changes in the optimal rates on food and non-food products are driven by the shape of the product-specific formal food Engel curves. As explained above the log-linearity of the food Engel curve implies that the optimal relative food subsidy will increase with development under fairly general conditions when all varieties are taxed. As long as the IEC for non-food products isn't much steeper than the IEC for food products this result follows through in the presence of an informal sector.

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 mechanisms.

5.1 Calibrated government and household preferences

We calibrate the optimal commodity tax rates defined in expressions (5) to (8) using the observed expenditure shares described in section 3 above. In each country we aggregate the data for each decile of the total expenditure distribution and assume 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.[35] We set the own-price elasticity to -1 for all goods at baseline, and allow the price elasticity to vary with the size of the formal sector as a robustness check. 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.

5.2 Results: optimal tax policy over the development path

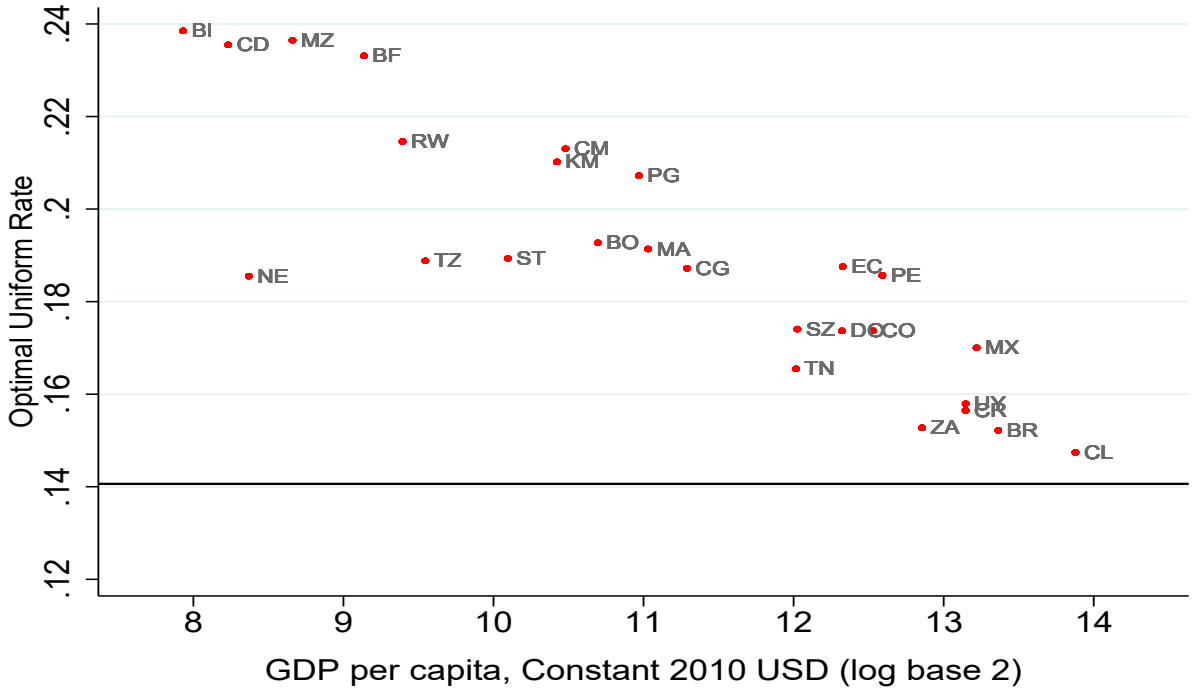
Figure 5 presents the optimal uniform tax rate for each country in our core sample, as a function of economic development. The black line indicates the optimal uniform rate obtained under the assumption that there is no informal sector (both varieties can be taxed) – expression (5) above. This does not vary with development, as expected.[36] The red dots are the optimal uniform rates obtained when we assume informal varieties are untaxed – expression (7). We see that these optimal rates are decreasing with economic development, as predicted by the model: from close to 24% in the poorest countries in our sample where most household expenditure is informal, to just above 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.

We then look at the optimal tax rates on food and non-food products – the optimal version of the commodity tax schedules used throughout the world, which set a different rate on food products. 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 (a counterfactual world without an informal sector – see expression (6) above). A ratio equal to one would indicate that the optimal policy is

[35] 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.

[36] Allowing for different levels of inequality introduces variations in the optimal rate even in the absence of an informal sector, see Appendix Figure A.7.

Figure 5: Optimal uniform rates over the development path



This figure shows the calibrated optimal uniform tax rate with an informal sector on countries GDP per capita. Optimal tax rates are measured using equation (7). The black horizontal lines shows the optimal rate without an informal sector. All optimal rates hold inequalities constant across countries.)

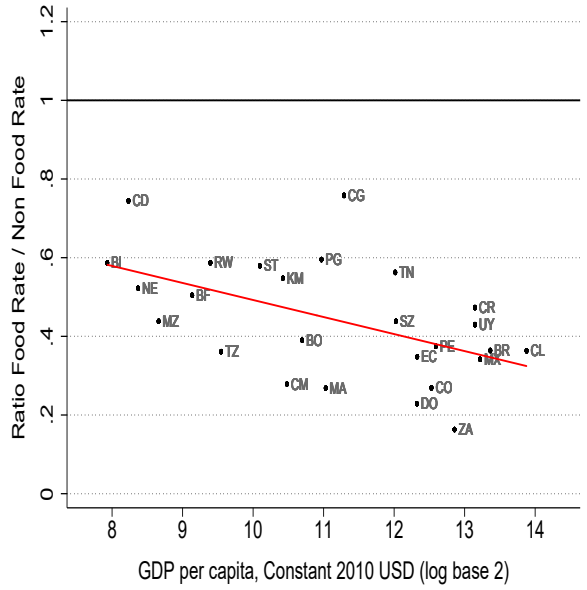
to tax all products uniformly. The ratio is less than one in all countries, indicating that subsidizing food relative to non-food products is always optimal in the absence of an informal sector – as expected given the downward-sloping food Engel curves described above. On average across all countries 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, indicating that the optimal rate on food is higher in those countries. As explained above, this can be explained by the approximate linearity of food Engel curves with respect to log expenditures which makes subsidizing food products less attractive in poorer countries.[37]

Figure 6, panel (b) presents ratios of optimal rates on food relative to non-food products under the more realistic assumption that informal varieties cannot be taxed. We see that the existence of an informal sector increases the ratios in most countries, and hence increases the optimal relative rate on food. This is particularly true in low-income

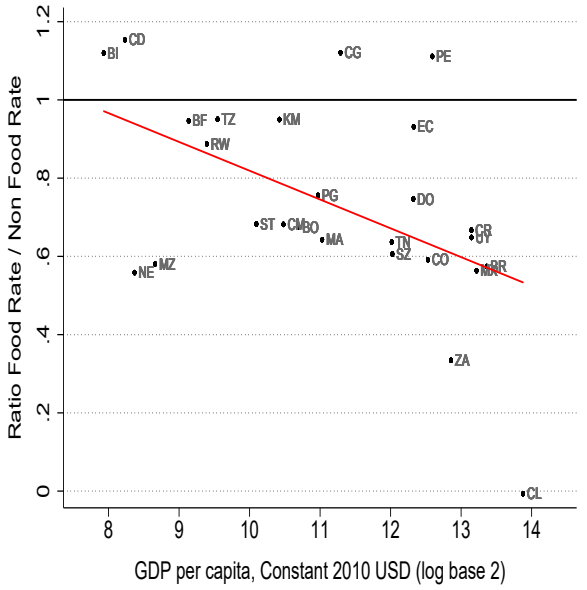
[37]This result would not hold for very high shares of food in total expenditures. In simulations we find that for food shares of 60% and above the relative food subsidy decreases with development. In practice only the very poorest countries in the world have food shares above 60%, as shown in Figure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

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

(a) Modern & Traditional Taxed



(b) Only Modern Taxed



This figure plots the calibrated relative food subsidy rates on countries GDP per capita under two scenarios. In panel (a) we assume that both the modern and traditional sector can be taxed. In panel (b) we assume that only the modern sector can be taxed. The red line is the fitted line. The relative food subsidy corresponds to the ratio of the optimal rate on non-food products on the optimal rate on food product. The optimal tax rates are measured following equation (6).

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. In these countries subsidizing food relative to non-food is simply not optimal.

The discussion thus far has ignored the possibility that the efficiency cost of taxation could change over the course of economic development as the informal sector shrinks. As explained above one possibility is that the efficiency cost falls with development as the informal sector shrinks and it becomes harder for households to substitute away from formal varieties to informal varieties when taxes increase. In Appendix Figure A.6 we plot optimal uniform rates obtained when we allow the price elasticity of demand to increase (in absolute value) with the size of the informal sector. We find that this decreases optimal rates in poorer countries, as expected, but that optimal rates are still decreasing with economic development when we assume that the elasticity governing the efficiency cost of taxation is 40% higher in our poorest country than in our richest country.

Appendix Figures A.7 and A.8 finally present optimal tax results when we allow

for different levels of inequality in each country. We see that optimal uniform rates differ across countries even when both varieties are assumed taxed when inequalities are introduced (more inequalities imply a higher tax rate, see expression (5) above), but the key message remains: when the existence of the informal sector is taken into account the optimal level of consumption taxes falls over the development path. Results regarding optimal rate differentiation are similarly unaffected.

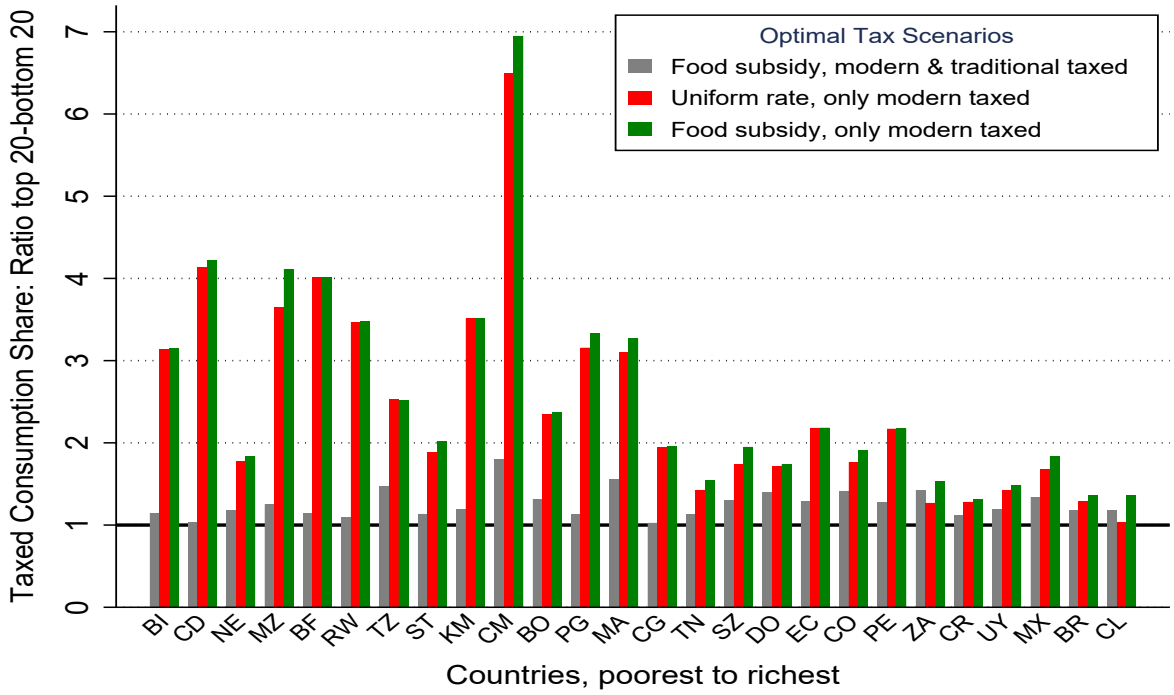
5.3 Progressivity of Optimal Consumption Taxes Across Countries

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 an effective tax rate that is 3 times higher than that of the bottom 20% on average. Taking into account patterns of informal sector consumption thus leads us to findings that differ greatly 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 moreover smallest in the poorest countries, in which optimally

Figure 7: Consumption Tax Progressivity Across Countries



This figure shows the progressivity of optimal consumption taxes (measured as the effective tax rate of the top 20% richest households on the bottom 20%) in each country under three different scenarios. The first scenario, in grey corresponds, allows for different rates on food vs non-food products, but assumes no informal sector. The second scenario, in red, only allows for a uniform tax rate on all products, but takes into account the informal sector. The third scenario in green, allows differentiated rates on food and non-food products with an informal sector. Countries are sorted in increasing order of their per capita GDP.

differentiating tax rates achieves very little 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.9 and A.10 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 using them to measure savings. A few countries collect instead dedicated consumer finance surveys aimed at directly measuring savings. Such data from the US indicates a range of savings from 0% for the bottom quintile to 12% for the top quintile.[38] We plug these distributional savings rates in our model, which makes the taxable base (consumption) smaller than income for rich households while it stays the same for poor 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 8 assesses the progressivity of consumption taxes when including distributional savings, equivalent to the main figure 7. Redistribution falls on average by 8% when considering savings: for example in the scenario with an optimal uniform rate and an informal sector the effective tax rate on of the top 20 to the bottom 20 fell from 2.6 to 2.4.

Figure 8: Consumption Tax Progressivity with Savings

This figure mirrors figure 7, assuming that savings rate increases with income, from 0 for the bottom quintile to 12% for the top quintile.

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

Pass-through of Taxes in the Informal Sector

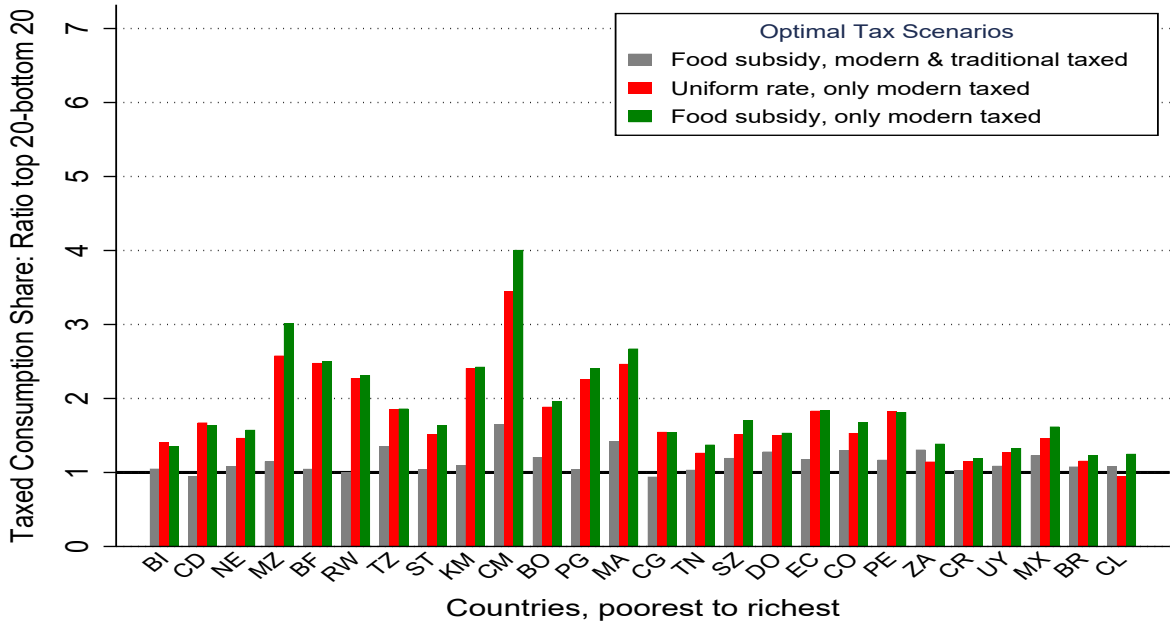
Our analysis thus far ignores the possibility that informal firms might pay some taxes by purchasing intermediate inputs from formal firms which charge commodity taxes on their products (in a VAT system), or through general-equilibrium competition effects. In practice, we know that formal firms partially 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 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.[39] Results are presented in Figure 9: 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.[40]

[39] 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.

[40] 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).

Figure 9: Consumption Tax Prog., 10% Informal Sector Pass-through



This figure mirrors figure 7, under the assumption that informal stores pay taxes on a share of their input purchases.

6 Conclusion

In this paper we study how consumption patterns vary with income within and across countries and derive implications for the design and redistributive potential of consumption taxes. We consider two channels through which consumption taxes can redistribute across households: the rate-differentiation channel, and the novel channel of informal consumption. We find that the informal expenditure share is decreasing with income in our core sample of 30 countries, and that Informality Engel Curves remain downward-sloping when controlling for narrow product choices and household location, used as a proxy for market access. We combine our data with a model of commodity taxation in the presence of an informal sector to calibrate optimal tax policies. We find that the informal sector makes consumption taxes progressive: setting a uniform rate on all products leads to the richest quintile paying effective tax rates that are on average 2.5 times higher than those faced by the poorest quintile. Progressivity is highest amongst low and lower-middle income countries. 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 subsiding food relative to non-food products only marginally increases progressivity, and is particularly ineffective in low income countries.

Our findings have clear policy implications. First, they imply that full enforcement of tax policies – eradicating the informal sector – may not necessarily be optimal, as informal expenditure patterns make consumption taxes progressive. This does not mean that policies aiming to reduce the size of the informal sector should be abandoned. Rather, our results caution that they could have adverse distributional consequences, as they may shift the burden of consumption taxes towards the poor. Second, our results have stark implications for the use of reduced tax rates on food products, observed in most countries in the world. We find that differentiating commodity tax rates has a limited redistributive impact; in some low-income countries setting a lower rate on food simply cannot be justified on equity grounds once informal expenditures are taken into account.[41] In practice, removing subsidies on food and other necessities is often met with fierce political resistance.[42] To be implementable, the removal of subsidies may have to be used to fund transfer programs (Hanna and Olken (2018)).

[41] This could still be justified on equity grounds if households are much more likely to turn to the informal sector when rates on food increases than when rate on non-food products increase.

[42] 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).

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.

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.

Kopczuk, W. (2001): "Redistribution when avoidance behavior is heterogeneous," *Journal of Public Economics*, 81, 51–71.

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.

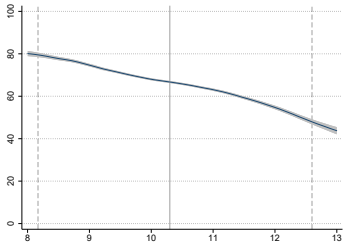
Ulyssea, G. (2018): "Firms, informality, and development: Theory and evidence from Brazil," *American Economic Review*, 108, 2015–47.

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.

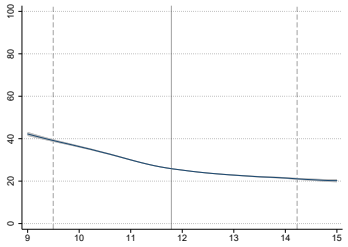
Working, H. (1943): "Statistical laws of family expenditure," *Journal of the American Statistical Association*, 38, 43–56.

A Additional Figures and Tables

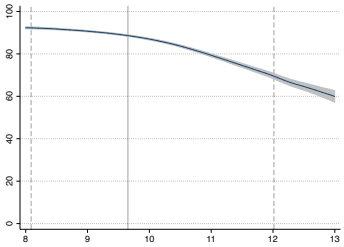
Figure A.1: Informality Engel Curves



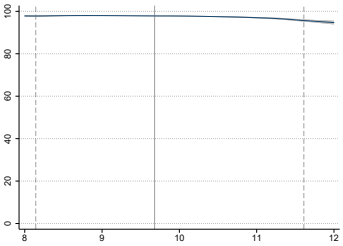
(a) Bolivia



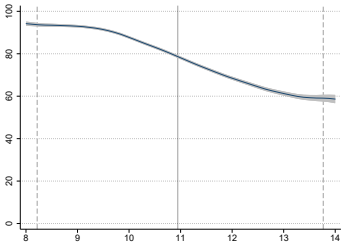
(b) Brazil



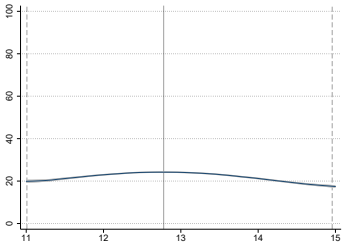
(c) Burkina Faso



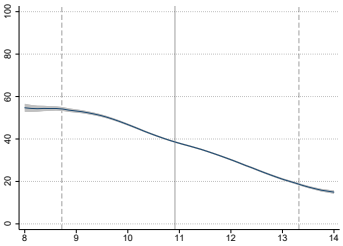
(d) Burundi



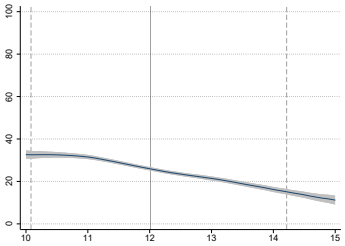
(e) Cameroon



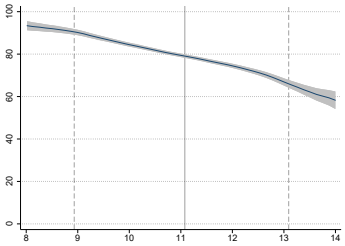
(f) Chile



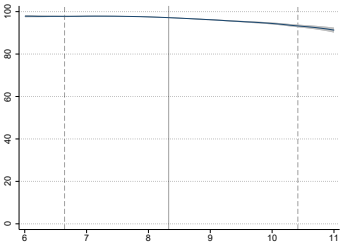
(g) Colombia



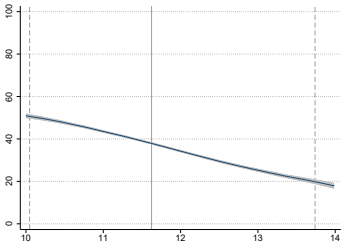
(h) Costa Rica



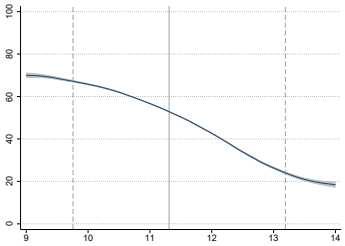
(i) Comoros



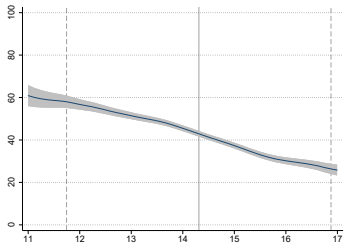
(j) Dem. Rep. of Congo



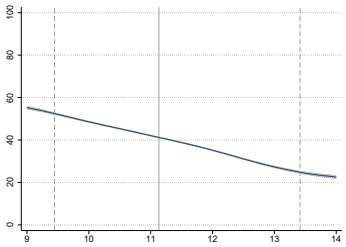
(k) Dominican Rep.



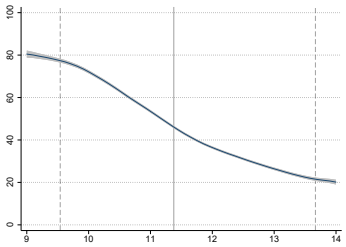
(l) Ecuador



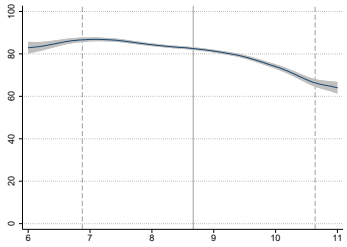
(m) Eswatini



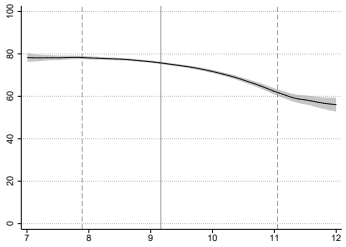
(n) Mexico



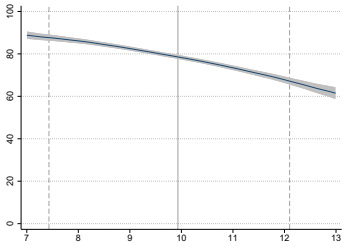
(o) Morocco



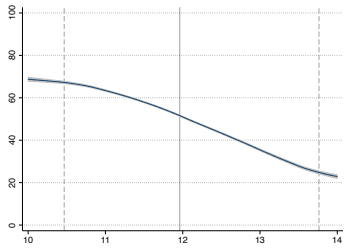
(p) Mozambique



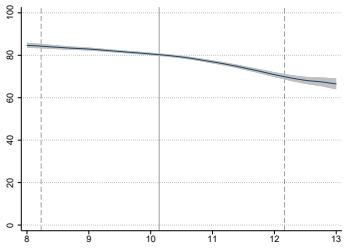
(q) Niger



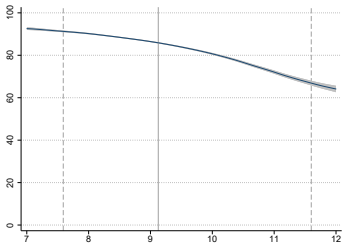
(r) Papua New Guinea



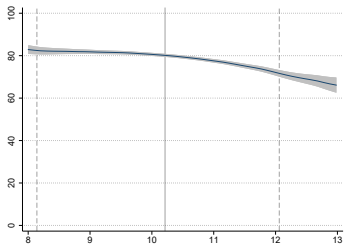
(s) Peru



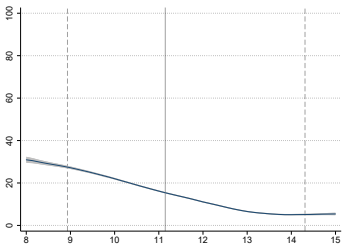
(t) Rep. of Congo



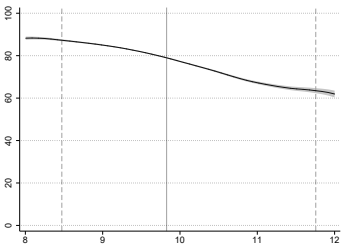
(u) Rwanda



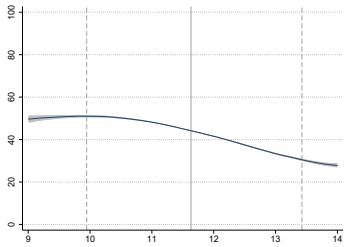
(v) Sao Tome



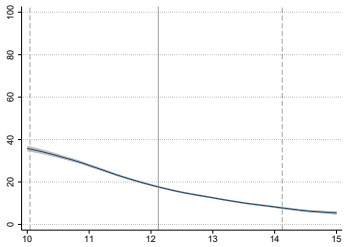
(w) South Africa



(x) Tanzania



(y) Tunisia

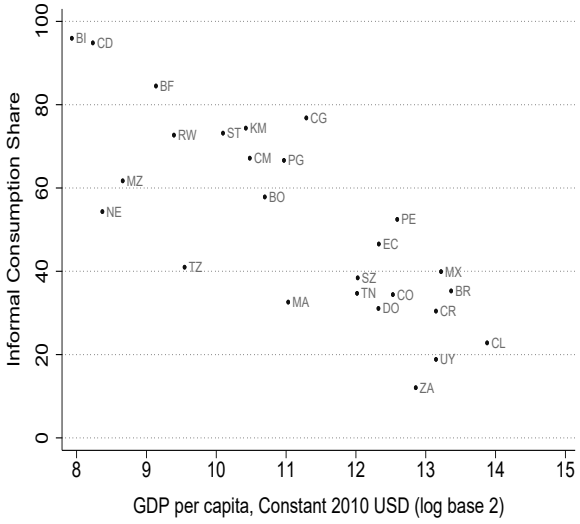


(z) Uruguay

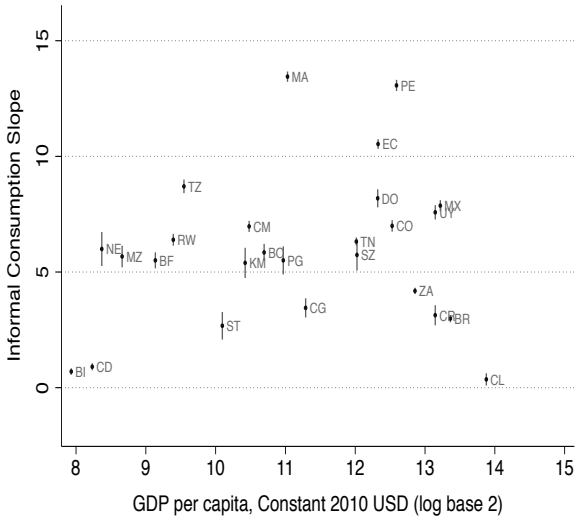
Local polynomial fit of the informality Engel curves for DR Congo and Mexico as presented in section 3.1. The informality Engel curves are defined as the share of informal consumption over the per capita expenditure of households. Expenditure is measured in log base 2, such that a one unit increase on the horizontal axis corresponds to a doubling of household's expenditure. 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 Scenarios for Informal Consumption

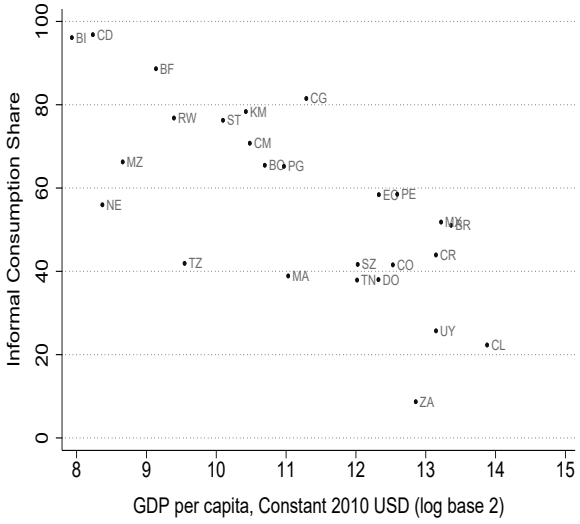
(a) Level (Probability Scenario)



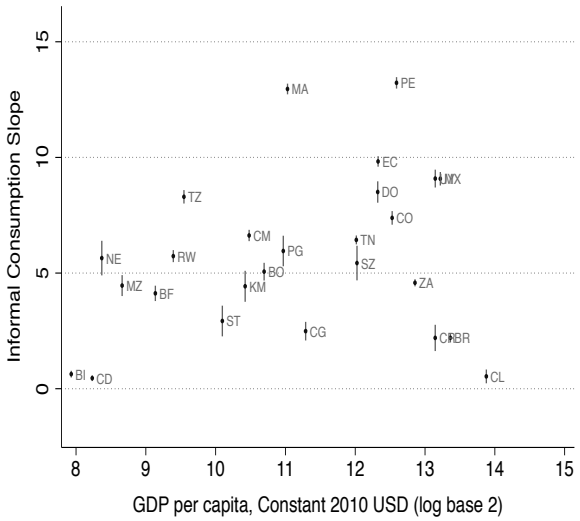
(b) Slope (Probability Scenario)



(c) Level (Specialized to Informal Scenario)



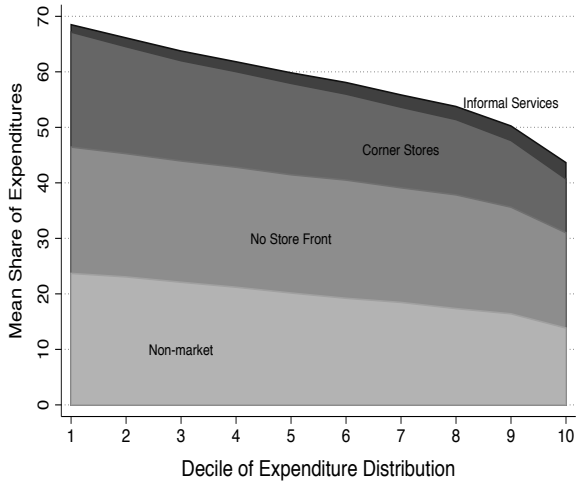
(d) Slope (Specialized to Informal Scenario)



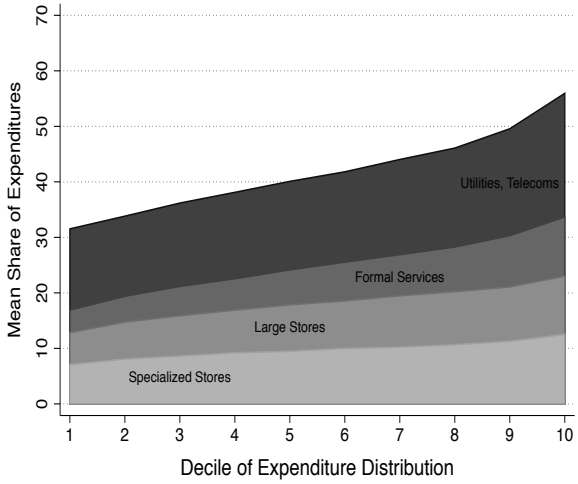
Panel (a) and (b) shows informal consumption as a share of total consumption, on per capita GDP. Panel (b) shows the slope of informal consumption with total per person household expenditure, on per capita GDP. The bars correspond to the 95% confidence interval of the slope coefficient. The slope measures the drop in informal consumption for a doubling of households' expenditure, within country. GDP per capita is in constant 2010 USD, and transformed in log base 2 (Source: World Bank WDI)

Figure A.3: Average Consumption of Each Decile by Place of Purchase

(a) Informal Places of Purchase

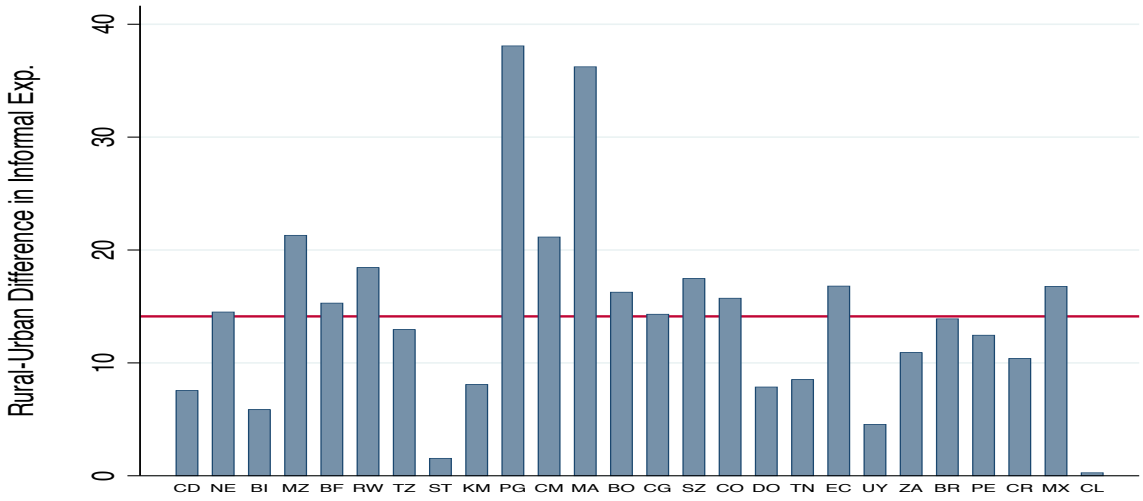


(b) Formal Places of Purchase



This figure shows the average consumption of each decile (across countries) by place of purchase. Panel A shows the places of purchase classified as informal and Panel B shows the places of purchase classified as formal.

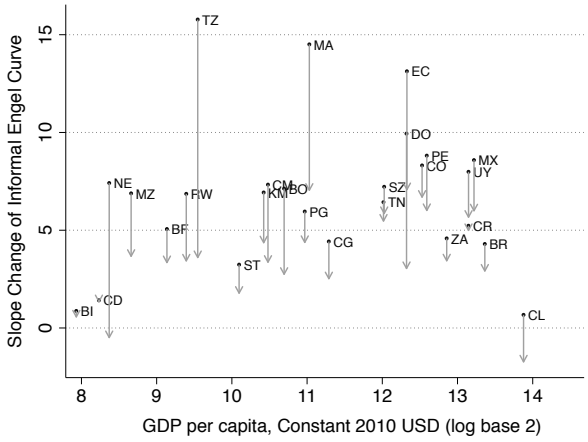
Figure A.4: Rural-Urban difference in Informal Expenditure Shares



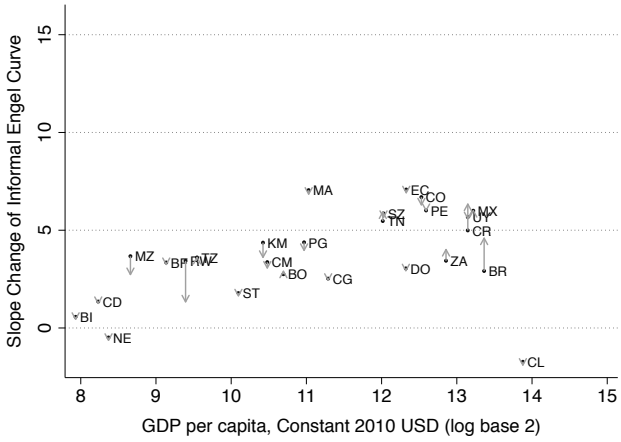
The figure shows the difference between the rural and urban share of informal expenditure across countries. The red line shows the sample average of 13%.

Figure A.5: Change in IEC Slopes from Product Composition

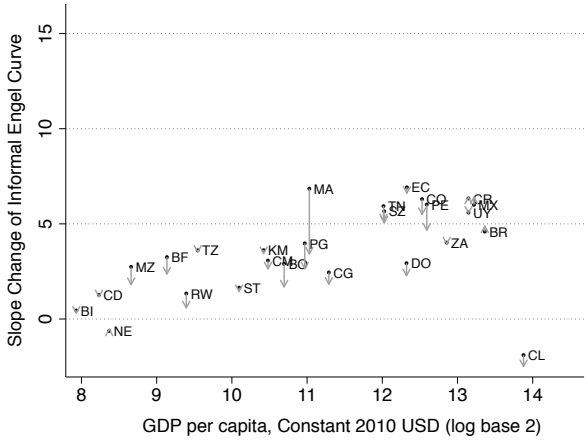
(a) No product controls to COICOP 2



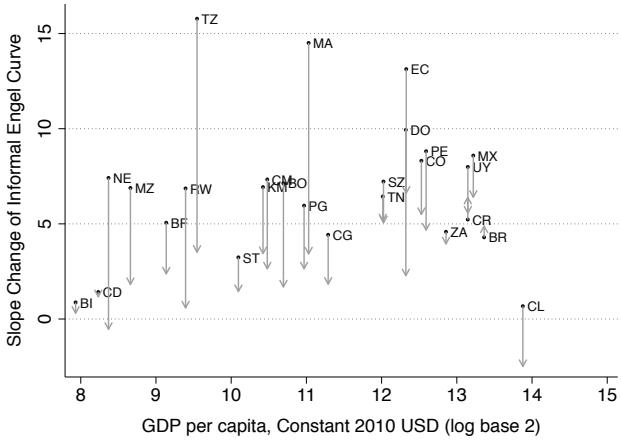
(b) COICOP 2 to COICOP 3



(c) COICOP 3 to COICOP 4



(d) No product controls to COICOP 4



This figure shows the change in informal Engel curves' slopes when controlling for increasingly narrow product codes.

Table A1: Average Informal Consumption Slopes - Scenario Proba

Specification:	**Main**		**Geography**		**Product Codes**			**All**
Avg. of 26 Countries	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Slope	6.1	6.5	5.6	4.9	3.3	3.3	2.7	2.4
Confidence Interval	[5.7,6.5]	[6.1,6.9]	[5.1,6.0]	[4.4,5.3]	[3.0,3.6]	[3.0,3.5]	[2.5,3.0]	[2.1,2.7]
# of p-values < 0.05	26	26	26	24	24	24	24	24
R^2 adjusted	0.19	0.20	0.25	0.41	0.55	0.59	0.59	0.63
Household Characteristics		X	X	X	X	X	X	X
Urban/Rural			X					
Census Blocks				X				X
COICOP 2-dig					X			
COICOP 3-dig						X		
COICOP 4-dig							X	X

Table A2: Average Informal Consumption Slopes - Scenario Robust

Specification:	**Main**		**Geography**		**Product Codes**			**All**
Avg. of 26 Countries	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Slope	5.8	6.1	5.2	4.5	2.9	3.0	2.5	2.3
Confidence Interval	[5.4,6.2]	[5.7,6.6]	[4.8,5.7]	[4.0,5.0]	[2.6,3.2]	[2.7,3.3]	[2.3,2.7]	[1.9,2.6]
# of p-values < 0.05	26	26	26	24	24	24	22	22
R^2 adjusted	0.17	0.18	0.22	0.38	0.54	0.59	0.60	0.62
Household Characteristics		X	X	X	X	X	X	X
Urban/Rural			X					
Census Blocks				X				X
COICOP 2-dig					X			
COICOP 3-dig						X		
COICOP 4-dig							X	X

Table A3: 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.

Figure A.6: Optimal uniform rates over the development path

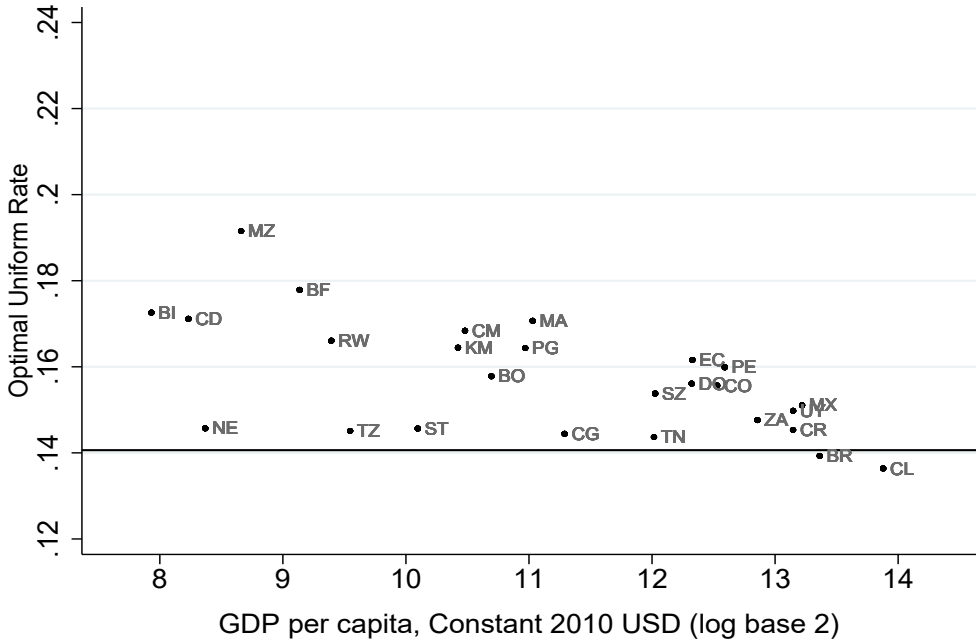
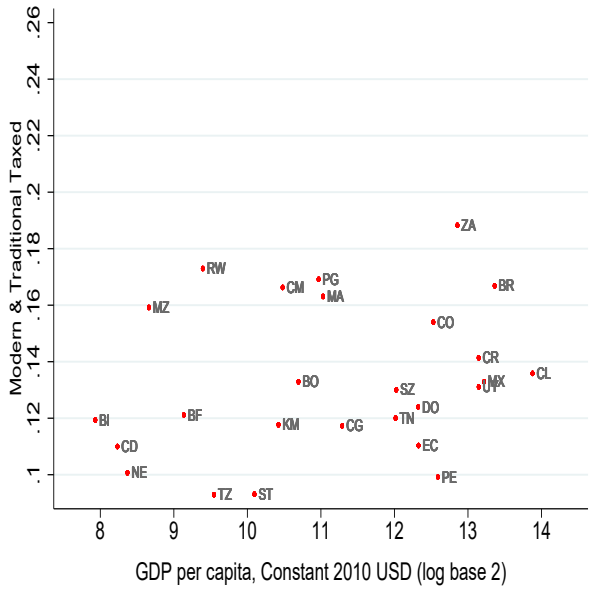


Figure A.7: Optimal uniform rates over development with inequalities

(a) Modern & Traditional Taxed



(b) Only Modern Taxed

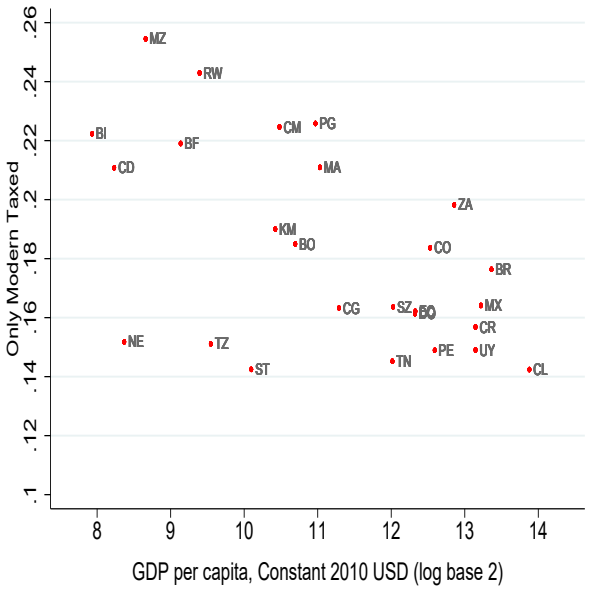
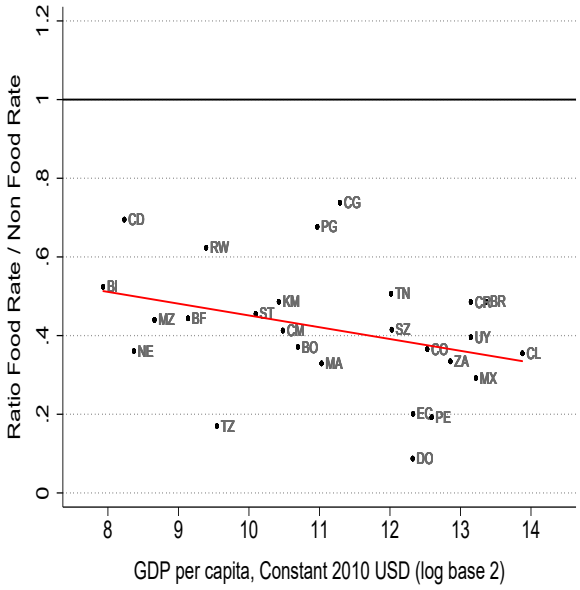


Figure A.8: Optimal Rates on Food and Non-Food, with inequalities

(a) Modern & Traditional Taxed



(b) Only Modern Taxed

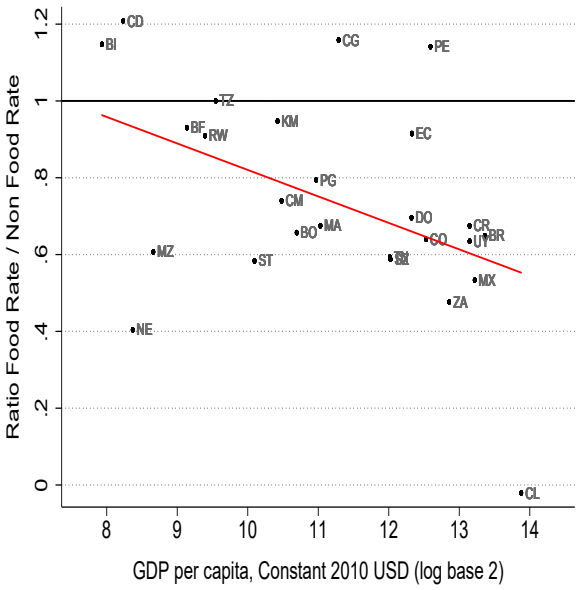


Figure A.9: Consumption Tax Progressivity, Robust Scenario

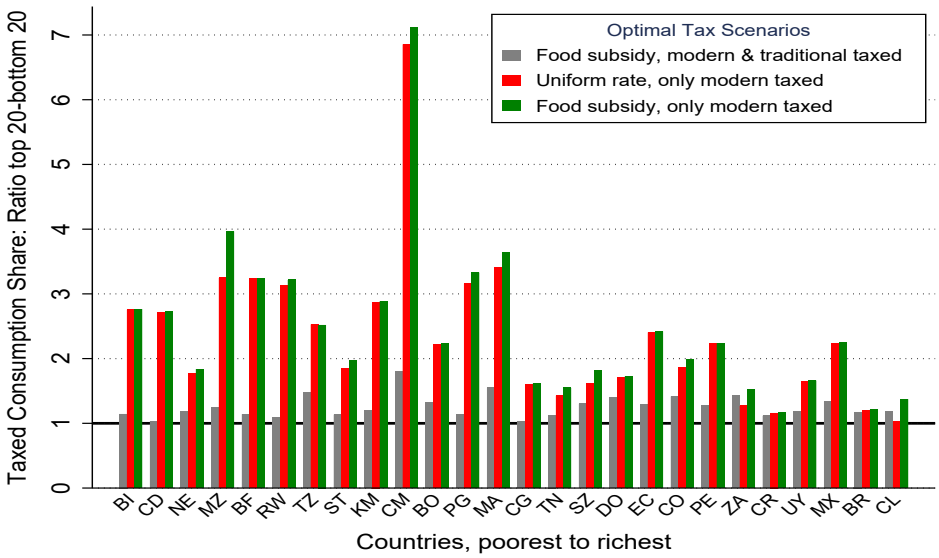
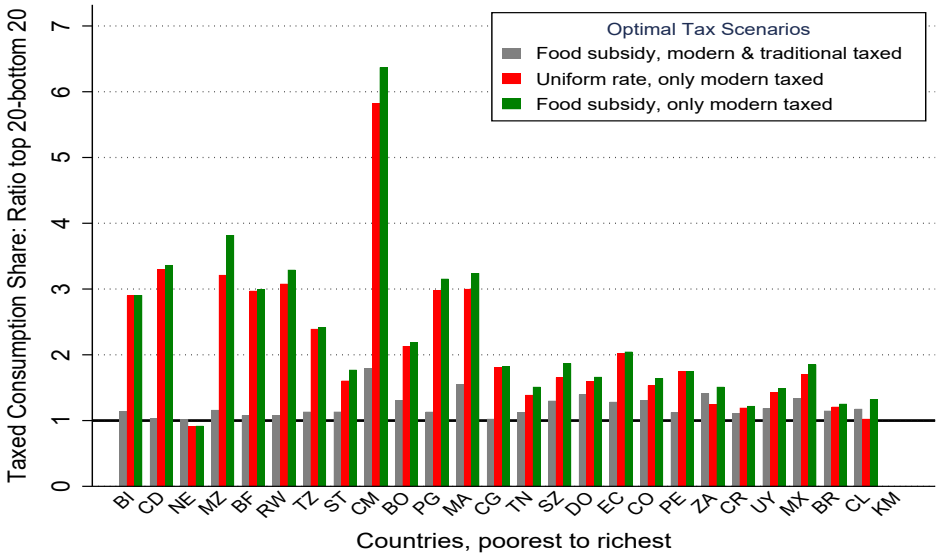


Figure A.10: Consumption Tax Progressivity, Probabilistic Scenario



B Data Appendix

B.1 Dataset Construction

Inclusion Criteria

The dataset is constructed based on 25 nationally representative household expenditure surveys. We retained surveys which satisfied the following five criteria:

1. The household expenditure survey is nationally representative and dates from the 21st century.
2. The expenditure survey includes a variable on the place of purchase: for each consumption item there is information on where each consumption item was purchased. The place of purchase names should have enough details for reliable classifications into formal or informal sectors, as further outlined in Appendix B.2. In particular, there should be enough detail to classify at least one place of purchase to the formal sector and at least one place of purchase to the informal sector.
3. The expenditure modules in the survey are structured as open consumption diaries, rather than pre-fill diaries for specific products.
4. The place of purchase variable rarely contains missing values; particularly for food, clothing, and household goods product categories.
5. We were able to obtain access to the data.

Table B1 lists the countries in the sample, with summary statistics and the structure of each survey's expenditure modules.

Data Sources and Coverage

We obtained the data principally from two sources: (i) the World Bank Microdata Library and (ii) national statistical agencies, with the exception of South Africa, for which the data came from the University of Cape Town (refer to column 3 of Table B1). The first step for accessing data started with the restricted-access World Bank Microdata Library, where we examined National Household Income and Expenditure, Living Standards, and Budget Surveys to see if criteria (1)-(4) above were satisfied. The datasets that satisfied such criteria ranged in their ease of access: while some countries' survey microdata

Table B1: Household Expenditure Surveys

			Household surveys						Structure of expenditure modules					
Country	**Survey-Year**	**Source**	**# HH**	**Avg HH Size**	**Avg Exp/HH** Cst. 2010 USD	**# Exp per HH**	**# PoP**	**Urban**	**# Modules**	**Module Freq**	**Durables**	**Self Prod.**	**Comments**	**Product Code**
Bolivia	Encuesta de Continua de los Hogares 2003-2004	Stat. Office	9149	4.18		49.42	24	60.7%	3	Daily, Monthly, Quarterly, Yearly	Included	Included		Country-specific
Brazil	Pesquisa de Orçamentos Familiares	IBGE website	56049	3.3	3891.96	47.99	753	84.4%	8	Weekly, Monthly, Quarterly, Yearly	Included	Included	Food, non-food modules separated	Country-specific
Burkina Faso	L'Enquête Intégrale sur les Conditions de Vie des Ménages	Stat. Office	8404	6.73	374.05	72.02	45	29.3%	1	Yearly	Included	Included		COICOP
Burundi	Enquete sur les conditions de vie des menages	MicroData Lib	6681	4.8	396.26	89.76	13	9.0%	23	Bi-Weekly, quarterly, Biannual, Yearly	Included	Included	Food, transport, clothing, comm. Modules separated	COICOP
Cameroon	Quatrieme Enquete Camerounaise Aupres des Menages (ECAM)	MicroDataLib	10303	4.57	4047.58	95.76	17	44.5%	1	Daily	Included	Included		COICOP
Chile	Encuesta de Presupuestos Familiares - EPF	Stat. Office	15237	3.29	6872.47	129.2	17	55.8%	1	Monthly	Included	Unavailable		COICOP
Colombia	Encuesta Nacional de Ingresos y Gastos - ENIG	Stat. Office	42733	3.82	1571.82	79.65	24	82.1%	5	Monthly, quarterly, yearly	Included	Included		COICOP
Dem. Rep. of Congo	Enquete 1-2-3 sur l'emploi, le secteur informel et les conditions de vie des menages 2004	MicroDataLib	12098	5.29	198.3	106.9	13	16.0%	1	Yearly	Included	Included		COICOP
Republic of Congo	Enquete Congolaise aupres des Menages pour L'Evaluation de la Pauvrete	MicroDataLib	5002	5.12	641.39	84.76	17	63.8%	1	Yearly	Included	Included		COICOP
Costa Rica	Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH)	Stat. Office	5705	3.36	4605.61	67.48	41	73.2%	1	Monthly	Included	Included		COICOP
Dominican Republic	Encuesta Nacional de Gastos e Ingresos de los Hogares (ENIGH) 2006-2007	Stat. Office	8363	3.67		89.07	88	67.6%	3	Weekly, Monthly, Quarterly, Yearly	Included	Included		COICOP
Ecuador	Encuesta Nacional de Ingresos y Gastos de Hogares Urbanos y Rurales	MicroDataLib	39617	3.88	1922.59	88.65	75	68.0%	7	Daily, Weekly, Biweekly, Monthly, Quarterly, Biannual, Yearly	Included	Included	Food, non-food modules separated	COICOP
Eswatini	Household Income and Expenditure Survey	MicroDataLib	3167	4.54		43.9	13	37.4%	2	Daily, Monthly, Yearly	Included	Included		Country-specific
Mexico	Encuesta Nacional de Ingreso-Gasto de los Hogares (ENIGH)	Stat. Office	19459	3.79	2271.83	57.36	19	64.5%	1	Quarterly	Included	Included		Country-specific
Morocco	Enquête nationale sur la consommation et les dépenses des ménages	MicroDataLib?	14243	5.89	1679.25	87.5	47	61.6%	17	Weekly, Monthly, Two-months, quarterly, Yearly	Unavailable	Included	Food, non-food modules separated	Country-specific
Mozambique	Inquérito ao Orçamento Familiar - IOF	MicroDataLib?	10809	4.67	662.42	48.75	6	28.9%	6	Daily, Monthly, Yearly	Included	Included	Self-consumption, transfer modules separated	COICOP
Niger	Enquête Nationale sur le Budget et la Consommation des Ménages 2007, Troisième Enquête	MicroDataLib	3980	6.36	333.16	61.05	15	17.1%	6	Daily, Weekly, Monthly, Quarterly, Biannual, Yearly	Unavailable	Included	Food, non-food modules separated	COICOP
Papua New Guinea	Household Income and Expenditure Survey 2009	MicroDataLib	3811	5.13	1002.03	111.2	6	11.3%	1	Bi-Weekly, quarterly, Biannual, Yearly	Included	Included		COICOP
Peru	Encuesta Nacional Hogares (ENAHO)	Stat. Office	43530	3.88	1344.37	56.43	41	76.8%	8	Yearly	Included	Included	Food, transport, clothing, rec., comm. modules separated	Country-specific
Rwanda	Integrated Household Living Conditions Survey	MicroDataLib	14419	1	1046.43	89.53	11	17.1%	8	Monthly, Yearly, 10 weeks	Included	Included	Food, non-food modules separated	COICOP
Sao Tome & Principe	Inquerito Orcamento Familiar	MicroDataLib	3145	3.82		105.9	21	68.1%	3	Daily, Bi-weekly, Monthly, Quarterly, Semi-annual, Yearly	Included	Included		
South Africa	Income and Expenditure Survey	DataFirst	25325	3.85	3561.11	44.2	6	67.3%	1	Yearly	Unavailable	Unavailable		COICOP
Tanzania	Household Budget Survey	MicroDataLib	10168	5.01	316.6	185.4	13	21.9%	2	Monthly	Unavailable	Included	Food, non-food modules separated	COICOP
Tunisia	Enquete Nationale sur le Budget, la Consommation et le Niveau de Vie des menages 2010	Stat. Office	11281	4.34		139.1	9	67.6%	1	Daily	Included	Included		Country-specific
Uruguay	Encuesta Nacional de Gastos e Ingresos de los Hogares (ENGIH)	Stat. Office	7042	2.96	2854.59	77.47	39	84.9%	1	Monthly	Unavailable	Unavailable		COICOP

were openly accessible to the public for download, others were licensed and required applications through the World Bank which would then sometimes contact the respective country's national statistical agency for approval.

If the applicable data was inaccessible through the World Bank Microdata Library, we searched the country-specific statistical agency websites. For some countries, this simply involved downloading the micro-data off of the website, while for others, we made formal data requests. This second step came with varying degrees of success: we obtained data through this route for most Latin American countries, however this process was not always successful elsewhere.

The countries that ultimately satisfied the criteria for inclusion span four regions of the world, with the greatest number of countries concentrated in Sub-Saharan Africa and Latin America and the Caribbean, as detailed in Table B2. Unfortunately we were not able to find any East Asian countries, with a question on the place of purchase in their household expenditure surveys. In Eastern Europe and Central Asia surveys sometimes ask this question, but the number of options for place of purchase are minimal, and the type of stores is often missing, thus not satisfying inclusion criteria 2 and 4.

Table B2: Regional Survey Representation

Region	**# Countries**	**Pop. of Surveyed Countries** (Millions)	**Total Pop.** (Millions)	**Proportion of pop.**
Sub-Saharan Africa	13	372	1078	34%
Middle East & North Africa	2	47	449	11%
Eastern Europe and Central Asia	0	0	918	0%
Latin America and Carribean	8	484	641	76%
East Asia & Pacific	1	7	2328	0.3%

While there were a number of surveys that outwardly appeared to satisfy all of the four main selection criteria (namely, Argentina, Belarus, Gambia, Ghana, and Turkey), we were constrained by issues of data access. Table B3 further details countries that were considered for inclusion in our sample, but were ultimately discarded for failing to satisfy any of (1) - (5) from the first page.

Table B3: Discarded Household Expenditure Surveys

Country	Survey	Year	Why Discarded
Argentina	Encuesta Nacional de Gastos de los Hogares	2013	(5) data access constraints
Armenia	Integrated Living Conditions Survey	2016	(4) PoP missing
Belarus	Household Sample Survey	2010	(5) data access constraints
Bosnia & Herzegovina	Household Budget Survey	2007	(3) PoP asked as purchasing habit
Chad	Enquete sur la Consommation et le Secteur Informel	2003	(2) PoP vague
El Salvador	Encuesta de Hogares de Propositos Multiples	2010	(4) PoP missing; limited consumption categories
The Gambia	Integrated Household Survey	2003	(5) data access constraints
Ghana	Living Standards Survey	2006	(5) data access constraints
Guatemala	Encuesta Nacional sobre Condiciones de Vida	2000	(4) PoP missing; limited consumption categories
Mauritius	Household Budget Survey	2012	(4) PoP missing
Montenegro	Household Budget Survey	2009	(4) PoP missing
Nicaragua	Encuesta Nacional de Hogares sobre Medicion de Nivel de Vida	2014	(3) PoP asked as purchasing habit; (5) limited consumption categories
Serbia	Living Standards Measurement Survey	2007	(4) limited consumption categories
Tajikistan	Household Budget Survey	2016	(4) limited consumption categories
Turkey	Household Income and Consumption Expenditures Survey	2009	(5) data access constraints
Ukraine	Household Living Conditions Survey	2012	(5) data access constraints; (4) limited consumption categories

Consumption Module Structure

While each country satisfies (1) to (5), consumption modules are structured differently across countries. Table B1 details their structure and how they can differ - we provide below a summary:

- **Number and frequency of modules**
  - The number of consumption modules range from 1 to 17 modules across countries in the sample. For example, while Costa Rica had one consumption module, Morocco had 17 consumption modules. Modules may also vary based on frequency of expenditures (daily, weekly, monthly, quarterly).
- **Durables**
  - Durable items, which do not have to be purchased frequently (e.g. furniture, motor vehicles, etc.), were included whenever available.
- **Self-production**
  - Self production was included as a "place of purchase" whenever available. In some cases, it was pre-coded in the raw PoP variable, while in others we added it as a PoP based on other variables, such as "mode of acquisition," which included the response of "self/home production."
- **Product codes**
  - Modules have product codes for each consumption item which either follow the official United Nations Classification of Individual Consumption According to Purpose (COICOP) standard or a nationally-specific product classification

scheme - which we harmonized with COICOP codes through a detailed cross-walk.

In spite of the diversity of frequency and number of modules, all surveys are nonetheless structured as open-fill diaries of consumption, in satisfaction of criteria (3) under Section A.1. Figure B.1 obtained from South Africa's 2010-2011 Income and Expenditure Survey serves as an illustrative examples of directions for filling out the household diary, as well as what a typical "open fill" diary of consumption looks like.

INCOME AND EXPENDITURE SURVEY 2010/11
GUIDELINES FOR FILLING IN DAILY ACQUISITIONS

TO BE RECORDED DAILY IN:

Form 1 (pages 6-11)

- All food and non-food items purchased by any member of the household, such as bread, milk, rice, furniture, electric appliances, wood, etc. for the household's consumption as well as to give away as a gift or maintenance.

  Households to record the full price of the item at the time of acquisition.

- All items acquired by the household without paying for them, such as items from own production, (e.g. from own garden or kraal) or from nature (e.g. items from hunting, fishing and gathering of vegetables).
  To be recorded when the item was consumed by the household.
- All items received as gifts or maintenance from someone who is not a household member.

  Households to record the estimated value of the item acquired.

Form 2 (pages 12-13)

- All food and beverages (such as hamburgers, fruit, soft drinks, etc) acquired at restaurants, canteens and other food outlet.
- "Small" acquisitions (such as cigarettes, newspapers, sweets, soft drinks, etc) by individual household members.

  To be recorded when purchased/acquired.

Form 3 (pages 14-15)

- All payments made by household members for services such as car insurance, telephone bills, DStv, bus and taxi fares, etc. whether paid for when receiving the service or paid monthly.

  Households to record the amount paid for the service.

SEE ALSO THE EXAMPLES ON THE NEXT TWO PAGES

NOT TO BE RECORDED IN THE DIARY:

- Items purchased for business purposes.

HOW TO COMPLETE THE DIARY
Instructions to the main respondent:

- The household member who knows the most about the household's acquisitions should take responsibility for completing the Household Diary.
- A notebook should be issued to individual household members to be carried when away from home, in order to record acquisitions which take place during the day. After completion of a week, tear out the used pages from the notebook and put inside the envelope provided, seal the envelope and give it to the interviewer at his/her next visit.
- Ask each member of the household about any transactions for the day which have not been recorded in the Household Diary.
- Please use the checklist as a reminder regarding items which are easily forgotten.
- ***In order to ensure complete recording, please keep receipts from all purchases.***

FORM 1

ITEMS *PURCHASED* BY THE HOUSEHOLD, OR *RECEIVED* AS GIFTS OR MAINTENANCE, OR *CONSUMED*

Note: *Items purchased from restaurants, fastfood outlets, etc. should be recorded in Form 2*

PAYMENTS FOR *SERVICES* (e.g. *taxi fares, airtime, telephone bills, internet and DStv subscriptions,*

FOR OFFICE USE					Value		
I	COICOP	CHK	Day	Description	Rand	Cent	
			0 1	Brown bread	8	5 0	101
			0 1	Long life full cream milk	7	0 0	102
			0 1	Newspaper	4	5 0	103
			0 1	Milk chocolate	1 1	9 5	104
			0 2	Rice	1 7	9 5	105
			0 5	Women's skirt	2 9 9	9 5	106
			0 5	Fresh potatoes	7	0 0	107
			0 6	Fish caught by household	2 0	0 0	108

	Source *1 = Shop* *2 = Own business* *3 = Own produce (from own garden, livestock, etc.)* *4 = Gift or maintenance received or from nature (hunted, gathered, fished, etc.)* *5 = Other*	**Was this for this household's own consumption?** *1 = Yes* *2 = No, gift given away* *3 = No. maintenance given away*	**Area of purchase** *1 = In a big city (Metro)* *2 = In another urban area (town/city/township)* *3 = In a rural area* 4 = NOT APPLICABLE 5 = DON'T KNOW	**Type of retailer** **Formal sector**: *1 = Chain store*, *2 = Internet*, *3 = Other retailer* **Informal sector**: *4 = Street trading*, *5 = Other* 6 = NOT APPLICABLE 7 = DON'T KNOW	***FOR OFFICE USE*** S	D
101	1	1	1	1		
102	1	1	1	1		
103	1	1	1	3		
104	4	1	2	4		
105	1	1	1	1		
106	1	2	4	6		
107	5	1	2	4		
108	5	1	4	6		

B.2 Assignment of Places of Purchase to Formality Status

This study involves the assignment of a formality status to the places of purchase (PoP) detailed in each line-item of the household consumption diaries. We structured the formality status classification around a number of guiding principles which we detail in this appendix. Although the number and names of the PoP can vary from one country to another, we aimed to limit the number of country-specific choices, and we provide in this appendix a transparent explanation for our choices. We drew upon two sources of information to guide our choices (i) for selected countries we used data from firm censuses to establish a relation between formality and store types or store sizes, and (ii) the International Price Comparison (ICP) project, which builds purchasing power parity indexes and provides a store type classifier for marketed consumption.

At a high level, we use the place of purchase to break down consumption into four categories:

(1) **Non-market consumption** $\Rightarrow$ Assigned to *informal*

(2) **Market consumption, non brick-and-mortar stores** $\Rightarrow$ Assigned to *informal*

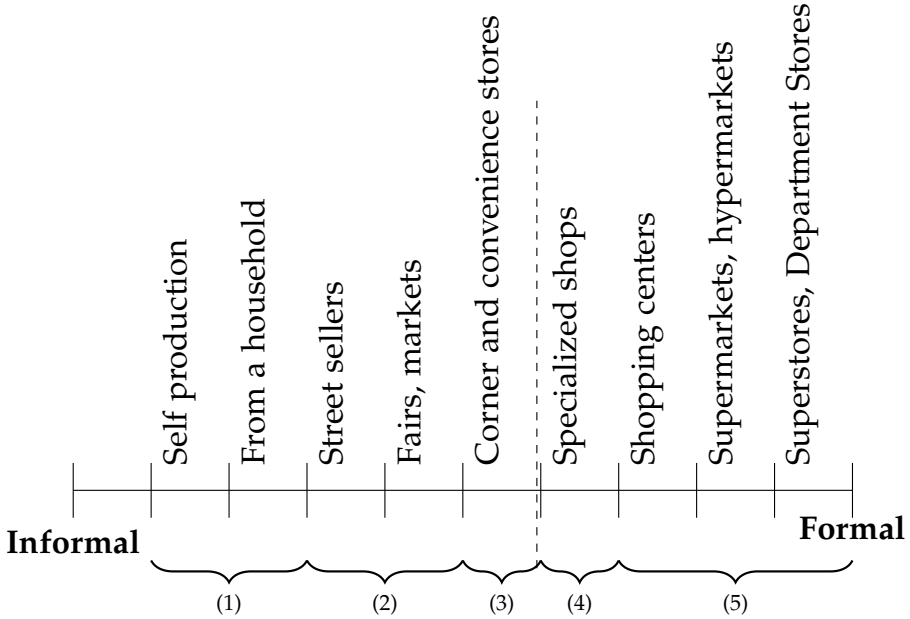
(3) **Market consumption, small brick-and-mortar convenience stores, corner shops** $\Rightarrow$ Depends on scenario. In core scenario (A) corner stores are assigned to *informal*.

(4) **Market consumption, brick-and-mortar specialized shops** $\Rightarrow$ Depends on scenario. In core scenario (A) specialized stores are assigned to *formal*.

(5) **Market consumption, large brick-and-mortar stores** $\Rightarrow$ Assigned to *formal*

Based on the above classification, Figure B.2 displays the most common place of purchase names, with our "formality" ordering.

Figure B.2: High-level Overview of Places of Purchase



Using this general categorization and drawing from the place of purchase definitions developed by the International Price Comparison project,[43] we assign the different places of purchase to categories (1) - (5) above. All PoP classified as (1) Non-market and (2) Market, non brick-and-mortar, are assigned to *informal*. These include self production, transfers between households, street selling, fairs, and markets. The reasoning is that small-scale retailers that operate out of temporary settlements are more likely to be informal and less likely to pay taxes than businesses with permanent, regulated locations. In contrast, all PoP classified as (5) Market consumption from large brick-and-mortar, such as supermarkets, shopping centers, and department stores are assigned to *formal*. These larger institutions, by virtue of their size, number of employees, and establishment size, among other factors, are likely to be registered and pay commodity taxes.

Categories (3) and (4) - Market consumption from small brick-and-mortar stores - is the category for which the formality status might be most context dependent. This category spans corner stores (3) which tend to be small, but also specialized shops (4) which might be more heterogeneous in size. In our central scenario (A) we assign corner and convenience stores to informal and specialized stores to formal. In Appendix A.2 we report results for two alternative scenarios: Scenario (B) where we assign both categories (3) and (4) to informal, and scenario (C), where we assign both categories (3) and (4)

[43] See Table **??** for details on the International Price Comparison project's classification

to formal. Table **??** outlines scenarios (A) to (C), as well as a comparison between our classification and the ICP classification for the most common places of purchases across the 20 countries.

While places of purchases in categories (1)-(5) characterize a large share of expenditure for the countries in our sample, some places of purchase do not fit in this nomenclature and are assigned to category (6) ”other places of purchases”. We list below the most common occurrences, which correspond mostly to services:

(6) **other places of purchase**

We list below some common occurrences:

- public institutions ⇒ Assigned to *formal*
- health, education, and financial institutions ⇒ Assigned to *formal*
- abroad ⇒ Assigned to *formal*
- online ⇒ Assigned to *formal*
- entertainment (hotels, restaurants, shows)
- other undefined services

As can be appreciated from the list above, most of these PoP are likely to be formal and taxable.[44] We assign these PoPs to formal unless there are additional details in the labels of the PoP about their relative size: for example, within health spending, there may be categories called “public health institution,” versus “traditional medicine.” In this example, the latter would be considered as an informal service and the former two would be considered as formal. Such a distinction is particularly common, though not present in the full sample of countries, within entertainment services: we always assign ”restaurant” to formal and ”canteen-truck” or ”bars/cafes” to informal. Besides, although most services mentioned among the category (6) are likely to be formal, for some of them the declared retailer signals the transaction was likely to be informal. Therefore we consider that category (6) can be split in the following way:

(6) **other formal places of purchase** ⇒ Assigned to *formal*

- public institutions

[44]Whether these place of purchases are taxed in practice, is a different matter. For example taxing online purchases is not technically difficult, but tax systems have been slow to adjust, in part due to issues of assessing the correct jurisdiction where the tax should be remitted.

- health, education, and financial institutions

- abroad

- online

(6) informal services (services from an undefined individual, individual transportation services)

(6) formal entertainment (hotels, restaurants, professional cafeterias, school canteen)

(6) informal entertainment (cafés, bars, other types of canteens, take-away)

- other undefined services

We provide further details on these nuanced scenarios in Table B4.

Table B4: Country-Specific Places of Purchase

	BRAZIL				BURKINA FASO			
Sector			new classification	recoded as			new classification	recoded as
Formal	2.70%	restaurant	5: other consumption	entertainment	2.10%	Autres service prives	5: other consumption	institutions
	0.70%	bank	5: other consumption	institutions	1.40%	Service de transport prive	5: other consumption	institutions
	2.10%	bar-caf_	5: other consumption	entertainment	1.80%	Bar, cafe, restaurant, hotel	5: other consumption	entertainment
	1.50%	health institution	5: other consumption	institutions	0.65%	Ecole, lycee, universite publics	5: other consumption	public sector
	2.50%	education institution	5: other consumption	institutions	0.61%	Cabine telephone privee	5: other consumption	institutions
	0.60%	internet	5: other consumption	internet	1.00%	Clinique, laboratoire medical public	5: other consumption	public sector
	11.50%	supermarket	4: market, larger stores	large stores	1.98%	Telephone, eau, electricite	5: other consumption	public sector
	0.70%	department store	4: market, larger stores	large stores	1.17%	Ecole, lycees, universite privas	5: other consumption	institutions
	3.40%	grocery store	3: market, small store-front	specialized stores	0.57%	Magasin de gros a petits prix	4: market, larger stores	large stores
	22.10%	specialized shop	3: market, small store-front	specialized stores	0.69%	Grands magasin	4: market, larger stores	large stores
	4.00%	pharmacy	3: market, small store-front	specialized stores	1.69%	Station service (lubrifiants)	3: market, small store-front	specialized stores
	6.50%	vehicle	3: market, small store-front	specialized stores	1.30%	Atelier, service reparation	3: market, small store-front	specialized stores
					1.03%	Pharmacie	3: market, small store-front	specialized stores
					0.53%	Quincallerie (petite taille)	3: market, small store-front	specialized stores
Informal	8.30%	person	2: market, no store front	individual	14.3%	Boutique de quartier	3: market, small store-front	convenience stores
	1.40%	street seller	2: market, no store front	street selling	0.88%	Kiosque ou echoppe quartier	2: market, no store front	street selling
	1.20%	fair	2: market, no store front	street selling	0.98%	Marchant ambulants	2: market, no store front	street selling
	1.40%	small market	2: market, no store front	street selling	40.5%	Marche	2: market, no store front	street selling
	4.80%	private service	2: market, no store front	individual	12.4%	Menage	1: non-market	from a household/transfers
	0.80%	small shop	2: market, no store front	street selling	10.1%	Bien ou service autoproduit	1: non-market	self production
					0.51%	Cadeau recu en nature ou en espace	1: non-market	from a household/transfers
Unspecified	20.80%	unspecified		not applicable/other				

	BURUNDI				CAMEROON			
Sector			new classification	recoded as			new classification	recoded as
Formal	2.10%	Secteur public ou parapublic	5: other consumption	public sector	7.00%	Hotels/bars/restaurants	5: other consumption	entertainment
	3.60%	Autre lieu d'achat formel	4: market, larger stores	large stores	2.90%	Presetation de services publics	5: other consumption	public sector
	0.80%	Magasin, atelier formel (societe)	3: market, small store-front	specialized stores	2.10%	Cliniques	5: other consumption	institutions
					7.50%	Secteur transport	5: other consumption	public sector
					1.00%	Supermarche/Grand magasin	4: market, larger stores	large stores
					3.80%	Magasin specialistes	3: market, small store-front	specialized stores
Informal	3.87%	Vendeur ambulant	2: market, no store front	street selling	2.50%	Prestation de services individuels	5: other consumption	individual
	13.87%	Autre lieu d'achat informel	2: market, no store front	other informal	10.70%	Epiceries/Boutiques/Echoppes	3: market, small store-front	corner shops
	32.92%	Marche public	2: market, no store front	street selling	0.80%	Vendeurs specialises hors magasins	2: market, no store front	street selling
	26.84%	Domicile du vendeur	1: non-market	other informal	26.40%	Marches	2: market, no store front	street selling
	14.30%	Bien ou service autoproduit	1: non-market	self production	3.40%	Kiosque de jeux et Call Box	2: market, no store front	street selling
	0.86%	Cadeau Recu	1: non-market	from a household/transfers	3.40%	Vente ambulante	2: market, no store front	street selling
					1.90%	Domicile de vendeur	1: non-market	from a household/transfers
					0.90%	Dans la nature/forit/brousse	1: non-market	self production
					3.60%	Auto production	1: non-market	self production
					14.80%	Don, cadeau recu	1: non-market	from a household/transfers
Unspecified					7.30%	Autre		not applicable/other

	CHILE				COLOMBIA		
Sector			new classification	recoded as		new classification	recoded as
Formal	1.98%	INTERNET	5: other consumption	internet	5.08% Restaurantes	5: other consumption	entertainment
	2.37%	HOSPITAL PñBLICO Y CONSULTORIOS	5: other consumption	public sector	1.05% Cafeteras y establecimientos de comidas rpidas	5: other consumption	entertainment
	5.39%	CLeNICAS	5: other consumption	public sector	0.96% Televentas y ventas por catlogo	5: other consumption	internet
	0.96%	RESTAURANTES Y BARES	5: other consumption	entertainment	0.00% A travs de Internet	5: other consumption	internet
	4.29%	DISTRIBUIDORAS - MAYORISTAS	4: market, larger stores	large stores	10.07% Almacenes o supermercados de cadena y tiendas por departamento	4: market, larger stores	large stores
	26.55%	SUPERMERCADOS	4: market, larger stores	large stores	2.05% Plazas de mercado y galeras	4: market, larger stores	large stores
	4.85%	FARMACIAS	3: market, small store-front	specialized stores	0.55% Hipermercados	4: market, larger stores	large stores
	0.53%	TIENDA ESPECIALIZADA	3: market, small store-front	specialized stores	11.12% Establecimientos especializados en la venta del artculo o la prestacin del servicio adquirido	3: market, small store-front	specialized stores
	1.61%	FERRETEReAS Y MULTIFERRETEReAS	3: market, small store-front	specialized stores	1.73% Farmacias y drogueras	3: market, small store-front	specialized stores
Informal	13.30%	ALMACN TRADICIONAL	3: market, small store-front	corner shops	0.85% Graneros	3: market, small store-front	corner shops
	2.85%	COMERCIO AMBULANTE	2: market, no store front	street selling	4.72% Supermercados de barrio	3: market, small store-front	corner shops
	3.27%	FERIAS LIBRES	2: market, no store front	street selling	13.55% Tiendas de barrio	3: market, small store-front	corner shops
	0.93%	VEGAS - MERCADOS	2: market, no store front	street selling	1.13% Persona particular	2: market, no store front	individual
					1.71% Vendedores ambulantes o ventas callejeras	2: market, no store front	street selling
					0.89% Transfers, from household	1: non-market	from a household/transfers
					10.31% Self production	1: non-market	self production
Unspecified	29.33%	NA		not applicable/other	10.58% [Unspecified]		not applicable/other
	0.93%	ESTABLECIMIENTOS IMPOSIBLES DE IDENTI..		not applicable/other	21.98% [Missing]		missing

	COSTA RICA				DEM. REPUBLIC of CONGO			
Sector			new classification	recoded as			new classification	recoded as
Formal	3.86%	Restaurante / soda / cafetera / heladera	5: other consumption	entertainment	3.07%	Achat secteur public	5: other consumption	public sector
	1.71%	Comedor en lugar de trabajo	5: other consumption	entertainment	0.54%	Achat supermarche	4: market, larger stores	large stores
	1.06%	En el exterior	5: other consumption	abroad	3.19%	Achat magasin indo-pakistanais	3: market, small store-front	specialized stores
	1.86%	Laboratorio / clnica / centro medico	5: other consumption	institutions	3.77%	Achat magasin non indo-pakistanais	3: market, small store-front	specialized stores
	17.11%	Supermercado	4: market, larger stores	large stores				
	1.22%	Tienda por departamentos	4: market, larger stores	large stores				
	3.42%	Tienda de ropa / zapatera / perfumera	3: market, small store-front	specialized stores				
	4.26%	Gasolinera y estacion de servicio	3: market, small store-front	specialized stores				
	1.10%	Carnicera / pescadera	3: market, small store-front	specialized stores				
	3.39%	Almacn de electrodomosticos y de tecnologas	3: market, small store-front	specialized stores				
	11.34%	Local especializado	3: market, small store-front	specialized stores				
	0.98%	Salones de esttica o belleza	3: market, small store-front	specialized stores				
Informal	6.21%	Pulperia o minisuper	3: market, small store-front	corner shops	10.07%	Achat Ambulant	2: market, no store front	street selling
	9.06%	Recibido o comprado a otros hogares	2: market, no store front	street selling	36.48%	Achat marche public	2: market, no store front	street selling
	0.80%	Local de artculos usados	2: market, no store front	street selling	5.76%	Autre lieu informel	2: market, no store front	other informal
	2.41%	Vendedor ambulante o a domicilio	1: non-market	self production	17.88%	Achat domicile	1: non-market	self production
	0.80%	Retiro del negocio	1: non-market	self production	17.53%	Bien ou service autoproduit	1: non-market	self production
					1.38%	Cadeau recu	1: non-market	from a household/transfers
Unspecified	3.72%	Otro		not applicable/other				
	1.38%	Imputado		not applicable/other				
	20.03%	[Missing]		missing				

	ECUADOR				MEXICO			
Sector			new classification	recoded as			new classification	recoded as
Formal	6.61%	Restaurantes, salones	5: other consumption	entertainment	0.72%	Diconsa	5: other consumption	public sector
	2.61%	Transporte de pasajeros	5: other consumption	public sector	2.62%	Loncherias, fondas, torterias , cocinas economicas, cenadurias	5: other consumption	entertainment
	0.74%	Establecimientos educativos	5: other consumption	institutions	2.35%	Restaurantes	5: other consumption	entertainment
	0.74%	Hipermercados	4: market, larger stores	large stores	2.08%	Tiendas departamentales	4: market, larger stores	large stores
	2.10%	Supermercados de cadena	4: market, larger stores	large stores	11.37%	Supermercados	4: market, larger stores	large stores
	0.63%	Ropa de todo tipo	3: market, small store-front	specialized stores	0.50%	Compras fuera del pai_s	4: market, larger stores	large stores
	1.08%	Tercena/carnicera	3: market, small store-front	specialized stores	0.97%	Tiendas con membresi_a	4: market, larger stores	large stores
	0.89%	Boticas y farmacias	3: market, small store-front	specialized stores	21.11%	Tiendas especi_ficas del ramo	3: market, small store-front	specialized stores
	5.16%	Panaderas	3: market, small store-front	specialized stores				
	2.35%	Gasolineras	3: market, small store-front	specialized stores				
	0.64%	Personas particulares	2: market, no store front	individual				
Informal	0.96%	Bodegas, distribuidores	3: market, small store-front	corner shops	0.58%	Tiendas de conveniencia	3: market, small store-front	corner shops
	30.38%	Tiendas de barrio	3: market, small store-front	corner shops	12.81%	Tiendas de abarrotes	3: market, small store-front	corner shops
	12.89%	Mercados	2: market, no store front	street selling	3.11%	Vendedores ambulantes	2: market, no store front	street selling
	0.62%	Ferias libres	2: market, no store front	street selling	5.61%	Persona particular	2: market, no store front	individual
	2.56%	Vendedores ambulantes	2: market, no store front	street selling	3.65%	Mercado	2: market, no store front	street selling
	23.04%	Productos autoconsumo, autosuministro	1: non-market	self production	1.99%	Tianguis o mercado sobre ruedas	2: market, no store front	street selling
					1.33%	Auto produccin	1: non-market	self production
Unspecified	1.83%	Otros		not applicable/other	28.79%	No aplica		not applicable/other

	MOROCCO				MOZAMBIQUE			
Sector			new classification	recoded as			new classification	recoded as
Formal	35.48%	Public and semi-public agencies	5: other consumption	public sector	8.80%	loja	3: market, small store-front	specialized stores
	4.11%	Private education institution	5: other consumption	institutions				
	1.50%	Regular transportation means (bus, train, plane)	5: other consumption	public sector				
	0.66%	Public baths, shower, swimming pool	5: other consumption	public sector				
	1.30%	Medical care in private institution	5: other consumption	institutions				
	1.69%	Public administration	5: other consumption	public sector				
	0.84%	Modern clothing shop	3: market, small store-front	specialized stores				
	1.83%	Pharmacy	3: market, small store-front	specialized stores				
	0.53%	Small Bookshop, kiosk	3: market, small store-front	specialized stores				
	1.22%	Craftsman's shop (hairdresser, tailor)	3: market, small store-front	specialized stores				
	3.12%	Butcher or retail chicken seller	3: market, small store-front	specialized stores				
	0.65%	Gas stations (benzine)	3: market, small store-front	specialized stores				
Informal	0.73%	Cafe, non-standing restaurant	5: other consumption	informal entertainment	18.65%	mercado informal	2: market, no store front	street selling
	9.69%	Neighborhood or village grocer	3: market, small store-front	corner shops	12.24%	mercado	2: market, no store front	street selling
	1.85%	Grocer's	3: market, small store-front	corner shops	31.49%	auto produo	1: non-market	self production
	2.63%	Neighborhood market	2: market, no store front	street selling				
	0.52%	City market or central market	2: market, no store front	street selling				
	10.73%	Weekly market	2: market, no store front	street selling				
	1.34%	Itinerant merchant selling on sidewalks	2: market, no store front	street selling				
	3.62%	Self production/consumption	1: non-market	self production				
Unspecified	4.70%	Other places		not applicable/other	9.65%	missing		missing
	0.66%	Unknown		not applicable/other	18.83%	outro		not applicable/other
	5.81%	Not relevant		not applicable/other				

	NIGER				PAPUA NEW GUINEA			
Sector			new classification	recoded as			new classification	recoded as
Formal	7.27%	Prestation services publiques	5: other consumption	public sector	34.45%	Supermarket	4: market, larger stores	large stores
	0.84%	Secteur transport	5: other consumption	public sector				
	0.91%	Hotel, bar restaurant	5: other consumption	entertainment				
Informal	45.54%	Epicerie, boutique	3: market, small store-front	convenience store	9.35%	Small shop, canteen, tuck shop	3: market, small store-front	corner shops
	4.81%	Vente ambulante	2: market, no store front	street selling	10.52%	Local market	2: market, no store front	street selling
	20.19%	Marche	2: market, no store front	street selling	3.76%	Street vendor	2: market, no store front	street selling
	4.93%	Prestation service individuels	2: market, no store front	individual	10.17%	Gift	1: non-market	from a household/transfers
	4.41%	Auto production	1: non-market	self production	14.17%	Home production	1: non-market	self production
	4.10%	Cadeau recu	1: non-market	from a household/transfers				
Unspecified	6.61%	Autre		not applicable/other	17.57%	Other		not applicable/other

	PERU				REPUBLIC of CONGO			
Sector			new classification	recoded as			new classification	recoded as
Formal	3.42%	Empresas de Transporte formales	5: other consumption	public sector	3.94%	Hotels, restaurants, bars, cafes	5: other consumption	entertainment
	0.54%	Talleres formales	5: other consumption	institutions	2.51%	Cliniques, laboratoires mdicaux et coles	5: other consumption	institutions
	0.62%	Clnica particular	5: other consumption	institutions	3.94%	Secteur transports	5: other consumption	public sector
	1.56%	Centro de estudios	5: other consumption	institutions	5.84%	Prestataires de services publics	5: other consumption	public sector
	0.96%	Restaurantes y/ bares	5: other consumption	entertainment	1.03%	Grands magasins	4: market, larger stores	large stores
	1.33%	Grifos de empresas	5: other consumption	public sector	3.43%	Epiceries modernes	3: market, small store-front	specialized stores
	1.05%	Bodega (por mayor)	4: market, larger stores	large stores	6.98%	Autres commerces modernes	3: market, small store-front	specialized stores
	3.42%	Supermercado	4: market, larger stores	large stores				
	3.63%	Farmacia	3: market, small store-front	specialized stores				
	5.78%	Tienda especializada al por menor	3: market, small store-front	specialized stores				
	0.91%	Librera	3: market, small store-front	specialized stores				
	0.79%	Panadera	3: market, small store-front	specialized stores				
	0.65%	Peluquera	3: market, small store-front	specialized stores				
Informal	14.62%	Bodega (por menor)	3: market, small store-front	corner shops	42.78%	Marches	2: market, no store front	street selling
	23.98%	Mercado (por menor)	2: market, no store front	street selling	6.17%	Marchands ambulants	2: market, no store front	street selling
	2.67%	Feria	2: market, no store front	street selling	8.44%	Echoppes sur marches et sur bord de route	2: market, no store front	street selling
	4.97%	Ambulante	2: market, no store front	street selling	5.50%	Prestataires de services individuels	2: market, no store front	individual
	3.34%	Mercado (por mayor)	2: market, no store front	street selling	4.55%	Produit autoconsommes	1: non-market	self production
					3.93%	Menages	1: non-market	self production
Unspecified	22.31%	Otro (Especifique)		not applicable/other				

	RWANDA				SOUTH AFRICA			
Sector			new classification	recoded as			new classification	recoded as
Formal	2.44%	Bar/restaurant	5: other consumption	entertainment	38.64%	Chain store	4: market, larger stores	chain stores
	0.59%	Supermarket/big shop	4: market, larger stores	large stores	11.18%	Other retailer	3: market, small store-front	other retailers
	4.63%	Specialized shop	3: market, small store-front	specialized stores				
Informal	13.55%	Small shop/boutique	3: market, small store-front	corner shops	2.72%	Other	2: market, no store front	other informal
	13.14%	Service provider	2: market, no store front	individual	0.88%	Street trading	2: market, no store front	street selling
	0.77%	Mobile seller	2: market, no store front	street selling	0.63%	From a household	1: non-market	from a household/transfers
	1.72%	Individual	2: market, no store front	individual				
	12.47%	Market	2: market, no store front	street selling				
	11.49%	Self production	1: non-market	self production				
	26.49%	From a household	1: non-market	from a household/transfers				
Unspecified	12.71%	Other		not applicable/other	0.55%	Not applicable		not applicable/other
					45.11%	Unspecified		not applicable/other

	TANZANIA				URUGUAY			
Sector			new classification	recoded as			new classification	recoded as
Formal	37.49%	Shop	3: market, small store-front	specialized stores	0.77%	Bar, Pizzeria	5: other consumption	entertainment
					0.52%	Cantina Trabajo Colegio	5: other consumption	entertainment
					0.85%	Restaurante, Parrillada	5: other consumption	entertainment
					11.73%	Autoservicio, Cadena de Supermercados	4: market, larger stores	large stores
					0.76%	Fuera del pais	4: market, larger stores	large stores
					0.97%	Shopping o galeria	4: market, larger stores	large stores
					2.29%	Merceria, Tienda	3: market, small store-front	specialized stores
					2.58%	Carniceria, Polleria, Pescaderia	3: market, small store-front	specialized stores
					1.49%	Panaderia, Confiteria	3: market, small store-front	specialized stores
					0.66%	Verduleria, Puesto, Fruteria	3: market, small store-front	specialized stores
					0.90%	Zapateria, Marroquineria, Talabarteria	3: market, small store-front	specialized stores
					1.28%	Casa de electrodomesticos, telefonos	3: market, small store-front	specialized stores
					0.71%	Farmacia, Perfumeria, Panalera	3: market, small store-front	specialized stores
Informal	2.44%	Street vendor	2: market, no store front	street selling	0.75%	Quiosco, Salon	3: market, small store-front	corner shops
	22.64%	Market	2: market, no store front	street selling	7.76%	Almacen	3: market, small store-front	corner shops
	4.73%	Other household	1: non-market	from a household/transfers	1.50%	Feria vecinal	2: market, no store front	street selling
	1.82%	Gift or free	1: non-market	from a household/transfers	1.01%	Vendedor ambulante, Puesto callejero, Carrito	2: market, no store front	street selling
	15.90%	Produced by household	1: non-market	self production				
Unspecified	13.77%	Other		not applicable/other	11.66%	Missing		missing
					47.97%	No corresponde		not applicable/other

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:

$$q_{jl} = \frac{\epsilon_{jl}}{\epsilon_{jl} - 1} \frac{w}{\phi_{jl}} \tag{12}$$

This implies the following expression for consumer prices:

$$p_{j1} = (1 + t_j) \frac{\epsilon_{j1}}{\epsilon_{j1} - 1} \frac{w}{\phi_{j1}} \tag{13}$$

and

$$p_{j1} = \frac{\epsilon_{j0}}{\epsilon_{j0} - 1} \frac{w}{\phi_{j0}} \tag{14}$$

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:

$$x_{jl} = \Big(\sum_k \alpha_{jlk} x_{jlk}^{\frac{\rho-1}{\rho}} \Big)^{\frac{\rho}{\rho-1}} \tag{15}$$

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:

$$p_{jl} = (1 + t_j f_{jl}) \frac{P_{jl}}{\phi_{jl}} \frac{\epsilon_{jl}}{\epsilon_{jl} - 1} \tag{16}$$

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:

$$P_{jl} = \Big(\sum_k \alpha_{jlk}^{\rho} p_{jlk}^{1-\rho} \Big)^{1/(1-\rho)} \tag{17}$$

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:

$$p_{jlk} = (1 + t_k f_k (1 - f_{jl})) w \frac{\rho}{\rho - 1} \tag{18}$$

Combining expressions (16), (17) and (18), 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:

$$\frac{\partial p_{j1}}{\partial t_j} \frac{1 + t_j}{p_{j1}} = 1 \tag{19}$$

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

$$\frac{\partial p_{j0}}{\partial t_j} \frac{1 + t_j}{p_{j0}} = s_{j0F} \tag{20}$$

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

$$s_{j0F} = \sum_k f_k \alpha_{j0k}^{\rho} P_{j0}^{\rho-1} p_{j0k}^{1-\rho} \tag{21}$$