DISCUSSION PAPER SERIES

DP14945

INFORMALITY, CONSUMPTION TAXES AND REDISTRIBUTION

Lucie Gadenne, Anders Jensen and Pierre Bachas

DEVELOPMENT ECONOMICS

PUBLIC ECONOMICS

ISSN 0265-8003

INFORMALITY, CONSUMPTION TAXES AND REDISTRIBUTION

Lucie Gadenne, Anders Jensen and Pierre Bachas

Discussion Paper DP14945
Published 25 June 2020
Submitted 18 June 2020

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programmes:

- Development Economics
- Public Economics

INFORMALITY, CONSUMPTION TAXES AND REDISTRIBUTION

Abstract

Can consumption taxes reduce inequality in developing countries? We combine household expenditure data from 31 countries with theory to shed new light on the redistributive potential and optimal design of consumption taxes. We use the type of store in which purchases occur to proxy for informal (untaxed) consumption. This enables us to characterize the informality Engel curve: we find that the budget share spent in the informal sector steeply declines with income, in all countries. The informal sector thus makes consumption taxes progressive: households in the richest quintile face an effective tax rate that is twice that of the poorest quintile. We extend the standard optimal commodity tax model to allow for informal consumption and calibrate it to the data to study the effects of different tax policies on inequality. Contrary to consensus, we show that consumption taxes are redistributive, lowering inequality by as much as personal income taxes. Once informality is taken into account, commonly used redistributive policies, such as reduced tax rates on necessities, have a limited impact on inequality. In particular, subsidizing food cannot be justified on equity or efficiency grounds in several poor countries.

JEL Classification: E26, H21, H23, O23

Keywords: Household Budget Surveys, inequality, Informality, redistribution, taxes

Lucie Gadenne - l.gadenne@warwick.ac.uk
University of Warwick, Institute for Fiscal Studies and CEPR

Anders Jensen - anders_jensen@hks.harvard.edu
Harvard Kennedy School, NBER

Pierre Bachas - pbachas@worldbank.org
World Bank

Informality, Consumption Taxes and Redistribution

Pierre Bachas, Lucie Gadenne & Anders Jensen*

Abstract

Can consumption taxes reduce inequality in developing countries? We combine household expenditure data from 31 countries with theory to shed new light on the redistributive potential and optimal design of consumption taxes. We use the type of store in which purchases occur to proxy for informal (untaxed) consumption. This enables us to characterize the informality Engel curve: we find that the budget share spent in the informal sector steeply declines with income, in all countries. The informal sector thus makes consumption taxes progressive: households in the richest quintile face an effective tax rate that is twice that of the poorest quintile. We extend the standard optimal commodity tax model to allow for informal consumption and calibrate it to the data to study the effects of different tax policies on inequality. Contrary to consensus, we show that consumption taxes are redistributive, lowering inequality by as much as personal income taxes. Once informality is taken into account, commonly used redistributive policies, such as reduced tax rates on necessities, have a limited impact on inequality. In particular, subsidizing food cannot be justified on equity or efficiency grounds in several poor countries.

JEL: E26, H21, H23, 023

Keywords: Household Budget Surveys, Inequality, Informality, Redistribution, Taxes.

*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 would like to thank Michael Best, Anne Brockmeyer, Roberto Fatal, Rema Hanna, Gordon Hanson, Xavier Jaravel, Michael Keen, Joana Naritomi, Henrik Kleven, Wojciech Kopczuk, Joel Slemrod, Johannes Spinnewijn, Mazhar Waseem and numerous seminar participants for helpful comments. We thank Eva Davoine, Elie Gerschel, Mariana Racimo, Roxanne Rahnama and Alvaro Zuniga for excellent research assistance. We gratefully acknowledge financial support from the Weatherhead Center for International Affairs at Harvard University, the World Bank Development Economics' Research Support Budget, and the IFS's TaxDev center. Replication codes for the paper are available here.

1 Introduction

Inequality in developing countries is higher than in most rich countries and has remained high over the past 30 years (Alvaredo and Gasparini, 2015). To what extent can tax systems redistribute income in these countries? In this paper, we combine microdata and theory to shed new light on the redistributive potential and optimal design of consumption taxes, the main source of government revenue in developing countries. We find that consumption taxes are redistributive, lowering inequality by as much as income taxes. These effects are primarily due to different informal (untaxed) consumption patterns along the household income distribution. We also show that, once informal consumption is accounted for, commonly used redistributive policies such as reduced rates on necessity goods hardly reduce inequality. Our results stand in contrast to the existing literature, which does not take into account the existence of informal consumption, and concludes that consumption taxes have a limited or negative redistributive effect in developing countries (see for example Lustig, 2018).

A major constraint in studying informality is that, by definition, informal sector purchases are hard to observe and to link to consumers' incomes. We innovate by using the types of stores in which households report purchasing items – ranging from home production to street vendors to supermarkets – to proxy for consumption from the informal sector. We create a new micro-database of expenditure surveys from 31 low-and middle-income countries which contains information on the store type at the transaction-level. The store-type classification is harmonized across countries and consistent with the macro-development taxonomy of traditional versus modern retailers (Lagakos, 2016). We assign each store type to the formal or informal sector, building on micro evidence from retail censuses and on the literature showing that large, modern retailers are much more likely to remit taxes than smaller, traditional retailers (Kleven et al., 2016). This consumption-based measure of the informal sector complements preexisting approaches which focus on informality at the firm and/or worker level.[1]

Our paper makes three contributions. First, we produce new stylized facts on consumption patterns across the income distribution and over development. We document the existence of downward-sloping Informality Engel Curves (IECs): within each country, the informal budget share declines steeply with household income. The IECs have a stable, log-linear functional form in all countries and their slopes remain negative even after controlling for household location and narrow categories of goods. We provide sug-

[1] Our measure is related to the literature in public finance which uses consumption data to infer evasion behavior (see Pissarides and Weber, 1989; Feldman and Slemrod, 2007; Morrow et al., 2019)

gestive evidence that the residual slopes may be explained by richer households valuing quality more and formal firms selling higher-quality products.[2] The shape of the IECs implies that the de facto exemption of the informal sector from taxes makes consumption taxes progressive (the taxed budget share increase with income). We find that with a simple uniform rate on all goods, the richest quintile pays twice as much in taxes as the poorest quintile in the average country. This 'progressivity dividend' from exempting the informal sector is largest in the poorest countries and decreases with development.

We similarly characterize food Engel curves and confirm that they are downward sloping in all countries, consistent with an extensive literature (for example Deaton and Paxson, 1998). This shape motivates the frequent use of de jure exemptions of food from consumption taxes. We find that food exemption makes consumption taxes progressive, but less so on average than the de facto exemption of the informal sector. The progressivity impact of exempting food is nil in the poorest countries, where poorer households' food consumption is almost exclusively informal, and increases with development.

Our second contribution is a simple model that derives the implications of these consumption patterns for tax policy. We extend the multi-person model of optimal commodity taxation of Diamond (1975) in two directions: we introduce formal and informal (untaxed) varieties of each good and we allow for changes in consumption patterns over the development path. Allowing for informal varieties increases the efficiency cost of consumption taxes, because households can substitute to informal varieties when these taxes increase. It increases their equity benefits, by making consumption taxes progressive as long as IECs are downward sloping. Calibrating the model to the data, we find that the optimal level of rate differentiation between food and non-food products increases with development.[3] This is due to the shape of the formal food Engel curve, which is flat in low-income countries but negative in middle-income countries. In some poorest countries, we find that subsidizing food relative to non-food cannot be justified on equity or on efficiency grounds.

Our third contribution is to investigate the impact of consumption tax policies on inequality by combining the calibrated optimal rates with the household data.[4] We find that setting a uniform tax rate on all goods while taking into account the de facto exemption of the informal sector achieves on average as much inequality reduction as existing

[2]This is in line with evidence in Faber and Fally (2017) and Atkin et al. (2018b).

[3]This is true for our baseline value of the elasticity of substitution in consumption between formal and informal varieties, the parameter which determines the size of the efficiency cost of taxation due to the existence of an informal sector. We use a baseline value taken from empirical estimates in the literature and discuss the robustness of our results to alternative values.

[4]The impact of a tax policy on inequality depends not only on its progressivity, but also on the level of tax rates and size of tax bases which determine the share of taxes in households' budgets.

direct tax policies (income taxes and social security) in developing countries. Moreover, this policy achieves 75% of the inequality reduction obtained in a counterfactual world with perfect enforcement in which governments can tax both formal and informal varieties (thereby taxing a much larger base) and optimally differentiate rates between food and non-food goods. Finally, in the realistic world in which informal varieties cannot be taxed, the de facto exemption of the informal sector achieves twice as much inequality reduction as the de jure optimal rate differentiation.

The results are robust to changing our main assumptions. First, we obtain similar results when changing the formality assignment for places of purchases for which there is uncertainty.[5] Second, we assume at baseline a zero pass-through of taxes to prices in informal stores. We use data on the universe of formal and informal retailers in Mexico to relax this assumption, and find that doing so only slightly reduces the redistributive effect of informality. Third, our baseline results use total household expenditure to proxy for household income.[6] We obtain similar results using income distributions constructed with estimates of distributional saving rates. Finally, we discuss the implications of our modeling assumption that no direct tax instrument is available to the government.

Our findings have several main implications. First, the result that consumption taxes reduce inequality in developing countries runs counter to the consensus view in the policy and academic literatures, which argue that these taxes have little or no redistributive potential (Sah 1983; Shah and Whalley 1991; Gemmell and Morrissey 2005; Coady 2006). These previous studies do not account for differences in informal consumption along the household income distribution, which our results show have large, positive impacts on redistribution. This negative prior may explain why studies of redistribution in developing countries often focus on the design of government transfers (Tanzi 1998; Clements et al. 2015). Our findings suggest more attention should be paid to the redistributive potential of tax design in these countries. Second, we obtain our results by measuring informality at the consumer level, while most studies focus on informality at the firm and/or worker level (DeSoto 1989; De Paula and Scheinkman 2010; Gerard and Gonzaga 2016; Ulyssea 2018). This approach leads to a more nuanced perspective on the welfare effects of the informal sector, which is often thought to trap the poor in low-productivity firms (La Porta and Shleifer 2014; Hsieh and Klenow 2014). In particular, our findings caution that while enforcement policies to reduce informality may yield production efficiency gains (Ulyssea, 2020), they might also have distributional costs by increasing the

[5]This is because the association between budget shares and household income is strongest for places of purchase which are clearly informal (non-market consumption) or clearly formal (large supermarkets).

[6]This is due to data limitations. Expenditure surveys in developing countries often do not attempt to directly measure income because of serious measurement issues - see Deaton (1997) for a discussion.

tax burden on poorer households. Third, by combining micro survey data from countries spanning a wide range of economic development (from Burundi to Chile), we document new stylized facts on consumption patterns across 'modern' and 'traditional' retailers as households get richer within and across countries. Our database could be used to test competing theories of development and global retail (Bronnenberg and Ellickson, 2015).

The rest of the paper is structured as follows. In the following sub-section, we summarize our contributions relative to existing literatures. Section 2 describes our data sources and methodology. Section 3 provides new stylized facts on consumption patterns across places of purchase and types of goods as households get richer within and between countries. Section 4 investigates the implications of these consumption patterns for tax progressivity. Section 5 develops a model to characterize optimal commodity tax policy with informal consumption. Section 6 calibrates the model using our data and investigates the impacts of consumption taxes on inequality. Section 7 concludes.

1.1 Related Literature

Our paper makes three main contributions to the literatures on tax policy and consumption in developing countries. First, we introduce differences in informal consumption along the income distribution as a novel channel through which consumption taxes are redistributive. This contrasts with the existing literature, which omits informal consumption and concludes that consumption taxes have, at best, limited redistributive impacts in developing countries (for recent studies, see Harris et al., 2018 and Lustig, 2018).[7] Our approach is thus related to recent papers on the redistributive implications of differences in consumption patterns across the income distribution, which focus on developed countries (Faber and Fally, 2017; Jaravel, 2018; Allcott et al., 2019).

Second, we contribute to the literature on the design of consumption taxes under limited enforcement capacity (Pomeranz, 2015; Naritomi, 2019; Waseem, 2020).[8] We revisit the redistributive potential of differentiating consumption tax rates across goods. Such policies are commonly implemented (Ebrill et al., 2001), but papers that study the redistribution achieved by optimal rate differentiation only do so on a country-by-country basis.[9] We combine theory and novel data to undertake the first systematic analysis

[7]Two exceptions are Muñoz and Cho (2003) and Jenkins et al. (2006) who use retailer information to classify expenditures as formal or informal, respectively in the Dominican Republic and in Ethiopia.

[8]More generally, we contribute to the growing literature on taxation under weak enforcement (Almunia et al., 2019; Basri et al., 2019; Brockmeyer and Hernandez, 2019; Weigel, 2019; Londoño-Velez, 2020).

[9]Including: In the United Kingdom, Sah (1983); in India, Ahmad and Stern (1984), Ray (1986), Srinivasan (1989); in Australia, Creedy (2001).

of optimal rate differentiation across goods and its impact on inequality across a large sample of low and middle-income countries.[10] More generally, we build on the literature on optimal tax design under imperfect enforcement which typically concentrates on the efficiency costs of taxation (Allingham and Sandmo, 1972; Cremer and Gahvari, 1993; Boadway and Sato, 2009); we consider instead both the equity and efficiency implications of informal consumption patterns.[11].

Finally, this paper speaks to the literature on consumption patterns and development. A large body of work documents how budget shares spent on specific goods vary with income, including the well-established food Engel curve (Deaton and Paxson, 1998; Anker et al., 2011; Pritchett and Spivack, 2013; Almås, 2012). We document, in addition, the association between household income and place of purchase both within and across countries. Thus, our analysis complements recent studies that document aggregate changes in the retail sector over development (Bronnenberg and Ellickson, 2015; Lagakos, 2016; Atkin et al., 2018b).[12]

2 Data and Measurement of Informal Consumption

We use two data sets to present new evidence on consumption patterns in developing countries and on the redistributive potential of consumption taxes. Our core sample consists of 31 countries for which the data enable us to proxy for informal consumption at the household level. Our extended sample consists of 80 countries for which the data allow us to document food consumption patterns.

2.1 Core Sample

We assemble our core dataset by combining household expenditure surveys from 31 countries that satisfy three selection criteria. First, the survey must be nationally representative. Second, the survey must record consumption from open diaries rather than pre-filled diaries, which only contain information on selected products. This helps to ensure that the survey covers all expenditures. Third, the diary must ask households to report the store type that each expenditure is purchased from - the *place of purchase* - and this information must be systematically reported in the diaries. This last criterion ensures that we can apply our method to robustly measure informal sector consumption,

[10] Our main results focus on food versus non-food rate differentiation, but we also show results for optimal differentiation between 12 large goods categories.

[11] See Kopczuk (2001) for an exception

[12] Our approach is related to a recent literature which compiles multi-country microdata to study macro changes over the development path (Bick et al., 2018; Donovan et al., 2018).

as described below.

Our core sample from 31 countries contains information on over 400,000 households. Table 1 lists alphabetically the countries in the core sample, with their survey name and year, the number of households, and the average number of purchases reported per household. Countries in the sample are principally located in Latin America and Sub-Saharan Africa. Unfortunately, most household expenditure surveys in Asia do not contain information on the place of purchase[13] Nonetheless, the core sample covers a wide range of development levels, from Burundi (GDP pc of 250 USD) to Chile (15,000 USD). In Section 3.3, we show that this core sample is very comparable to the large set of developing countries in the extended sample, along detailed expenditure dimensions that can be observed in samples.

Appendix B provides further details on the data sources used. Table B1 shows the geographical coverage of the core sample, table B2 lists the surveys considered for inclusion but ultimately discarded due to selection criteria, and table B3 provides further details on the structure of the surveys for each country in the core sample.[14]

2.2 Method: Proxy for Informal Consumption Using Places of Purchase

Our main methodological innovation is to use the place of purchase reported for each expenditure to assign a probability that it was obtained from a formal (tax remitting) source. Most recorded expenditures can be classified by place of purchase into seven categories. The first five pertain to purchases of goods. Ranked by ascending order of retailer size, these are: (1) non-market consumption (e.g. home production); (2) non brick-and-mortar stores (e.g. street stalls, public markets); (3) corner and convenience stores; (4) specialized stores (e.g. clothing stores); and, (5) large stores (e.g. supermarkets, department stores). Purchases of services can be allocated to two main categories: (6) services provided by an institution (e.g. banks, hospitals); and, (7) services provided by an individual (e.g. domestic services).[15] Combined, these categories account for 86% of total household expenditure. The remaining 14% are items for which no place of purchase is specified, primarily utilities, fuel and telecommunication (see Figure A3).[16]

[13]Survey design appears correlated across countries within regions, showing the influence of regional development partners and historical ties. Our data contain one Asia/Pacific country - Papua New Guinea

[14]Replication codes for the paper can be accessed here. This includes the cleaning files of each expenditure survey and the files generating the tables and figures of the paper. The data are not provided as we had to require access to the World Bank or Stat Agency for each country. Yet, these data are often accessible, allowing for a replication of the paper on a subset of countries of interest.

[15]We also use two smaller categories to classify entertainment services into (8) entertainment (e.g. restaurants) and (9) informal entertainment (e.g. food trucks).

[16]We exclude housing expenditure due to limited data on owner-occupied imputed rents.

We assign each category to either the formal or informal sector. We define a category as belonging to the formal sector if it is likely that consumption taxes are remitted on most purchases from that category. According to this definition, many retailers are informal because they evade taxes due on their sales. Alternatively, small retailers may be informal because they are not required to register for consumption taxes due to their size. We do not distinguish between these channels, as both imply that these retailers do not remit consumption taxes. In addition, these concepts are closely related: in countries with low enforcement capacity, the scope for evasion among small firms is such that the net revenue from bringing them into the tax net is small and outweighed by administrative and compliance costs (Ebrill et al., 2001; Keen and Mintz, 2004).

The key assumption behind our assignment method is that larger retailers are more likely to be formal. This is in part mechanically true, as a frequently used criteria for compulsory tax registration is firm size. In addition, a large literature argues that larger firms are less likely to evade taxes. Theoretically, Kleven et al. (2016) develop a model of tax evasion in which informality must be sustained by collusion between firm managers and their employees, and show that larger firms are more compliant since collusion costs increase with firm size. Hsieh and Olken (2014) similarly argue that the effective burden of taxation in developing countries falls more heavily on larger firms. Empirically, Kumler et al. (2015) find that compliance indeed increases with firm size in Mexico. Naritomi (2019) provides evidence suggesting that larger retail stores have more customers and are therefore less able to sustain evasion. Both Bronnenberg and Ellickson (2015) and Lagakos (2016) categorize retailers as either 'traditional' (small, labor intensive retailers - our categories 1 to 3) or modern (large, capital intensive retailers - our categories 4 and 5), and argue that traditional retailers are mostly informal.[17]

We provide direct evidence that formality and firm size are positively correlated. We use data on formality status and firm size available in the retail firm censuses of four countries in our core sample (Cameroon, Mexico, Peru and Rwanda). Panel A in Figure 1 shows the share of formal firms as a function of log employment in each country.[18] In all countries, retailers with a few employees are overwhelmingly informal, but more than 80% of retailers with 20 or more employees are formal. In Mexico, the census classifies retailers in categories which are similar to our broad place of purchase categories. This enables us to go one step further and directly observe retailer size and formality status.

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

[18] Formality is defined as 'being registered with tax authority' in Cameroon and Rwanda, and 'paying value-added-taxes on sales' in Mexico and Peru.

Figure 1 shows for our categories (1) to (5) the log median number of employees (Panel A) and the share of firms paying value-added taxes on their sales (Panel C). We observe that non-brick-and-mortar stores and convenience stores are small and rarely formal, whereas nearly all large stores are formal.

Given this evidence, our baseline formality assignment 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 are therefore untaxed. Categories (2) and (3) (non brick and mortar stores and corner stores) are likely very small and mostly informal, whereas category (5) (large stores) consists mostly of supermarkets which are unlikely to be non-compliant. Our approach assigns retailer-types to informality which are all classified as 'traditional' tax-evading retailers in Lagakos (2016).

For services, we assume that institutions pay taxes while individual providers do not. This leads us to assign category (6) to the formal sector and category (7) to the informal sector. We follow the same logic in assigning expenditures in the unspecified category to the formal sector: the bulk of these expenditures are utilities typically provided by large institutions which cannot evade taxes (Figure A3). Appendix B provides more details on the methodology. Table B4 shows for each country the original names of the places of purchase, their expenditure shares and our formality assignment.

We will investigate the robustness of our results to two alternative formality assignment rules. The first alternative is to assign a non-binary probability to each category that it pays VAT on its sales. We use the share of retailers that pay VAT in each category in the Mexican census to obtain a probability for each category. These probabilities are depicted in Panel C of Figure 1.[19] This enables us to take into account the fact that some small retailers pay taxes – either because a subset of small retailers are fully formal or because all small retailers pay VAT on a subset of their transactions. The second alternative changes specialized stores (category 4) to be fully informal (leaving all other categories unchanged). This is the category for which there is arguably the most uncertainty about its true formality status.

Our formality assignment rule enables us to measure the informal budget share of each household, defined as the share of its total expenditure purchased from informal retailers. In what follows we also consider within-product informal budget shares. We use the UN's detailed COICOP classification of products, which is available at the 2-digit

[19] We assign the following probabilities that the place of purchase is formal to each category: 10% in category 2; 20% in category 3; 50% in category 4; and 90% in category 5. Other categories are unchanged from the baseline assignment rule.

level (12 products), 3-digit level (47 products) and 4-digit level (117 products).[20]

2.3 Extended Sample for Food Expenditures

We complement our core dataset with microdata from the World Bank's Global Consumption Database (GCD). The GCD compiles household expenditure surveys across countries and harmonizes product categories across all surveys at the 2-digit COICOP product level. [21] Merging this dataset with our core sample, we obtain consumption data for 89 low and middle income countries which account for over 50% of the world population. While this extended sample does not contain place of purchase information, we can use it to characterize food consumption levels and food Engel curves. Food products are defined as all items pertaining to 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 consumption taxes more progressive. We also compare food consumption trends in the core and extended samples and find very similar patterns. This helps us argue that the documented informal consumption facts, while only observed in the core sample, may be relevant for the broader sample of developing countries.

3 Engel Curves of Informality and Food across Development

3.1 Informality Engel Curves

The informality Engel curve (IEC) traces the relationship between the informal budget share and total household expenditure within a given country. As is commonly done in developing countries, we use household expenditure to proxy for household income because of issues with measuring income (Deaton and Paxson, 1998; Atkin et al., 2018a). We use the logarithm of total expenditure, in line with the large literature on product-specific Engel curves (starting with Working, 1943; see review in Deaton, 1997).

For illustrative purposes, Figure 2 plots the IEC for one low-income country (Rwanda) and one middle-income country (Mexico). These graphs show a local polynomial fit of household budget share spent in the informal sector on log of household expenditure per capita in 2010 constant USD. The solid and dashed lines represent the median and

[20] We use crosswalks to convert survey-specific product categories to COICOP categories when necessary. This could not be done for three countries: Brazil, Chad and Peru. For these countries, we use survey-specific product categories at the 3 and 4 digit levels.

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

top/bottom 5 percentiles of the expenditure distribution, respectively. To investigate the functional form flexibly, we plot the non-parametric IEC constructed from kernel-weighted polynomial local regressions. In both countries, the IEC is downward sloping and approximately linear. In Rwanda, the informal budget share falls from 90% for the poorest decile of households to 70% for the richest decile. In Mexico, the IEC is steeper, falling from 55% to 25%. Figure A1 plots the IEC for all 31 countries. We find two empirical regularities. First, IECs are downward sloping everywhere. Second, IECs are approximately linear in log expenditure. This suggests there exists a stable functional form relationship between informal budget share and household expenditure in developing countries.[22]

To summarize the information contained in the country-level IECs, we focus on two empirical moments: i) the aggregate informal budget share; ii) the level-log slope of the IEC. In Section 4, we explain how these two moments are sufficient to characterize how consumption patterns affect tax progressivity. We obtain the country-specific slopes from the following regression:

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

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

In Figure 3, we plot the aggregate informal budget share (Panel A) and the estimated IEC slope (Panel B) against countries' per capita GDP. Panel A reveals a large drop in the aggregate informal budget share, from over 90% in the poorest countries to 20% in upper-middle income countries. This decrease in the size of the informal sector over development, obtained using our novel approach based on consumer shopping behavior, is consistent with patterns observed using alternative informality measures, based on labor markets (La Porta and Shleifer, 2014; Morrow et al., 2019) or money demand (Enste and Schneider, 2000). In Panel B, we see that the negative IEC slope first increases in magnitude, between low-income and lower-middle income countries, and then decreases, between lower-middle and upper-middle income countries. The average IEC slope is 9.8, implying a nearly 1 percentage point reduction in informal budget share when household expenditure increases by 10%.

[22] Almås (2012) similarly finds a stable Engel relationship between food budget share and household income around the world. For more dis-aggregated expenditure categories, however, Engel curves have been found to be non-linear and vary across settings (Banks et al., 1997; Atkin et al., 2018a).

Robustness Our results are robust to using the alternative formality assignment rules outlined in section 2.2. We re-estimate the informal budget shares and IEC slopes using these two alternative rules and present results in Figure A2. Our key findings are unchanged: over development, the aggregate informal budget share decreases steeply and the IEC slope first increases then decreases in magnitude. In Figures A4 and A5, we show that this robustness is driven by the fact that those expenditure categories with least uncertainty surrounding formality status are also those with the steepest Engel slopes (including non-market purchases, large stores, and institutional services).

3.2 Understanding Differences in Informal Consumption across Households

Why do poorer households consume a higher share of their budget from the informal sector? This question is important, for at least two reasons. First, by attempting to answer it we investigate the mechanisms behind the shape of IECs. Second, if differences in informal consumption across households can be explained by characteristics that governments can easily observe and target tax reductions or exemptions on, then de jure redistributive policies and de facto informality exemption may in practice have very similar redistributive impacts.

3.2.1 Observable Characteristics

To measure how much of the association between household income and informal consumption shares can be explained by observable characteristics, we estimate the following regression separately for each country:

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

where i indexes a household, X_i are household characteristics and each observation is weighted by the relevant household survey weight. Table 2 shows the average of the slope coefficients β across countries, 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 5% level. Column 1 displays results from the specification without controls. Column 2 controls for household demographics: household size and age, education and gender of the household head.[23] We find that these characteristics do not explain the correlation between informal expenditure and income and, if anything, the IEC slopes become slightly steeper.

[23]Household size controls for economies of scale across households of different size which could affect where households choose to shop - see Deaton and Paxson (1998).

Columns 3 and 4 add controls for households' location, either with an indicator for whether the household lives in a rural area (column 3) or with survey block fixed effects (column 4).[24] These controls allow us to test whether the association between informal budget shares and household expenditure is due to poorer households living in areas with worse access to formal retailers. Despite large differences in average informal budget shares between rural and urban households, controlling for rural locations only explains 13% of the slopes.[25] Controlling for survey blocks explains just over 20% of the differences in informal budget shares between poor and rich households.

In columns 5 to 8 we test whether non-homothetic preferences across goods play a role – when richer households spend more on goods predominantly sold in formal stores. We run a product-level version of specification (2) with goods fixed effects and compute an average goods-level estimate of β for each country.[26] We gradually consider variations within narrower goods categories: we first consider food vs non-food (column 5), then the 12 good categories of the COICOP 2-digit level classification (column 6), then the 47 categories of the COICOP 3-digit level (column 7) and finally the 117 categories at the COICOP 4-digit level (column 8). Preferences across goods explain part of the association: controlling for food goods alone explains 35% of the slopes, while controlling for the 12 broad goods categories explains 41% of the variation. Controlling for narrow goods categories only slightly reduces the slope further.[27]

Finally, column 9 shows the average IEC slope with all controls included. The average IEC slope is 4.3 and remains statistically significant in all but three countries. Overall, observable characteristics explain 54% of the association between informal expenditure shares and household income. We reproduce these results for our two alternative assignment rules and find similar results (see Tables A2 and A3).

3.2.2 Quality-Price Trade-off between the Formal and Informal Sectors

The previous analysis shows that observable location and preferences across goods explain half of the IEC slopes on average, but they remain significantly different from zero in most countries after including these controls. In six countries in our sample the ex-

[24] The survey block is the most granular location information and contains on average 74 households in our surveys. The median survey block is representative on average of 52,900 households.

[25] The informal budget share is on average 67% in rural areas versus 52% in urban areas and IEC slopes are steeper in urban locations. Figure A7 shows the IECs separately for rural and urban areas.

[26] Formally we run the regression: $Share\ Informal_{ig} = \beta * ln(expenditure_g) + \alpha_g + \Gamma X_i + \varepsilon_{ig}$ where $Share\ Informal_{ig}$ is the share of household i's informal expenditure on good g and α_g are goods fixed effects. Each observation is weighted by household survey weights and goods expenditure shares.

[27] Figure A8 displays visually these results for each country by showing the residual IEC slopes when controlling for increasingly narrow goods categories.

penditure module asks households their main reason for choosing a place of purchase for each item; the possible reasons are access, price, quality, store attributes and other.[28] Table 3 reports the average frequencies for each reason. Column 1 shows that across all store types, access is chosen for 41% of purchases, suggesting that controlling even at the survey block might not capture fully the local nature of shopping preferences.[29]

Columns 2 and 3 show the same frequencies separately for informal and formal stores. The key difference that emerges is that households visit informal stores for their prices and formal stores for their quality.[30] This result is robust to a set of controls and to the inclusion of household fixed effects.[31] Figure A9 shows that in each of the six countries, this taste for quality is more prevalent for richer households, as they are up to four times more likely to report quality as the reason for choosing any type of store.

These results suggest that part of the remaining association between informal budget shares and household expenditure in the last column of Table 2 could be due to richer households valuing high-quality goods more and such goods being sold mainly in formal stores. This is in line with results in Faber and Fally (2017) who show that richer households spend more on larger brands in the United States, and Atkin et al. (2018b) who find that richer Mexican households spend more on high-quality products sold by foreign retailers. This explanation could imply that formal varieties of a given good should be more expensive than informal varieties, reflecting quality differences. We test this hypothesis in the 20 core sample countries which report unit values for each purchase. We use unit values to proxy for non-quality adjusted prices. We study price differences between formal and informal stores within the most narrow good classification available, location and measurement units, leading us to interpret price differences as reflecting quality differences (similar to Atkin et al. (2018b)). We limit our analysis to food products since this mitigates measurement issues and because food items are typically exempt from consumption taxes, so that any price difference between formal and informal varieties cannot be due to taxes. Formally, we estimate the price premium in the formal sector in each country separately as follows:

[28] In all surveys, seven reasons are listed which we classify into five categories: access is defined as "The retailer is closer or more convenient" and "The good or service cannot be found elsewhere"; price as "The good or services are cheaper"; quality as "The goods or services are of better quality"; store attributes as "The retailer offers credit" and "The retailer is welcoming or is a friend"; and, other as "Others reasons".

[29] This could reflect the fact that many households in our sample cannot invest in costly durables, such as cars, which may give them access to a wider variety of stores (Lagakos, 2016).

[30] Access is slightly more frequently reported as a reason to visit an informal store, which could reflect a lack of store choices in poorer rural locations.

[31] Within a given household, formal store purchases appear motivated by higher quality while informal store purchases appear driven by lower prices.

$$ln(unit\ value)_{igmu} = \beta\ Formal_{igmu} + \mu_{gmu} + \epsilon_{igmu} \quad (3)$$

where $ln(unit\ value)_{igmu}$ is the unit value reported by household i, for good g, in location m, in units u, and $Formal_{igmu}$ equals one if the good is purchased in a formal store. We add fixed effects at the good * location * unit of measurement level.

Table A5 shows that on average, food unit values are 6.7% higher in formal than informal stores. This formal store premium result is robust to outliers, excluding self-production, and controlling for household characteristics. These results are consistent with the hypothesis that formal stores offer high quality varieties at a higher price.[32] Formal stores might of course differ in other ways reflected in prices, such as a higher productivity, which would make it less likely to find positive price differences.

3.3 Food Engel Curves

In this sub-section, we document the shape of the food Engel curve in developing countries. The shape of the IEC determines how much redistribution can be achieved by de facto exemption of informal consumption; the shape of the food Engel curve similarly determines how much de jure exemption of food can redistributive. Because food Engel curves are typically downward sloping, many governments set reduced rates or fully exempt food products for redistributive purposes.[33] Previous research has studied the magnitude and approximate log-level linearity of the food Engel curves.[34] We combine our core data with microdata from the Global Consumption Database (discussed in section 2.3) to estimate food Engel curves in a uniquely large sample of 89 low and middle income countries.

In the top panels of Figure 4, we present the aggregate food budget share (Panel A) and food Engel slope (Panel B) against countries' GDP per capita. The aggregate budget share spent on food decreases with development, but the percentage-point drop over development is less pronounced for food than for informal consumption. Food Engel curves are typically downward sloping and there is no relationship between the magnitude of the slope and development. In Panels C and D of4 we show the *formal* food aggregate budget share and Engel curve slopes, respectively. The aggregate formal food

[32]Consistent with a quality-gradient in size, we also find that within formal retailers, the larger stores (category 5) charge higher prices than smaller specialized stores formal category (category 4).

[33]Some countries apply reduced rates or exempt all food goods, while other countries target 'basic' food items. We follow the former approach; targeting narrow items may improve redistribution, but also increases the scope for cross-goods misreporting and distortions.

[34]Recent studies include 10 countries in Almås (2012), 22 in Anker et al. (2011) and 38 in Pritchett and Spivack (2013).

budget share is small on average (11%) but increases with development. The slope of the formal food Engel curve is small in magnitude but positive on average in low-income countries; it becomes negative in upper-middle income countries.

Finally, Panels A and B show that the levels and slopes of food Engel curves are very similar between our core sample and extended sample, for countries at the same level of development. This could suggest that while we can only characterize informal consumption patterns in our core sample, our results may be relevant for developing countries more broadly.

4 Progressivity of Consumption Taxes

The consumption patterns described in Section 3 determine the progressivity of consumption taxes both in the average developing country and across development, which we turn to in this section. We say a tax policy is progressive if the effective tax rate (ratio of taxes paid to household income) increases with household income.

4.1 Progressivity in the Average Developing Country

Set-up We study the progressivity of three tax policy scenarios. Scenario #1 imposes a uniform rate on all goods, but assumes informal varieties are not taxed (de facto exemption). Scenario #2 sets a zero rate on food goods (de jure exemption), but assumes formal and informal varieties are taxed. Scenario #3 implements both de facto and de jure exemptions by setting a zero rate on food and assuming informal varieties are not taxed. Scenario #1 illustrates the progressivity of our new informality channel. Scenario #2 corresponds to a counterfactual setting with perfect enforcement capacity; while practically implausible, it provides an unconstrained benchmark against which to compare the informality-constrained scenarios (#1 and #3). Scenario #3 captures the combined progressivity impacts of the government exemption policy and of informality. Importantly, the difference between scenarios #3 and #1 shows the actual impact of governments implementing a de jure exemption, conditional on the de facto informality exemption. We assume for each scenario that the government sets rates such that it collects 10% of total consumption in taxes, thus maintaining total revenue collected constant across scenarios.[35] Finally, we assume full pass-through of taxes to final consumers at baseline, but relax this assumption below.

[35]Distributional analyses are based on the first order impacts of small changes in tax rates, which are captured by the mechanical effects. Households' behavioral responses to tax changes are second order.

To build intuition for our results, we rely on the empirical evidence that the Engel curves of both tax bases (formal goods, non-food goods) are upward sloping and approximately linear with respect to log household income, as shown in Section 3. With a log-linear Engel curve, the progressivity of a tax scenario is decreasing in the aggregate budget share of the tax base, and increasing in the magnitude of the slope of its Engel curve. Consider two countries with the same positive slopes for a good, but different aggregate budget shares. When Engel curves are log-linear, the difference in budget shares spent on the taxed good between rich and poor households is more pronounced in the country with the lower aggregate share. In other words, for the same Engel curve slope the lower the average budget share of a taxed good, the more likely is it that any given purchase of that good is made by a rich household. This means that a purchase of that good becomes a better tag for household income, leading to a more progressive tax system. In addition, when two countries have the same budget share of that good and Engel curves are upward sloping, a steeper slope increases the difference in budget shares between rich and poor households, making the tax system more progressive.

Results Figure 5 shows the progressivity of each scenario for the average developing country in our sample. We obtain three main results. First, the existence of the informal sector makes consumption taxes progressive. Under scenario #1 (red circle line), the effective tax rate increases sharply with household income and the richest quintile pays twice as much taxes (as a share of income) as the poorest quintile. This large progressivity is explained by the steep increase of formal expenditure with household income (Figure 3.) Second, the de facto exemption of the informal sector is more progressive than the de jure exemption of food goods in the counterfactual setting with perfect enforcement. This can be observed by comparing scenario #1 to #2 (green cross line): the ratio of effective tax rate paid by the richest quintile to that of the poorest quintile is almost 50% larger under #1 versus #2. This difference is primarily driven by the fact that formal expenditure constitutes on average a smaller budget share of households' incomes than non-food expenditure among countries in our sample (Panel A, Figure 4 versus Figure 3). Third, the progressivity achieved by the de jure exemption conditional on the de facto exemption is small. We can see this by comparing scenario #3 (orange square line) to scenario #1: exempting food from taxation barely increases progressivity once the exemption of the informal sector is taken into account.

4.2 Progressivity across Development

The evidence presented in Figure 5 for the average country masks considerable heterogeneity across countries. We now turn to characterizing how the progressivity of our policy scenarios changes over the development path. To summarize the progressivity of a scenario we use the ratio of the effective tax rate paid by the richest quintile to that paid by the poorest quintile. The higher this metric, the more progressive the tax policy (a value > 1 implies a progressive tax policy). This metric does not depend on the tax rate used, and is frequently used in the literature (Sah, 1983; Srinivasan, 1989).

Figure 6 plots the country-level progressivity for each scenario as a function of countries' economic development. Panel A shows that the de facto exemption of informal consumption leads to the most progressivity in the poorest countries, where the top 20% pay on average 3 times as much in taxes as the bottom 20%. In these countries formal consumption is rare and therefore a strong tag for household income. Over development, progressivity decreases. This is primarily driven by the steep increase in aggregate formal budget shares (Figure 3, Panel A), which makes formal consumption a worse tag for income as countries become richer. Between low and lower-middle income countries, this negative budget share effect dominates the positive effect on progressivity from a rise in the slope of the formal Engel curve (Figure 3, Panel B). Between lower-middle and upper-middle income countries, the formal Engel slope falls, contributing to the decrease in progressivity.

In Panel B of Figure 6, we study progressivity of the de jure food exemption in the unrealistic setting with perfect enforcement and no informal consumption (scenario #2). The de jure food exemption is substantially less progressive than the de facto informal sector exemption (scenario #1) in the poorest countries, while both scenarios achieve similar levels of progressivity in upper-middle income countries. This is because the non-food budget share is much larger than the formal budget share in the poorest countries (making formal consumption a much better tag for household income than non-food consumption), and grows much less than formal budget shares over development (Figure 4). Finally, by comparing Panel C (scenario #3) to Panel A (scenario #1), we find that de jure exemptions have no effect on progressivity in the poorest countries once the de facto exemption of the informal sector is taken into account. It has a positive impact however in upper-income-countries. This is because the formal food Engel curve is roughly flat in poor countries, and downward sloping among the richest countries in our sample (Figure 4).

4.3 Extensions and Robustness

Pass-through of taxes in the informal sector Our analysis thus far assumes zero pass-through of taxes to prices of informal varieties. This assumption may not hold, for several reasons. First, if the consumption tax is a VAT, informal retailers may pay taxes on their inputs from formal suppliers. Second, competition between formal and informal retailers could lead informal retailers to pass through tax increases to prices. We relax this assumption by allowing some pass-through of taxes to informal prices. In Appendix C.6, we show that under a VAT system the pass-through of taxes to informal prices is equal to the share of formal input costs in informal retailers' total input costs. Using the Mexican census data described above we find that this share is on average 10% among all informal retailers.[36]

Table A6 summarizes our progressivity results under the assumption of a 10% pass-through of taxes to informal prices.[37] The results are partially affected but our key findings remain. First, we find that scenario #1 (exemption of informal varieties) remains very progressive, with the richest quintile paying over 70% more in taxes than the poorest quintile on average. Second, the de facto exemption of the informal sector continues to be more progressive than de jure exemption of food (scenario #2), although the difference in progressivity has decreased. Third, the progressivity impact of the de jure exemption, conditional on allowing for informal consumption, remains smaller than that of the de facto exemption.

Distributional savings rates Our baseline results use total expenditures to proxy for household income, assuming households do not save. Intuitively, allowing for savings both decreases effective tax rates (as savings are not taxed) and decreases the progressivity of all tax scenarios if saving rates increase with income.[38] The distribution of savings across income levels is hard to obtain from expenditure surveys, especially in develop-

[36] In the 2013 Census, 85% of retailers are informal by not reporting any payment on VAT. Among informal retailers, only 8% report paying VAT on inputs, which applies on average to 40% of their intermediate input purchases. The informal retailers that report positive VAT on inputs account for 25% of all informal sales. Combined, this leads to our estimate of 10%.

[37] The share of formal inputs used by informal firms is likely to be an upper bound. First, the 10% number is applied to both fixed and non-fixed establishments, while the latter category (which includes street stalls) is likely to source fewer inputs from formal firms. For this reason, we maintain a 0% pass-through for home production. Second, in Mexico, it is the large informal stores which use formal inputs: The use of formal inputs for smaller stores is closer to 5% on average. Third, segmentation between formal and informal firms is likely to be even larger in poorer countries, leading to lower average pass-through in those countries than in Mexico.

[38] Calculations based on annual income may overstate the regressive nature of consumption taxes since consumption depends on lifetime income, which is less volatile than annual income (Poterba, 1989; Caspersen and Metcalf, 1994).

ing countries where income is hard to measure. To assess how savings could affect our results, we use data from the US Consumer Finance Survey, in which savings rates range from 0% for the poorest households to 15% for the richest quintile.[39] Results are presented in Table A6: allowing for distributional savings decreases the progressivity of all scenarios, as expected, but our main findings are unchanged.

Alternative formality assignment Finally Table A6 presents our progressivity results under the two alternative rules for categorizing places of purchases as formal or informal, as described in Section 2.2 (the probabilistic formality assignment based on the 2013 Mexican Census of retailers and the assignment of specialized stores to the informal sector). Our three main take-aways are unchanged.

5 Optimal Consumption Tax Policy with an Informal Sector

5.1 Set-up

This section studies the implications of the novel consumption facts for optimal tax policy. We extend the multi-person Ramsey model of commodity taxation (Diamond, 1975) to a context in which informal varieties of each good cannot be taxed. We then derive optimal tax rates for three policy scenarios and study the changes in optimal rates as consumption patterns change over development. The scenarios are closely related to those in Section 4, but studying optimal tax policy enables us to relax the assumption that differentiating tax rates between food and non-food goods must take the form of full tax exemption of food. We study instead the optimal level of rate differentiation between these two goods categories.[40] Proofs of results are in Appendix C.

Household preferences There is a continuum of mass 1 of households i with heterogeneous exogenous incomes y^i. Households have preferences over j goods, and for each good over two varieties v, which we assume are imperfect substitutes. The subscript $v = 0$ indicates a variety produced in the informal sector, $v = 1$ a variety produced in the formal sector. In most of what follows we assume informal varieties cannot be taxed. Producer prices q_{jv} are exogenous, consumer prices are given by $p_{j1} = q_{j1}(1 + t_j)$, where

[39] Source: 1985 US Consumer Finance Survey. Savings rates are in the same range in the few developing countries (China and Chile) for which similar consumer finance surveys are available.

[40] We focus in this section on optimal rate differentiation between these two categories only for simplicity, and because of the policy relevance of this scenario. Our expressions can however be used to consider a government that optimally differentiates rates across a large number of categories; we show results from this full rate differentiation in Section 6.2 below.

t_j is the tax on good j, and $p_{j0} = q_{j0}$. These consumer prices reflect the commonly made assumption that there is full (no) pass-through of taxes to formal (informal) consumer prices.[41] We write $v(p, y^i)$ the indirect utility of household i, s^i_{jv} the budget share that household i spends on variety v of good j, $s^i_j = s^i_{j0} + s^i_{j1}$ the budget share it spends on good j, and ϵ_j the price elasticity of demand for good j.

We impose additional structure on household preferences to characterize how the efficiency cost of taxation changes along the development path. We assume that compensated price elasticities of demand for all goods are equal across households, across products and with development. We set elasticities of substitution across goods equal to zero but allow a positive cross-price elasticity of demand across varieties. This enables us to focus on households' responses that arise in the presence of an informal sector due to substitution across varieties within each good. This substitution is governed by the cross-variety price elasticities of demand, which are assumed equal across all goods and invariant along the development path. We allow for differences in income elasticities across goods and varieties but assume income effects are fixed across development.

Imposing this structure on preferences allows us to clearly determine how uncompensated price elasticities, which drive the efficiency costs of taxation, vary across products and along the development path. In Appendix C, we show that the uncompensated price elasticity of demand for a formal variety of a good, denoted ϵ_{j1}, can be expressed as a function of compensated price elasticities, income elasticities, and budget shares:

$$\epsilon_{j1} = \epsilon^C - \eta_{j1} s_{j1} - 2\tilde{\epsilon}^C \alpha_j \tag{4}$$

This elasticity captures the efficiency cost of taxing only the formal variety of good j and is composed of three components. The first is the compensated price elasticity of demand for a good, ϵ^C. The second is an income effect driven by the income elasticity of demand for the formal variety η_{j1} and its budget share s_{j1}. The third is a function of the compensated cross-price elasticity of demand, denoted $\tilde{\epsilon}^C$, and the share of informal consumption in total consumption of good j, denoted α_j.[42] Intuitively, as the price of formal varieties increases, households can substitute to informal varieties: this increases the price elasticity of the formal variety, the more so the more households are willing to substitute across varieties (higher $\tilde{\epsilon}^C$).

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

[42] This expression also assumes variety-level own-price elasticities of demand are equal across varieties for each good, and differences in prices across varieties are negligible, such that $p_{j0} \approx p_{j1}$. Section 3 presents evidence regarding the value of $\frac{p_1}{p_0}$. In all countries the difference between prices of formal and informal varieties is small, around 5%.

Government preferences The government chooses the tax rates t_j levied on each good j to maximize:

$$W = \int_i G(v(p, y^i))di + \mu \sum_j t_j q_{j1} x_{j1} \tag{5}$$

where $x_{j1} = \int_i x^i_{j1}(p, y^i)$ is total consumption of the formal variety of good j. 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 $\bar{g}$ is the average social marginal welfare weight (see Saez and Santcheva, 2016).[43] We assume that g^i is decreasing with household income and $\mu = \bar{g}$. The latter simplifies expressions and corresponds to a government that has no preference assumption for taxation unless it enables redistribution.

5.2 Optimal Tax Policy over the Development Path

In this sub-section, we consider how optimal tax policies vary with development. We model development as an increase in all households' income by the same proportional amount, so that the distribution of income across households does not change, and assume it leads to the changes in budget shares and Engel curve slopes documented in Section 3. To build intuition, and consistent with our empirical evidence, we assume that the Engel curves of all taxed goods are approximately linear with respect to log household income. We relax this assumption when we calibrate the model to our data. We first consider optimal uniform taxation in a world in which only formal varieties can be taxed. We then consider optimal rate differentiation between food and non-food goods: first in a counterfactual setting with perfect enforcement where all varieties can be taxed; then, in the realistic setting in which only formal varieties can be taxed.

5.2.1 Optimal Uniform Commodity Taxation

We start by assuming that the government levies a uniform tax rate on all products, $t_j = t$, $\forall j$, but cannot tax informal varieties. Writing $\tau = \frac{t}{1+t}$, welfare maximization yields the following expression for the optimal uniform rate:

$$\tau^* = \frac{\int_i (\bar{g} - g^i)\phi^i \frac{s^i_1}{s_1} di}{-\epsilon_1 \bar{g}} \tag{6}$$

[43] 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 $s_1 = \sum_j \int_i s_{j1}^i di$ is the aggregate budget share spent on all formal varieties, $\phi^i = \frac{y^i}{\bar{y}}$ is the ratio of household i's income relative to the average income $\bar{y}$ and ϵ_1 is the uncompensated price elasticity of demand for all formal varieties. Equation (6) shows that the optimal uniform rate is increasing in the co-variance between household income and formal budget shares: the more richer households spend on formal varieties relative to the poor, the more redistribution is obtained from taxing only formal varieties and the higher the optimal rate on those varieties. The optimal rate is also decreasing in the absolute value of the uncompensated price elasticity of demand for formal varieties: the more households respond to changes in formal prices by consuming fewer formal varieties, the higher the efficiency cost of taxing only those formal varieties.

The change in the optimal uniform rate over the development path is given by:

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

The first and second terms capture, respectively, the change in the redistributive effect and efficiency cost of taxing only formal varieties. The direction of these changes is summarized in the following proposition.

Proposition 1. ***Optimal uniform commodity taxation when only formal varieties can be taxed***

- *The redistribution gain from taxing all products uniformly is decreasing over the development path as long as: i) the formal Engel curve is upward sloping, ii) the aggregate formal budget share increases more than the slope of the formal Engel curve.*
- *The efficiency cost of taxing all products uniformly is decreasing over the development path as long as, in addition, $\tilde{\epsilon}^C > \frac{\eta_1}{2}$, where η_1 is the income elasticity of demand for all formal varieties $\tilde{\epsilon}^C$ is the cross-variety price elasticity of demand.*

Proof: see Appendix C.

The first part of Proposition 1 formalizes the intuition (outlined in Section 4) for how changes in the aggregate formal budget shares and slope of the formal Engel curve affect the redistributive effect of taxing formal varieties. As shown above, formal Engel curves are upward sloping in all countries. Among poorest countries, the likelihood that a formal variety purchase is made by rich households is high because the aggregate formal budget share is small. As this budget share increases with development, formal

purchases become a worse tag for higher household income. This decreases the redistribution gain from taxing all formal varieties, and therefore pushes the optimal rate on these varieties downwards, as long as the slope of the formal Engel curve does not increase substantially with development.

The second part of Proposition 1 states the conditions under which efficiency considerations will on the contrary push the optimal uniform rate upwards over development. The increase in aggregate formal consumption share over development lowers the opportunities for substitution towards informal varieties; this decreases the efficiency cost of taxing all formal varieties (see equation 4). At the same time, the growth in formal consumption share increases the responses to changes in prices due to income effects, which leads to higher efficiency costs. The first effect dominates as long as $\tilde{\epsilon}^C > \frac{\eta_1}{2}$. Overall, the presence of large informal sectors in poorer countries tends to increase both the redistributive gain and the efficiency cost of taxing consumption relative to richer countries with smaller informal sectors.

5.2.2 Optimal Rate Differentiation When All Varieties Can Be Taxed

We now turn to a government which sets a different rate on food and non-food goods. We start by considering rate differentiation under the assumption that the government has perfect enforcement capacity and can therefore tax both formal and informal varieties.[44] This unrealistic assumption enables us to consider how optimal rate differentiation would change over the development path in the absence of an informal sector and provides a 'no enforcement constraint' benchmark against which to compare more realistic scenarios in the following section. The optimal rate on product j is given by:

$$\tau_j^* = \frac{\int_i (\bar{g} - g^i)\phi^i \frac{s_j^i}{s_j} di}{-\epsilon_j \bar{g}} \tag{8}$$

This expression shows that the optimal rate is increasing in the co-variance between household income and budget share spent on good j. We know from Section 3 that the slopes of all non-food (food) Engel curves are positive (negative). Holding efficiency considerations constant, this implies that the optimal policy taxes food less than non-food goods.

The following proposition characterizes the change in the optimal tax on food relative to non-food over the development path.

[44] Formally we assume that t_j is levied on both x_{j1} and x_{j0}, so that $p_{j0} = q_{j0}(1 + t_j)$.

Proposition 2. ***Optimal rate differentiation when all varieties can be taxed***

- *The redistribution gain from taxing food less than non-food goods is increasing over the development path as long as: i) food Engel curves are downward sloping, ii) aggregate food budget shares decrease, iii) the aggregate food budget share is lower than the aggregate non-food budget share, iv) food Engel curves do not flatten too much.*
- *In addition, the efficiency cost of taxing food less than non-food products increases with development as long as non-food budget shares increase.*

Proof: see Appendix C.

The first part of proposition 2 states the conditions under which the redistribution gain from rate differentiation increases over development. Intuitively, for a given Engel curve slope, this redistribution gain is minimized when food and non-food aggregate budget shares are equal: in this case, observing a food or non-food purchase yields little information about a household's income, such that differentiating rates across these categories has little redistributive effect. This is the situation in the poorest countries in our sample where food and non-food goods are consumed in roughly equal proportions. As countries grow the food budget shares fall, so food purchases become a better tag for household income. This increases the optimal level of rate differentiation (decreasing the optimal rate on food relative to non-food) over the development path, as long as food Engel curves do not flatten too much. [45]

The second part of Proposition 2 states that efficiency considerations will on the contrary increase the optimal rate on food relative to non-food over development (decrease rate differentiation). The intuition for this stems from the fact that the average budget share spent on food falls while that spent on non-food products increases as countries grow. This decreases the efficiency cost of taxing food relative to non-food products due to income effects.

5.2.3 Optimal Rate Differentiation When Only Formal Varieties Can Be Taxed

Finally, we consider optimal rate differentiation under the more realistic assumption that informal varieties cannot be taxed. The optimal rate on product j when only variety $j1$ can be taxed is now given by:

[45] Condition iii) is not strictly necessary, but constitutes the relevant empirical setting we observe in Section 3: food budget shares are typically smaller than non-food budget shares in our sample. In Appendix Section C, we show that proposition 2 holds even when the food budget share is higher that the non-food budget share, as long as the slope of the food Engel curve increase sufficiently in magnitude as countries develop.

$$\tau_j^{**} = \frac{\int_i (\bar{g} - g^i)\phi^i \frac{s_{j1}^i}{s_{j1}} di}{-\epsilon_{j1}\bar{g}} \tag{9}$$

The following proposition characterizes the change in the optimal tax on food relative to non-food products over the course of development when only formal varieties can be taxed.

Proposition 3. ***Optimal rate differentiation when only formal varieties can be taxed***

- *The redistribution gain from taxing food less than non-food products is increasing over the development path as long as: i) the slope of the Engel curve for formal food varieties decreases relative to that for formal non-food varieties, ii) the aggregate budget share of formal non-food varieties does not increase much faster than the aggregate budget share of formal food varieties.*

- *The efficiency cost of taxing food less than non-food products increases with development as long as, in addition, the informal share of food consumption falls faster than that of non-food consumption.*

Proof: see Appendix C.

The first part of Proposition 3 states under what conditions equity considerations push the optimal rate of food down relative to that on non-food goods over the development path. As discussed above, the Engel curve slopes for formal food varieties are very close to zero in the poorest countries in our sample. In these countries subsidizing food relative to non-food products will therefore not necessarily be equity-improving. As countries grow however the Engel curve slopes of formal non-food varieties grow, whilst those of formal food varieties fall (see Figure 4). This change in slopes increases the redistributive gain from subsidizing food relative to non-food over the development path, as long as the aggregate budget share of formal non-food products does not increase too much with development relative to that of formal food products.

The second part of Proposition 3 states that, in contrast, efficiency considerations tend to push the optimal relative rate on food down over development. Recall from expression (4) that the efficiency cost of taxing only the formal variety of a good is increasing in the share of informal consumption in total consumption of that good. Over the development path these informal shares fall, lowering the efficiency cost of taxing both food and non-food products. They fall faster for food products than for non-food products, however

(see Figure A12). This implies that the efficiency cost of taxing formal food varieties drops faster than that of taxing non-food formal varieties as countries grow. [46]

6 Implications for Redistribution and Inequality

This section presents the implications of our results for the impact of consumption tax policy on disposable income inequality in developing countries. The extent to which a tax policy redistributes across households depends on both its progressivity (studied in Section 4) and the average effective tax rate levied. The latter is itself a function of the level of the statutory tax rates, which we obtain from our model, and the size of the tax base, which varies across tax scenarios. We calibrate the optimal tax rates in section 6.1, and then calculate the effect of different tax policy scenarios on inequality in section 6.2.

6.1 Calibrated Optimal Tax Rates

This sub-section calibrates the optimal tax rates defined in expressions (6), (8) and (9). Table 4 summarizes our choice of calibration parameters. We calibrate several parameters directly from our data: we use the observed budget shares described in section 3, total household expenditure to proxy for household income, the slopes of the Engel curves to obtain income elasticities, and the observed informal budget shares for each good and country.[47] We relax our theoretical assumptions that Engel curves are log-linear and that inequality is fixed, using instead the observed budget shares and income distributions in each country. We consider a range [1,2] for the cross-variety compensated price elasticity. This is in line with estimates in Faber and Fally (2017) and Atkin et al. (2018b), and we use 1.5 as our baseline value, while setting a value of -0.7 for the own-price compensated elasticity of goods. Together, these parameters yield values for the own-price uncompensated elasticity of goods that are in the $[-2.2, -0.7]$ range, in line with estimates from the literature (Deaton et al., 1994). Finally, we specify government preferences by setting a social welfare weight for households in each decile of the household expenditure distribution in each country. Our specification implies that governments place 10 times more weight on income received by households in the poorest decile than in the richest decile. Our calibration choices yield optimal uniform rates in

[46] In addition, it must be that the aggregate budget share of formal non-food products does not increase much faster than that of formal food products over development. This ensures that behavioral responses to taxation through income effects do not increase much faster for food products.

[47] Our model calls for using budget shares observed under a counterfactual 'no tax or transfers' scenario. We do not attempt to adjust observed budget shares to take into account the fact that they are affected by current tax systems as this would require an in-depth understanding of the tax and transfer system in each country in our sample which is beyond the scope of this paper.

the 10% to 25% range, in line with the range of statutory consumption tax rates set by developing countries. Appendix D details our calibration choices further.

Figure 7 plots the country-level ratio of optimal food to non-food rate (the relative food subsidy) as a function of economic development. The left two panels refer to the counterfactual scenario where all varieties are taxed; the right panels refer to the more realistic scenario where only formal varieties are taxed. The top panels show calibrated rates holding uncompensated price elasticities constant over development: in these figures all the variation across countries is due to varying redistribution gains from differentiating rates. The bottom panels allow price elasticities to vary with development: here, the variation is due to both redistribution gains and efficiency costs.[48]

There are two main take-aways from Figure 7. First, consistent with our model predictions, we find that over development, equity considerations tend to decrease the rate on food relative to non-food goods (thus increasing rate differentiation) but efficiency considerations tend to increase it. Equity effects dominate, so that optimal relative rates on food fall with development when both effects are taken into account. Second, comparing panels B and D, we see that once we take into account the fact that only formal varieties can be taxed, the optimal policy no longer subsidizes food relative to non-food in some poor countries (the ratio of food to non-food rates is higher than 1). Once the impossibility of taxing informal varieties is accounted for, taxing food less than non-food goods cannot by justified on equity or on efficiency grounds in these countries. Appendix Figure A14 shows that both of these findings are robust to changing values of the cross-variety price elasticity.[49]

6.2 Effect of Consumption Tax Policy on Inequality

Figure 8 presents the effects of different consumption tax scenarios on income inequality in the average country in our sample. In Panel A we use the calibrated optimal tax rates for each scenario (allowing efficiency costs to vary with development) and use these rates to calculate the net of tax income of each household in our data. Our redistribution metric is the percent change in Gini from the pre-tax income distribution to the net-of-tax distribution. To benchmark our results against estimates in the literature, Panel

[48]Formally, in the top panels we set the uncompensated price elasticities in expressions equal to -1 for all countries when calibrating expressions (6), (8) and (9). In the bottom panels we calibrate these elasticities using expression (4) and the parameter values detailed in Table 4.

[49]In Appendix Figure A13 we also show the calibrated optimal uniform rates on all formal varieties. Consistent with our theory, the optimal uniform rate falls over development for redistributive purposes. As predicted, allowing efficiency costs to vary with development lowers the optimal rates, especially in the poorest countries and more so the higher the cross-variety elasticity of substitution.

B reports the inequality impacts of actual tax policies in place in a comparable sample of developing countries obtained by Lustig (2018). To calculate these effects, we use their Commitment to Equity (CEQ) database which contains information on income and estimates of taxes paid (consumption taxes, direct taxes and social security) for each income decile in 25 developing countries.[50] Importantly, their methodology does not systematically consider the possibility that some places of purchase may be informal, and therefore cannot take into account the redistributive effect of de facto exemption of informal consumption.

Several key results emerge from Figure 8. First, our different optimal consumption tax policy scenarios achieve substantial inequality reduction. The inequality effects are large compared to estimates of the effect of current consumption tax policies, which suggest consumption taxes in developing countries achieve very little inequality reduction (0.3% reduction in Gini, row 1 of Panel B), despite taking into account reduced rates and exemptions. On the contrary, when we consider the de facto exemption of informal varieties, simply setting a uniform rate on all goods achieves a non-trivial inequality reduction (a 2.3% decrease in the Gini, row 1 of Panel A). Second, the amount of redistribution achieved by this scenario is still large when we compare it to a scenario with perfect enforcement in which governments optimally differentiate rates on food and non-food products (row 2 of Panel A). Comparing rows 1 and 2 of Panel A, we see that taxing only formal varieties achieves 75% of the inequality reduction obtained under the perfect enforcement and de jure rate differentiation scenario.[51] Third, introducing rate differentiation on top of de facto exemption of the informal sector further reduces inequality (row 3 of Panel A). Comparing rows 1 and 3, however, we see that simply setting a uniform rate on all goods achieves two-third of the redistribution obtained by optimally differentiating rates once we assume that only formal varieties can be taxed.

Finally, by comparing panel A to row 2 of Panel B, we find that our estimated effects of consumption tax policy on inequality have the same magnitude as the effect of

[50] In the first row of Panel B, we calculate the change in Gini from applying general consumption and excise taxes to disposable income plus indirect subsidies. In the second row of Panel B, we calculate the change in Gini from applying the direct tax and social security contributions to market income plus direct cash transfers. These exercises allow us to calculate the marginal Gini impacts of the indirect and direct tax systems, respectively.

[51] Optimally differentiating rates when all varieties are taxed reduces inequality substantially, despite the relatively small progressivity achieved by differentiating rates on food and non-food described in section 4.3. It occurs because this (unrealistic) scenario assumes the government can tax the entire consumption base, which yields much higher effective tax rates.

the direct tax system (income taxes and social security) in developing countries.[52] We note however that the redistributive potential of direct tax systems in these countries is constrained because they only cover the small share of the workforce which is not self-employed (Jensen, 2019). Direct tax systems that do not face this constraint have a much larger effect on inequality.[53]

6.3 Extensions and Discussion

Full rate differentiation We have thus far only considered scenarios in which governments set at most two tax rates, on food and non-food goods. How much more inequality reduction can be achieved if we allow for different rates on each of the 12 large COICOP 2-digit good categories? To answer this, we calibrate optimal tax rates at the level of each good and for each country and recompute the changes in Gini.[54] The last two rows of Table A7 display the results. In the realistic case with an informal sector (last row), the inequality reduction achieved by full rate differentiation is 30% higher than that achieved by simply differentiating rates between food and non-food goods. Given that administrative costs from managing multiple rates may be high (Ebrill et al., 2001), this result further suggests a limited role for rate differentiation across goods in developing countries.

Robustness Table A7 reports a large range of robustness checks on our inequality results. First, we show that results are broadly robust to the checks implemented in Section 4 for the progressivity results. In particular, allowing for some pass-through of taxes to prices in the informal sector has two opposite effects on inequality which tend to cancel each other out. On the one hand, it decreases the progressivity of consumption tax scenarios in which the informal sector is exempt, as seen in Section 4. On the other hand, it increases the tax base (and therefore the aggregate effective tax rates), which reduces inequality. The first effect dominates slightly, leading to marginally lower effects on inequality. Allowing for distributional saving rates reduces the inequality effects, especially in the counterfactual scenario in which all varieties are assumed to be taxed, but our main results are unchanged. Changing the rule used to assign a place of purchase to the formal or informal sector similarly leaves our key results qualitatively unchanged.

[52] Since direct taxes are generally considered most strongly suited for redistribution, our results are large in magnitude among tax policy impacts. They are, however, smaller in magnitude than the inequality reduction achieved in CEQ-countries by in-cash and in-kind transfers (Lustig, 2018).

[53] Using data from the OECD Income Distribution Database, we calculate that direct taxes in developed countries achieve a 11.2% Gini-reduction in inequality.

[54] Formally, we re-calibrate expressions (8) and (9) for each good and country.

Second, in the final two rows of Table A7 we present results using alternative values of the cross-variety price elasticity of demand. This is the parameter which governs the strength of the efficiency cost of taxation. Our results are largely unchanged.

Absence of direct tax instruments Our result that optimal indirect taxes are robustly redistributive is derived in a model where no direct tax instrument is available. A central result in public finance is that redistribution is better achieved through direct rather than indirect taxes (Atkinson and Stiglitz, 1976; Jacobs and Boadway, 2014). However, this theoretical result relies on the assumption that income taxes cannot be evaded, which is at odds with reality in developing countries (Jensen, 2019). When income taxes can be evaded, a greater redistributive role is found for indirect tax instruments (Boadway et al., 1994; Huang and Rios, 2016).[55] This discussion suggests that an extended model with direct tax instruments, even constrained, would lead to less optimal redistribution through consumption taxes. Jointly studying the optimal direct and indirect tax instruments over development would, however, require additional empirical moments, including how opportunities to evade income taxes vary along a country's income distribution and as the country develops. Such an undertaking is beyond the scope of this paper.

7 Conclusion

This paper studies how consumption patterns vary with household income both within and across countries and derive implications for the optimal design and redistributive potential of consumption taxes. We consider two channels for redistribution: the de facto tax exemption of informal expenditure and the de jure tax exemption of necessities, in particular food. To measure informal expenditure, we harmonize expenditure surveys across 31 developing countries that contain information on the store type for each transaction. We assign each store type to the informal or formal sector using a robust assignment rule, and calculate the informal budget share at the household level. This enables us to characterize the Informality Engel Curves: we find that informal budget shares decrease with household income in every country. This implies that the de facto exemption of the informal sector is progressive. We then extend the standard optimal commodity tax model to allow for informal consumption and calibrate it to our data to study the effects of different tax policies on inequality. Contrary to the consen-

[55] For example, Huang and Rios (2016) find that incorporating income tax evasion and the existence of a consumption tax reduces the optimal income tax by 28%. Note that Boadway et al. (1994) and Huang and Rios (2016) both assume consumption taxes are perfectly enforceable. Given our findings, incorporating consumption tax evasion in their models would further reinforce the redistributive role of indirect taxes.

sus view, we show that consumption taxes are redistributive and lower inequality by as much as personal income taxes in developing countries. This effect is mainly driven by the existence of the informal sector. Once informal consumption is accounted for, reduced rates on necessities have a limited impact on inequality.

Our findings have sharp implications for the use of reduced rates on necessities and on food items in particular. We find that differentiating rates across goods has limited redistributive potential once informal consumption is accounted for. In particular, these policies have no redistributive impact in some of the poorest countries. As practice shows, removing reduced rates on food and other necessities is often met with fierce resistance. An equity-improving policy would most likely have to combine the removal of reduced rates with further investments in transfer programs and social protection (Hanna and Olken, 2018).

Our results do not imply that tax administrations should abandon their efforts to reduce the size of the informal sector. Rather, they caution that any benefits from reducing the informal sector's size should be weighed against potential equity costs. Our findings call more generally for future research on tax enforcement policies to take into account not only their impact on efficiency but also their distributional effects.

Finally, we note that in most countries firms below a size threshold are exempt from taxation. This policy is motivated by the large enforcement costs tax administrations incur when trying to tax these firms, and the compliance costs to the firms themselves (Keen and Mintz, 2004). The growing availability of digital technologies lowers these costs and makes it possible to bring increasingly smaller firms into the tax net, thus removing the administrative rationale for tax exempting small firms (Gupta et al., 2017). However, our results suggest that this policy can still be justified on equity grounds.

References

Ahmad, E. and N. Stern (1984): "The Theory of Reform and Indian Indirect Taxes," *Journal of Public Economics*, 25, 259–298.

Allcott, H., B. B. Lockwood, and D. Taubinsky (2019): "Regressive Sin Taxes, with an Application to the Optimal Soda Tax," *The Quarterly Journal of Economics*, 134, 1557–1626.

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.

Almunia, M., J. Hjor, J. Knebelmann, and L. Tian (2019): "Strategic or Confused Firms? Evidence from "Missing" Transactions in Uganda," *Working Paper, Mimeo, Columbia University*.

Alvaredo, F. and L. Gasparini (2015): *Recent Trends in Inequality and Poverty in Developing Countries*, vol. 2 of *Handbook of Income Distribution*, Elsevier.

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," *NBER Working Paper*, No. 26890.

Atkin, D., B. Faber, and M. Gonzalez-Navarro (2018b): "Retail Globalization and Household Welfare: Evidence from Mexico," *Journal of Political Economy*, 126, 1–73.

Atkinson, A. and J. Stiglitz (1976): "The Design of Tax Structure: Direct versus Indirect Taxation," *Journal of Public Economics*, 6.

Banks, J., R. Blundell, and A. Lewbel (1997): "Quadratic Engel Curves and Consumer Demand," *Review of Economics and Statistics*, 79, 527–539.

Basri, M. C., M. Felix, R. Hanna, and B. A. Olken (2019): "Tax Administration vs. Tax Rates: Evidence from Corporate Taxation in Indonesia," *NBER Working Papers*, No. 26150.

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., M. Marchand, and P. Pestieau (1994): "Towards a Theory of the Direct-Indirect Tax Mix," *Journal of Public Economics*, 55.

Boadway, R. and M. Sato (2009): "Optimal Tax Design and Enforcement with an Informal Sector," *American Economic Journal: Economic Policy*, 1, 1–27.

Brockmeyer, A. and M. Hernandez (2019): "Taxation, Information, and Withholding: Evidence from Costa Rica," *Working Paper, The World Bank*.

Bronnenberg, B. J. and P. B. Ellickson (2015): "Adolescence and the Path to Maturity in Global Retail," *Journal of Economic Perspectives*, 29, 113–34.

Caspersen, E. and G. Metcalf (1994): "Is a Value-Added Tax Regressive? Annual versus Lifetime Incidence Measures," *National Tax Journal*, 47.

Clements, B., R. Mooij, S. Gupta, and M. Keen (2015): *Fiscal Redistribution in Developing Countries: Overview of Policy Issues and Options*, International Monetary Fund, chap. 4.

Coady, D. (2006): *The Distributional Impacts of Indirect Tax and Public Pricing Reforms*, vol. 2 of *Analyzing the Distributional Impact of Reforms: A Practitioner's Guide to Pension, Health, Labor Market, Public Sector Downsizing, Taxation, Decentralization and Macroeconomic Modeling*, The World Bank.

Creedy, J. (2001): "Indirect Tax Reform and the Role of Exemptions," *Fiscal Studies*, 22, 457–486.

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. (1997): *The Analysis of Household Surveys: a Microeconometric Approach to Development Policy*, The World Bank.

Deaton, A., K. Parikh, and S. Subramanian (1994): "Food Demand Pattern and Pricing Policy in Maharashtra: An Analysis Using Household Level Survey Data," *Sarvekshana*, 17, 11–34.

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.

Ebrill, L. P., M. Keen, J.-P. Bodin, V. P. Summers, et al. (2001): *The Modern VAT*, International Monetary Fund.

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," *NBER Working Paper*, No. 23101.

Feldman, N. E. and J. Slemrod (2007): "Estimating Tax Noncompliance with Evidence from Unaudited Tax Returns," *The Economic Journal*, 117, 327–352.

Gemmell, N. and O. Morrissey (2005): "Distribution and Poverty Impacts of Tax Structure Reform in Developing Countries: How Little We Know," *Development Policy Review*, 23, 131–144.

Gerard, F. and G. Gonzaga (2016): "Informal Labor and the Efficiency Cost of Social Programs: Evidence from the Brazilian Unemployment Insurance Program," *NBER Working Paper*, No. 22608.

Gordon, R. and W. Li (2009): "Tax Structures in Developing Countries: Many Puzzles and a Possible Explanation," *Journal of Public Economics*, 93, 855 – 866.

Gupta, S., M. Keen, A. Shah, and G. Verdier, eds. (2017): *Digital Revolutions in Public Finance*, International Monetary Fund.

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," *IFS Working Paper*, W18/11.

Hsieh, C.-T. and P. Klenow (2014): "The Life Cycle of Plants in India and Mexico," *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.

Jacobs, B. and R. Boadway (2014): "Optimal Linear Commodity Taxation under Optimal Non-Linear Income Taxation," *Journal of Public Economics*, 117.

Jaravel, X. (2018): "The Unequal Gains from Product Innovations: Evidence from the U.S. Retail Sector," *The Quarterly Journal of Economics*, 134, 715–783.

Jenkins, G. P., H. P. Jenkins, and C. Y. Kuo (2006): "Is the Value Added Tax Naturally Progressive?" *Working Paper, Economics Department, Queen's University*, No. 1059.

Jensen, A. (2019): "Employment Structure and the Rise of the Modern Tax system," *NBER Working Paper*, No. 25502.

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 Paper*, No. 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.

Londoño-Velez, J. (2020): "Can Wealth Taxation Work in Developing Countries? Quasi-Experimental Evidence from Colombia," *Working Paper*.

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*, No. 03/232.

Naritomi, J. (2019): "Consumers as Tax Auditors," *American Economic Review*, 109, 3031–72.

Pissarides, C. A. and G. Weber (1989): "An Expenditure-Based Estimate of Britain's Black Economy," *Journal of Public Economics*, 39, 17 – 32.

Pomeranz, D. (2015): "No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax," *American Economic Review*, 105, 2539–69.

Poterba, J. (1989): "Lifetime Incidence and the Distributional Burden of Excise Taxes," *American Economic Review, Papers and Proceedings*, 79.

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*, No. 339.

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. (1983): "How Much Redistribution is Possible through Commodity Taxes?" *Journal of Public Economics*, 20, 89–101.

Shah, A. and J. Whalley (1991): "Tax Incidence Analysis of Developing Countries: An Alternative View," *The World Bank Economic Review*, 5, 535–552.

Srinivasan, P. (1989): "Redistributive Impact of 'Optimal' Commodity Taxes: Evidence from Indian Data," *Economics Letters*, 30, 385–388.

Tanzi, V. (1998): "Fundamental Determinants of Inequality and the Role of Government," *International Monetary Fund*.

Ulyssea, G. (2018): "Firms, Informality, and Development: Theory and Evidence from Brazil," *American Economic Review*, 108, 2015–47.

——— (2020): "Informality: Causes and Consequences for Development," *Annual Review of Economics*, 12, null.

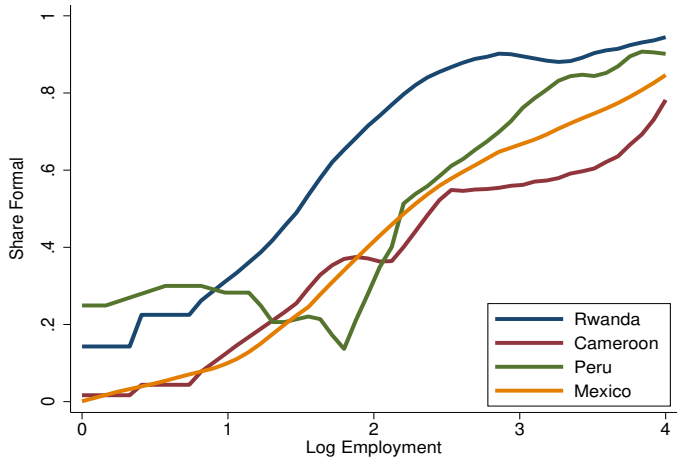
Waseem, M. (2020): "Overclaimed Refunds, Undeclared Sales, and Invoice Mills: Nature and Extent of Noncompliance in a Value-Added Tax," *CESifo Working Paper*, No. 8231.

Weigel, J. L. (2019): "The Participation Dividend Of Taxation: How Citizens In Congo Engage More with the State When it Tries to Tax Them," *forthcoming Quarterly Journal of Economics*.

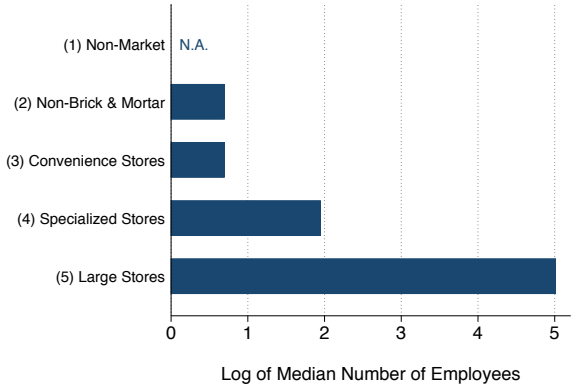
Working, H. (1943): "Statistical Laws of Family Expenditure," *Journal of the American Statistical Association*, 38, 43–56.

Figure 1: Employment Size, Store Types & Formality

(a) # Employees on Formality in Retail Censuses



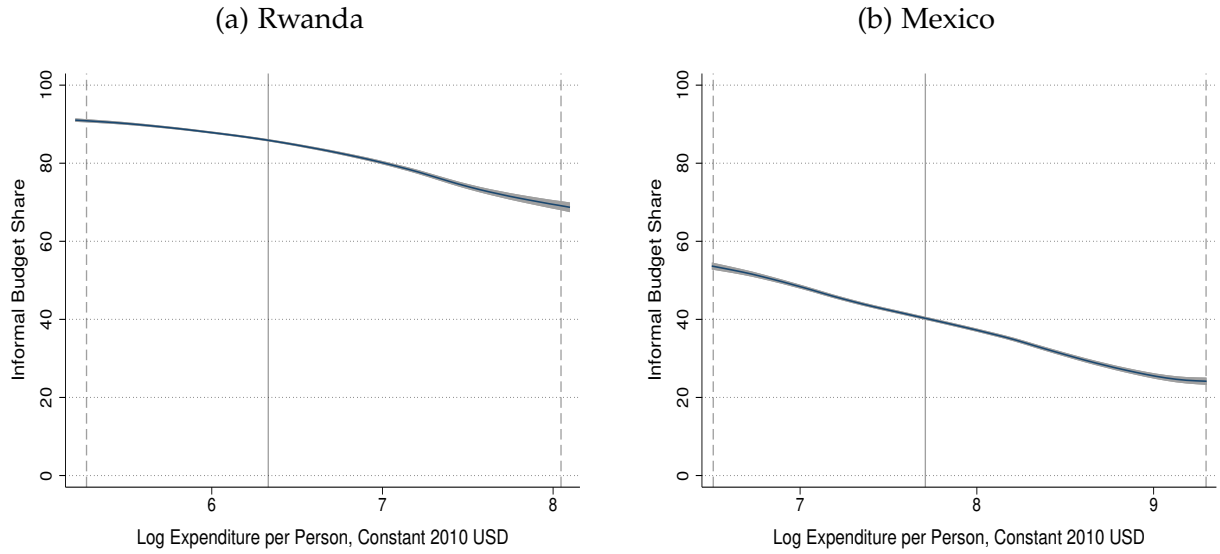
(b) # Employees by Store in Mexico



(c) % Paying VAT by Store in Mexico

These panels shows the association between formality status, employment and firm type. Panel A shows the share of formal firms as a function of log employment, using retail censuses of four core sample countries (Cameroon, Mexico, Peru and Rwanda). Formality is defined as 'being registered with the tax authority' in Cameroon and Rwanda, and 'paying Value-Added-Taxes on sales' in Mexico and Peru. Panels B and C use the 2013 Mexican retail census, which classifies retailers in similar categories as our broad place of purchase taxonomy. The figures show the log median number of employees by place of purchase (Panel C) and the share of firms paying Value-Added-Taxes on sales (Panel C). The data comes from the following firm censuses, keeping only firms which operate in the retail sector: Recensement Général des entreprises 2016 (Cameroun), Censo Económico 2014 (Mexico), Censo Nacional Económico 2008 (Peru), Establishment Census 2011 (Rwanda).

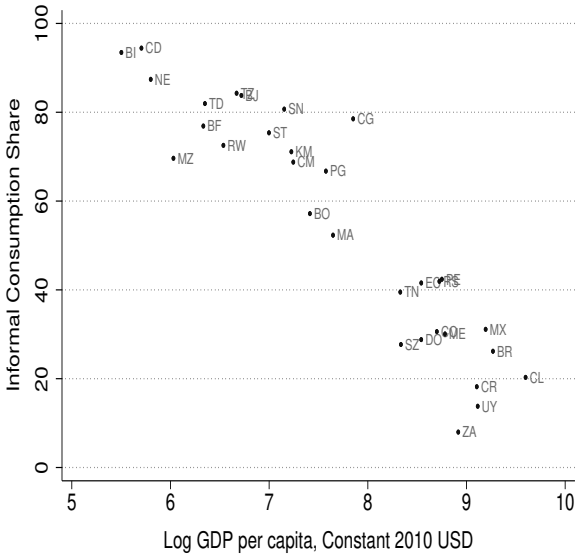
Figure 2: Selected Informality Engel Curves



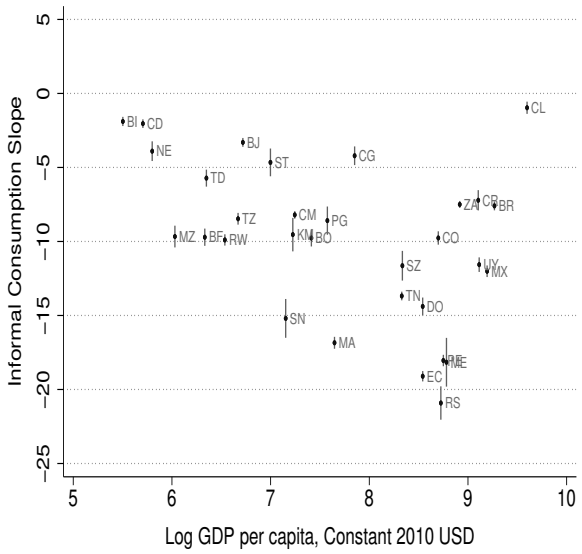
These panels show the local polynomial fit of the Informality Engel Curves in Rwanda (Panel A) and Mexico (Panel B). Per person total expenditure on the horizontal axis is measured in log. Informal budget share is on the vertical axis. The shaded area around the polynomial fit corresponds to the 95% 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 A1 for each core sample country's informality Engel curve.

Figure 3: Informal Expenditure Across Countries

(a) Informal Budget Share



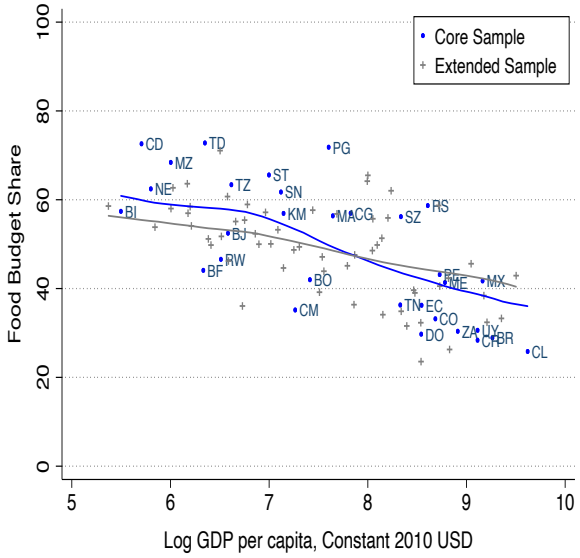
(b) Informality Engel Curve Slope



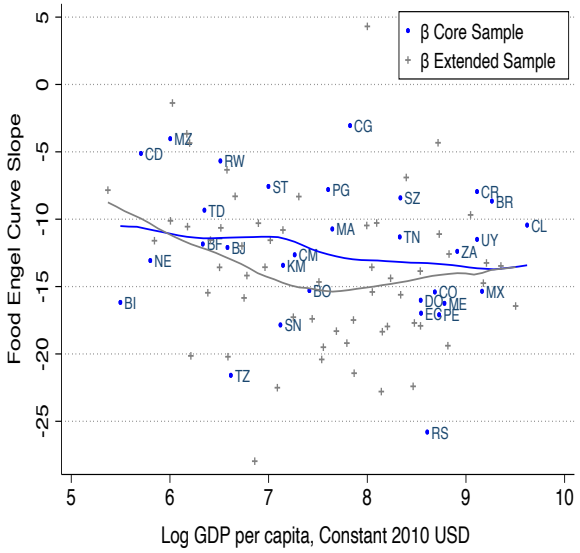
Panel A plots country-level informal budget shares on log per capita GDP. The average informal budget share is 52%. Panel B shows the slope of the informality Engel curves on log per capita GDP. The average slope is 9.8. The bars correspond to the 95% confidence interval of the slope coefficient. GDP per capita is in constant 2010 USD (Source: World Bank WDI).

Figure 4: Food Expenditure Across Countries

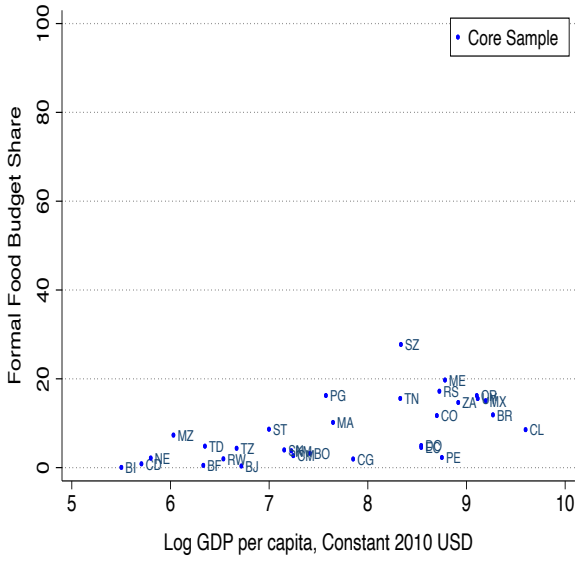
(a) Food Budget Share



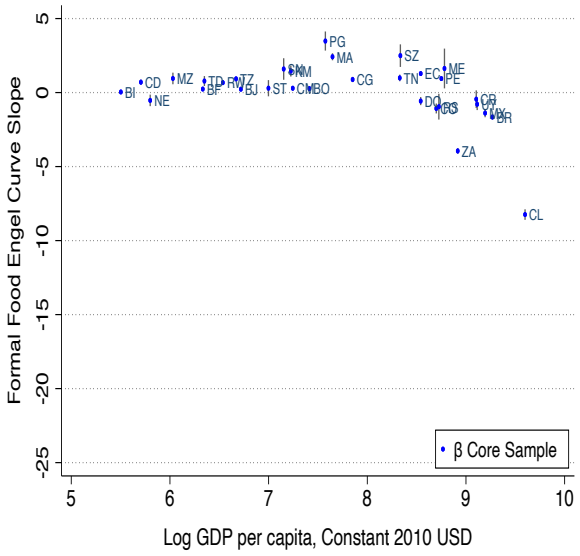
(b) Food Engel Curve Slope



(c) Formal Food Budget Share

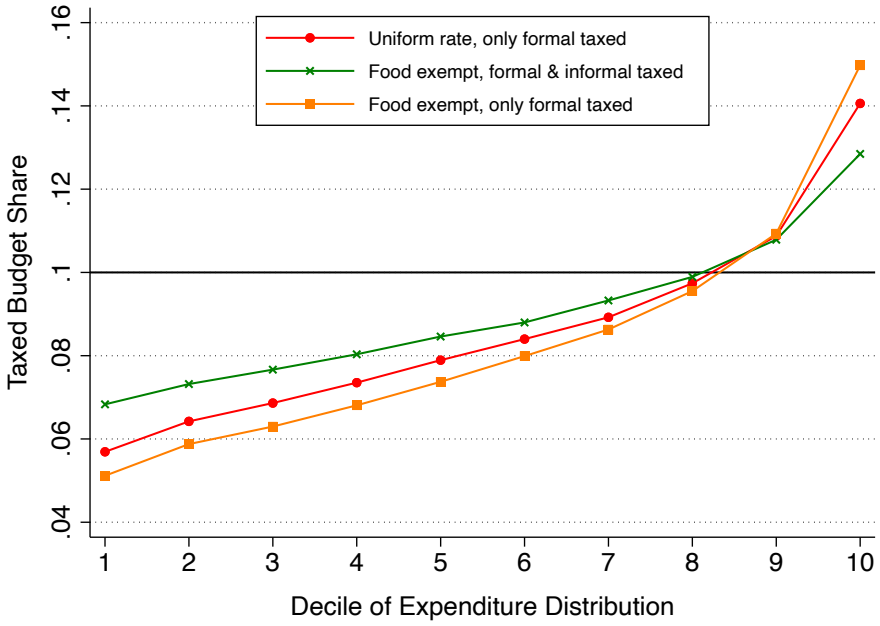


(d) Formal Food Engel Curve Slope



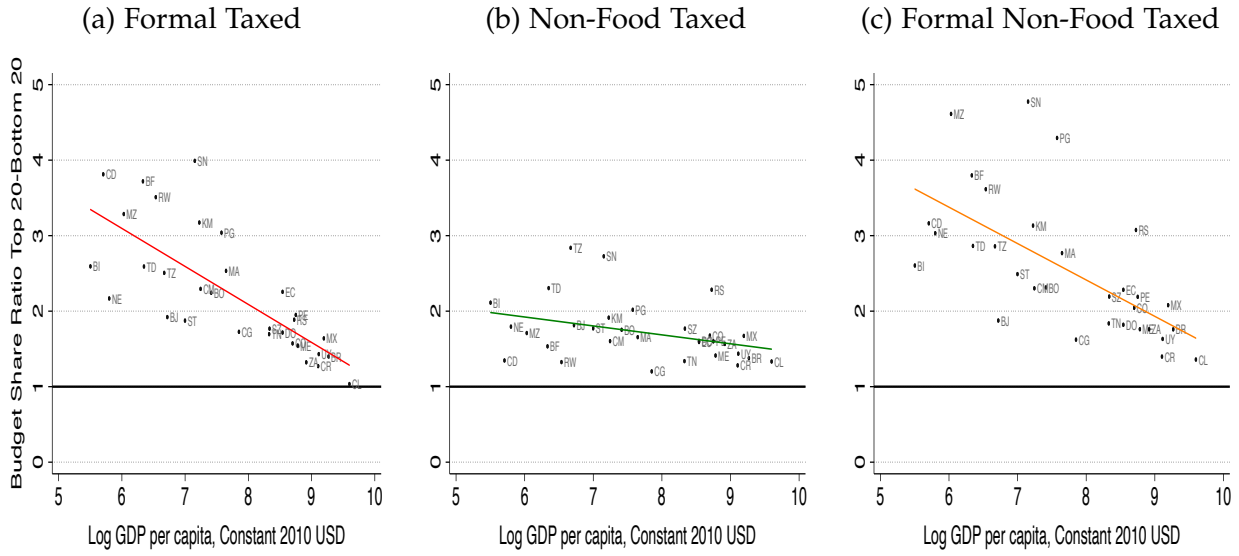
This figure combine two sources: data from the core-sample of 31 countries and data from the Global Consumption Database (GCD) which adds 58 developing countries not included in the core-sample. Panel A shows each country's food budget share, plotted against log per capita GDP. The average food budget share in the core sample is 49%, while the average in the GCD sample is 48%. Panel B shows the country-specific slope of the food Engel curve, plotted against log per capita GDP. The average slope in the core sample is 12.5, while the average in the GCD sample is 13. The lines correspond to local polynomial fits. GDP per capita is in constant 2010 USD (Source: World Bank WDI). Panels C and D are constructed similarly to Panel A and B, but for formal food expenditure which can only be measured in the core sample.

Figure 5: Progressivity of Tax Policy Scenarios



This figure plots the share of expenditures that is paid in taxes (effective tax rates), by decile, for each tax scenario. The three scenarios are simulated in all 31 countries, each point corresponds to the average effective tax rates of each decile across all countries in our sample. The red-circle line corresponds to a tax scenario where a uniform tax is levied on all goods, but where purchases in informal stores are de facto not taxed (scenario #1 defined in Section 4). The green-cross line corresponds to a scenario where food purchases are de jure exempt, but where formal and informal stores are taxed (scenario #2 in Section 4). The green-square line corresponds to a tax scenario where both de facto informality exemption and de jure food exemption are present (scenario #3 in Section 4).

Figure 6: Progressivity over Development

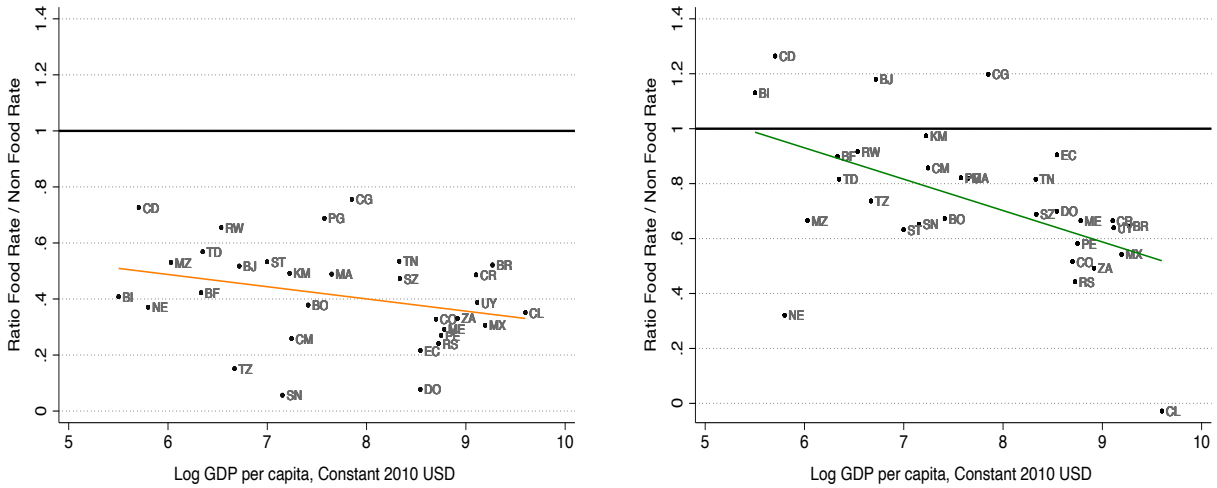


These panels plot the ratio of effective tax rate paid by the richest quintile relative to the poorest quintile in each country as a function of the country's level of economic development. The left panel corresponds to a scenario where a uniform tax is levied on all goods, but purchases in informal stores are de facto not taxed (scenario #1 in Section 4). The middle panel corresponds to a scenario where food goods are de jure exempt, but purchases in both formal and informal stores are taxed (scenario #2 in Section 4). The right panel combines de facto informality exemption and de jure food exemption (scenario #3 in Section 4). In each panel, the ratio is plotted for all 31 countries of our sample against log GDP per capita.

Figure 7: Optimal Rate Differentiation over Development

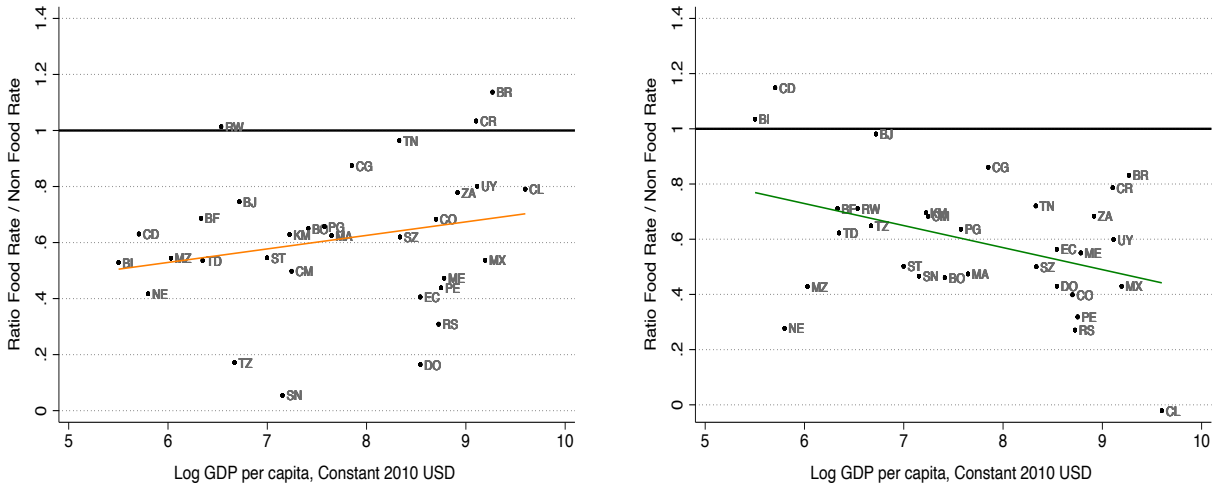
(a) All Varieties Taxed, No Efficiency Change

(b) Only Formal Varieties Taxed, No Efficiency Change



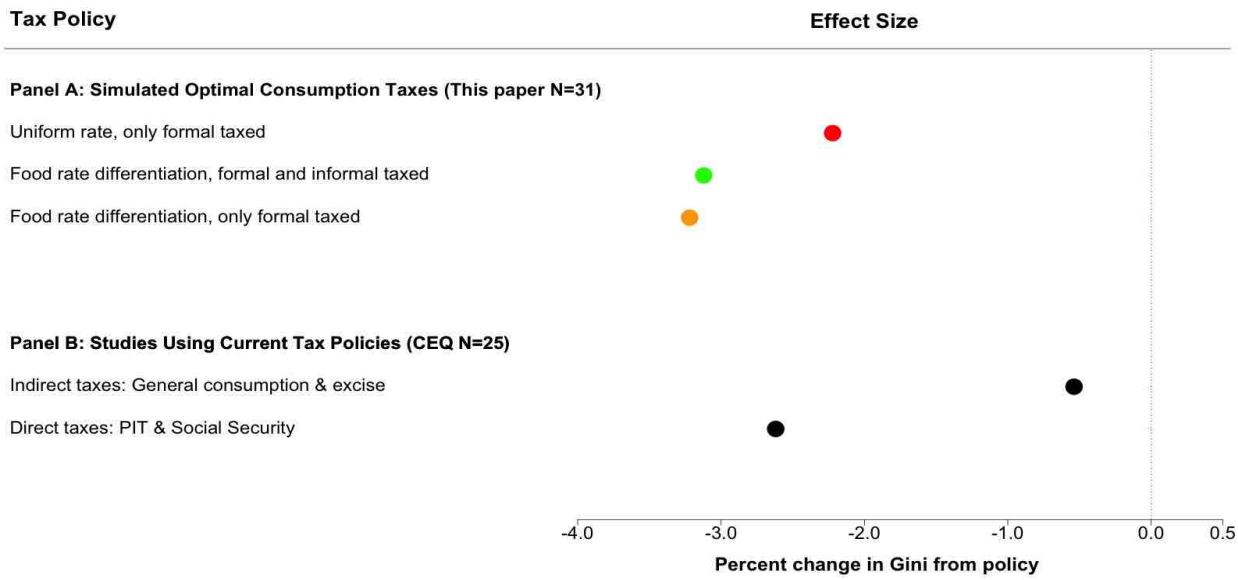
(c) All Varieties Taxed, With Efficiency Change

(d) Only Formal Varieties Taxed, With Efficiency Change



All panels plot for each country the ratio of the calibrated optimal rate on food products relative to the optimal rate on non-food products, as a function of that country's log GDP per capita. Optimal rates are calibrated using expressions (8) (panels A and C) and (9) (panels B and D). A value equal to 1 indicates that both optimal rates are set at the same level, a lower value indicates that it is optimal to subsidize food products relative to non-food products. In panels A and C we assume that all varieties can be taxed; in panels B and D, we assume that only formal varieties can be taxed. In panels A and B we hold efficiency considerations constant by assuming that uncompensated price elasticities of demand are equal across goods and countries, while in panels C and D we allow price elasticities to vary across goods and countries by calibrating values using expression (4).

Figure 8: Inequality Reduction: Optimal Policy and Current Tax Policies



Panel A shows the average percent-change in Gini for different scenarios applied to the countries of this paper's core sample. The red dot represents the scenario where a uniform rate is implemented but only the formal sector is taxed. The green dot represents the scenario where only non-food items, but both sectors are taxed. The orange dot represents the scenario where only non-food items, and only the formal sector are taxed. The reported effects reflect the change in Gini from the pre-tax income distribution to the post-tax distribution. Panel B show the percent-change in Gini using data from the Commitment to Equity Institute (CEQ). In the first row of Panel B, the pre-tax income measure is disposable income, and actual general consumption and excise taxes are applied; in the second row of Panel B, the pre-tax income measure is market income plus direct transfers, and actual personal income taxes and compulsory social security contributions are applied. Adding transfers to the pre-tax income measure is commonly done in analyses of the marginal distributional effects of tax policies. Effects calculated in Panel B are based on country and income-decile data, publicly available, and released under Lustig (2018).

Table 1: Household Expenditure Surveys for Core Sample

Country	Code	Survey	Year	GDP per capita	Sample size	Nb items/HH
Benin	BJ	EMICOV	2015	828	19871	32
Bolivia	BO	ECH	2004	1658	9149	49
Brazil	BR	POF	2009	10595	56025	41
Burkina Faso	BF	EICVM	2009	563	8404	152
Burundi	BI	ECVM	2014	245	6681	90
Cameroon	CM	ECAM	2014	1400	10303	81
Chad	TD	ECOSIT	2003	572	6697	94
Chile	CL	EPF	2017	14749	15239	129
Colombia	CO	ENIG	2007	5999	42373	60
Comoros	KM	EDMC	2013	1373	3131	82
Congo DRC	CD	E123	2005	301	12098	107
Congo Rep	CG	ECOM	2005	2569	5002	85
Costa Rica	CR	ENIGH	2014	8994	5705	68
Dominican Rep	DO	ENIGH	2007	5121	8363	88
Ecuador	EC	ENIGHUR	2012	5122	39617	89
Eswatini	SZ	HIES	2010	4169	3167	44
Mexico	MX	ENIGH	2014	9839	19479	61
Montenegro	ME	HBS	2009	6516	1223	149
Morocco	MA	ENCDM	2001	2095	14243	90
Mozambique	MZ	IOF	2009	416	10832	221
Niger	NE	ENCBM	2007	330	4000	192
Papua NG	PG	HIES	2010	1949	3810	111
Peru	PE	ENAHO	2017	6315	43545	78
Rwanda	RW	EICV	2014	690	14416	54
Sao Tome	ST	IOF	2010	1095	3545	100
Senegal	SN	EDMC	2008	1278	2503	299
Serbia	RS	HBS	2015	6155	6531	105
South Africa	ZA	IES	2011	7455	25328	44
Tanzania	TZ	HBS	2012	788	10186	318
Tunisia	TN	ENBCNV	2010	4142	11281	139
Uruguay	UY	ENIGH	2005	9079	7043	77

This table lists alphabetically the 31 countries in the core sample, the survey names and years. GDP per capita is in PPP USD in the year of the survey, obtained from the World Bank Development Indicators. Code refers to the country acronym which we use in the figures. 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.

Table 2: Average Slopes of the Informality Engel Curves

Specification:	**Main**		**Geography**		**Product Codes**				**All**
Avg. of 31 Countries	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(Negative of) Slope	9.8	10.6	9.2	8.5	6.9	6.3	6.1	5.4	4.3
Confidence Interval	[9.2,10.4]	[9.9,11.2]	[8.5,9.9]	[7.7,9.2]	[6.2,7.4]	[5.7,6.7]	[5.5,6.5]	[4.8,5.7]	[3.7,4.7]
# of p-values < 0.05	31	31	31	30	30	29	30	29	28
R^2 adjusted	0.19	0.21	0.25	0.41	0.43	0.51	0.51	0.50	0.54
Household Characteristics		X	X	X	X	X	X	X	X
Urban/Rural			X						
Survey Blocks				X					X
Food Products					X				
COICOP 2-dig						X			
COICOP 3-dig							X		
COICOP 4-dig								X	X

This table shows the (negative) average slope of the Informality Engel Curves across countries for different specifications. In column 1, we report the slopes that are estimated from the following regression: $Share\ Informal_{ip} = \beta_0 + \beta_1 ln(expenditure_i) + \varepsilon_{ip}$ where $Share\ Informal_{ip}$ is the share of household i's informal expenditure on product p. We weigh each observation using household survey weights and the expenditure share of the product. Average of lower and upper bound of 95% confidence intervals in brackets, from robust standard errors. In column 2, we augment this regression with controls for household characteristics (household size, age, gender, education of head). In column 3 (4), we instead add fixed effects for urban/rural (survey enumeration blocks). In column 5, we instead add fixed effects for food versus non-food products. In column 6/7/8, we instead add fixed effects for product codes at 2nd/3rd/4th level of the COICOP classification. In column 9, we add household characteristics, as well as fixed effects for survey blocks and COICOP-4.

Table 3: Main Reason for Choosing Place of Purchase

	Outcome: Share of purchases (in %)		
Reason	**All Stores**	**Informal Stores**	**Formal Stores**
Access	41.5	42.1	31.3
Price	28.6	29.4	17.7
Quality	13.4	11.8	40.6
Store Attributes	6.9	6.9	5.0
Other	9.6	9.8	5.5

The table reports for each potential reason households report for using a particular place of purchase, the share of purchases associated with this reason. Each number is an average across the six countries in our core sample in which the household survey asks this questions. These countries are Benin, Burundi, Comoros, Congo Rep., Morocco and RD Congo. In all surveys seven reasons are listed which we classify into five categories as follows: access is defined as "The retailer is closer or more convenient" and "The good or service cannot be found elsewhere", price as "The good or services are cheaper", quality as "The goods or services are of better quality", store attributes as "The retailer offers credit" and "The retailer is welcoming or is a friend" and other as "Others reasons". Note that Morocco has a few additional small categories, which pertain to attributes of retailer. The table lists the frequency for all purchases of goods and excludes services, which are less comparable along these dimensions, although their inclusion does not impact the results.

Table 4: Baseline Calibration Parameters

Parameter	Value	Justification
Budget shares s_j^i and s_{j1}^i	Varying	Observed in our data
Household income (scaled) ϕ^i	Varying	Observed in our data
Income elasticities of goods η_j	Food: 0.65, Non-food: 1.2	From our data, using $\eta_j = 1 + \frac{\beta_j}{s_j}$[1]
Income elasticities of formal varieties η_{j1}	Food: 1.14, Non-food: 1.31, All goods: 1.25	From our data, using $\eta_{j1} = 1 + \frac{\beta_{j1}}{s_{j1}}$[1]
Informal share of consumption α_j	Varying	From our data
Cross-variety compensated elasticity $\tilde{\epsilon}^C$	1.5	Faber and Fally (2017); Atkin et al. (2018b)[2]
Own-price compensated elasticity ϵ^C	-0.7	Deaton et al. (1994)[3]
Government preferences g^i	1-10	Uniform tax rates in the $[0.10, 0.25]$ range[4]

[1] The parameter β_j (β_{j1}) refers to the estimated slope of the Engel curve for good j (variety $j1$).
[2] For the cross-variety price elasticity we use estimates of the elasticity of substitution σ across store types in consumption obtained by Faber and Fally (2017); Atkin et al. (2018b) which are in the $[2, 4]$ range. With a CES utility function $\tilde{\epsilon}^C = \sigma s_0$ where s_0 is the aggregate budget share spent in the informal sector, equal to 0.5 on average in our sample.
[3] Our choice of value for ϵ^C together with our estimated income elasticities and observed budget shares yield uncompensated own-price elasticities for goods in the $[-2, -0.5]$ range, in line with estimates obtained by Deaton et al. (1994) in the developing country context.
[4] We set $g^i = 10$ for the first decile, $g^i = 9$ for the second decile, $g^i = 8$ for the third decile, ..., $g^i = 1$ for the tenth decile. This, together with our other calibration choices, yields optimal uniform rates when we assume only formal varieties can be taxed in the 10-25% range, in line with observed consumption tax rates in developing countries.

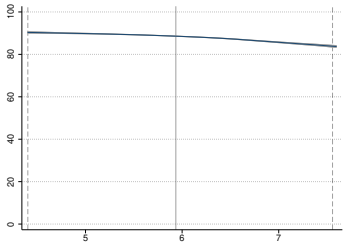
Online Appendix for 'Informality, Consumption Taxes and Redistribution'

Pierre Bachas, Lucie Gadenne & Anders Jensen

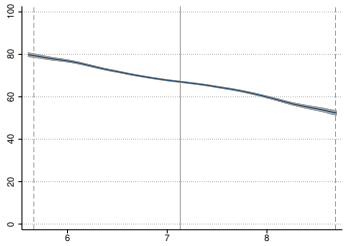
May 2020

A Additional Figures and Tables

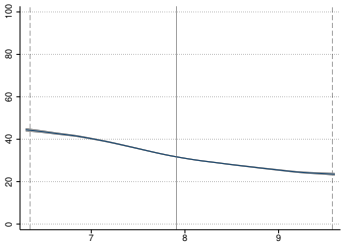
Figure A1: Informality Engel Curves



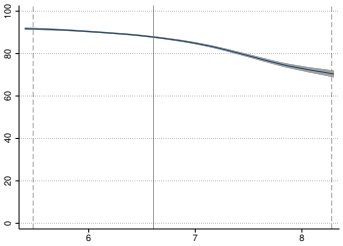
(a) Benin



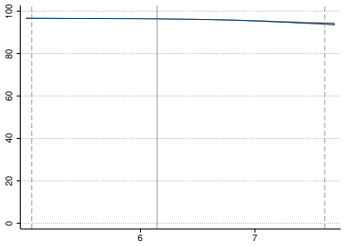
(b) Bolivia



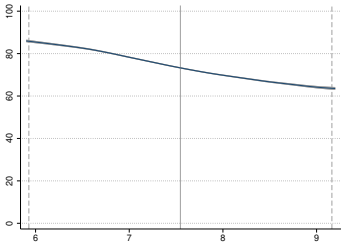
(c) Brazil



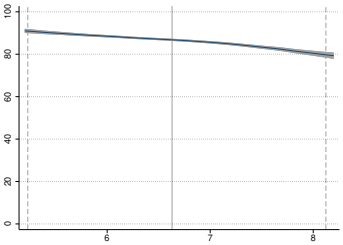
(d) Burkina Faso



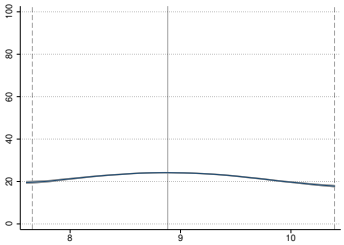
(e) Burundi



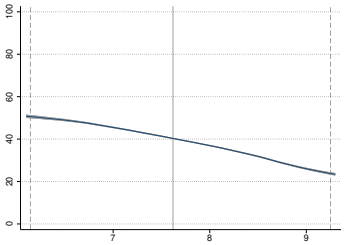
(f) Cameroon



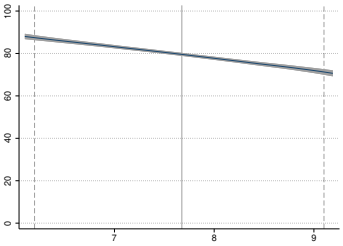
(g) Chad



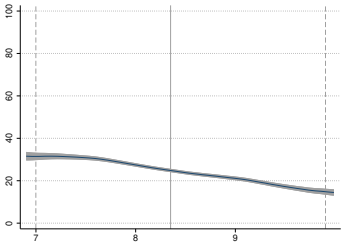
(h) Chile



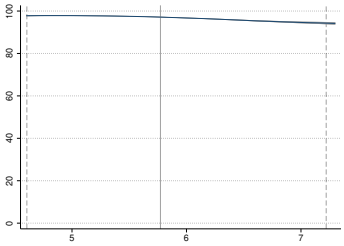
(i) Colombia



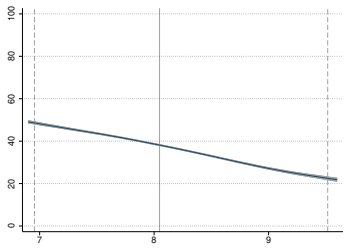
(j) Comoros



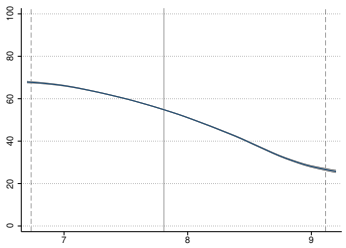
(k) Costa Rica



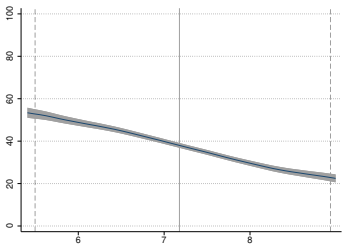
(l) Dem. Rep. of Congo



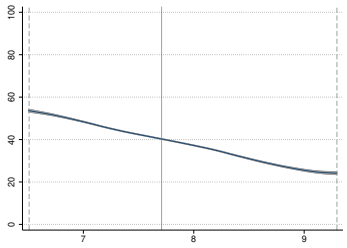
(m) Dominican Rep.



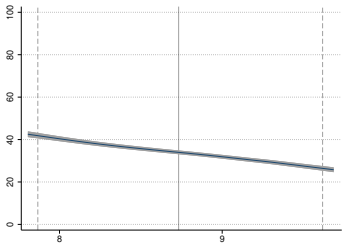
(n) Ecuador



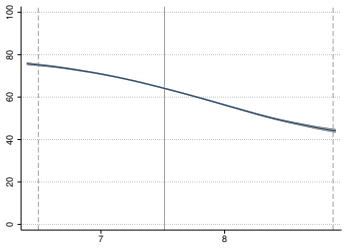
(o) Eswatini



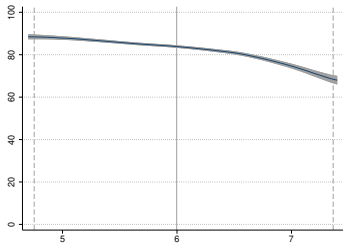
(p) Mexico



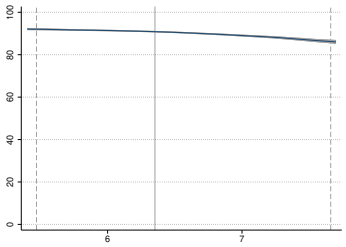
(q) Montenegro



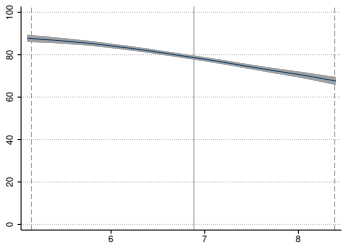
(r) Morocco



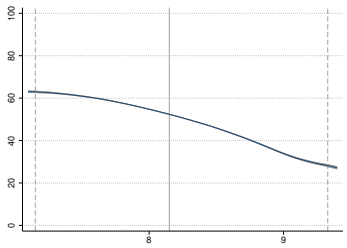
(s) Mozambique



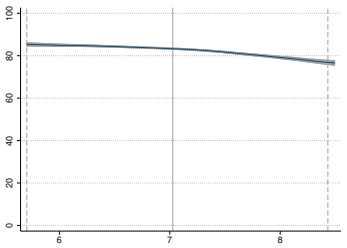
(t) Niger



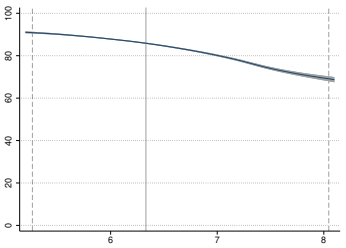
(u) Papua New Guinea



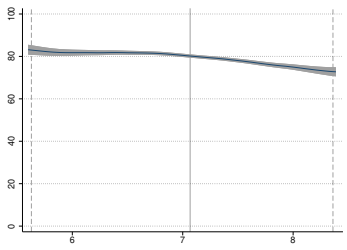
(v) Peru



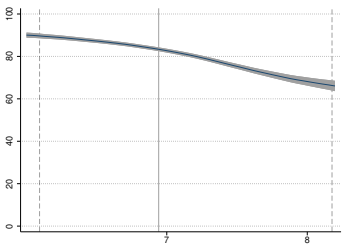
(w) Rep. of Congo



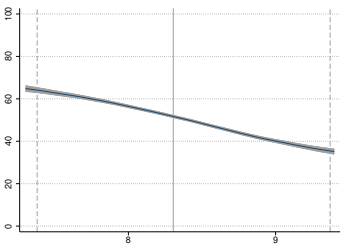
(x) Rwanda



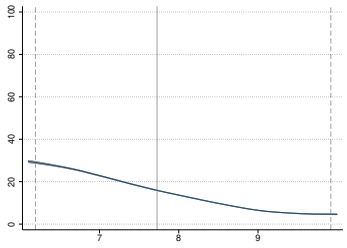
(y) Sao Tome



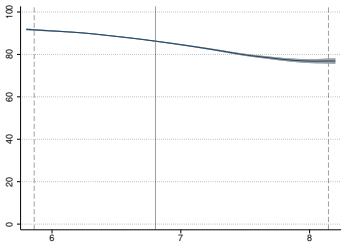
(z) Senegal



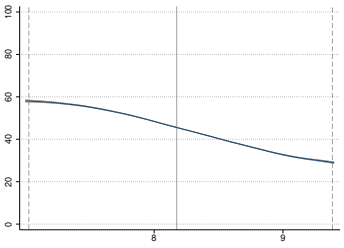
(aa) Serbia



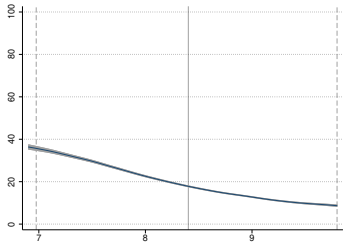
(ab) South Africa



(ac) Tanzania



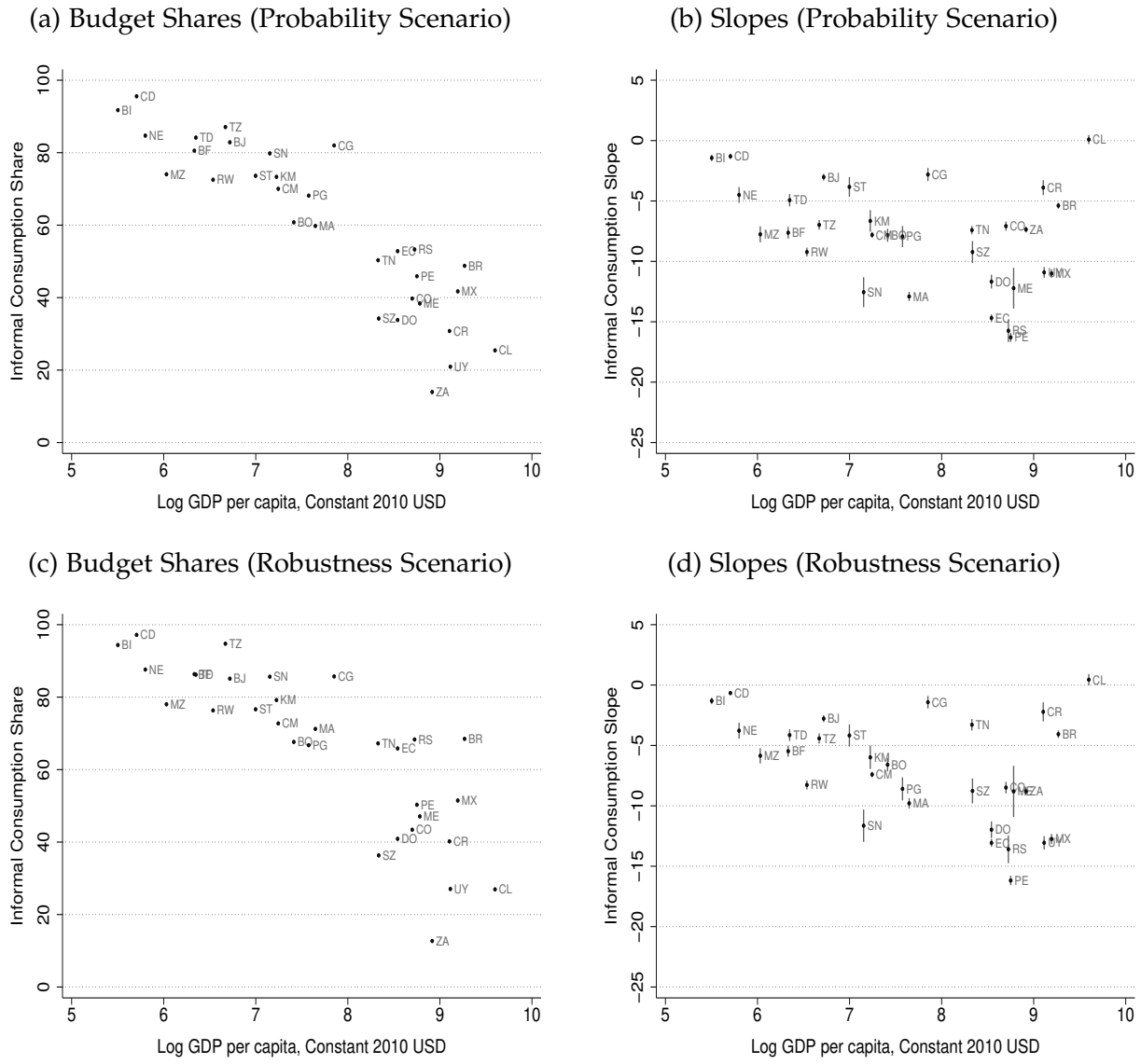
(ad) Tunisia



(ae) Uruguay

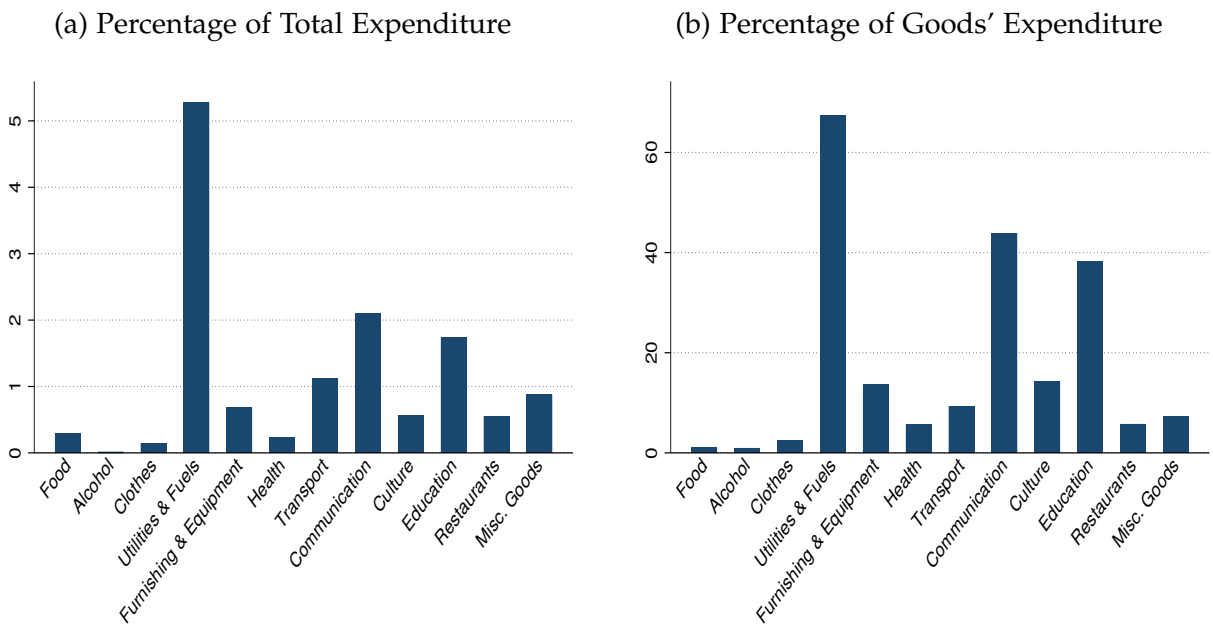
Local polynomial fit of the Informality Engel Curves in all 31 core sample countries. Per person total expenditure on the horizontal axis is measured in log. Informal budget share is on the vertical axis. The shaded area around the polynomial fit corresponds to the 95% confidence interval. The solid grey line corresponds to the median of each country's expenditure distribution, while the dotted lines correspond to the 5th and 95th percentiles. The construction of informality Engel curves is presented in section 3.1.

Figure A2: Alternative Assignment Scenarios for Informal Expenditure



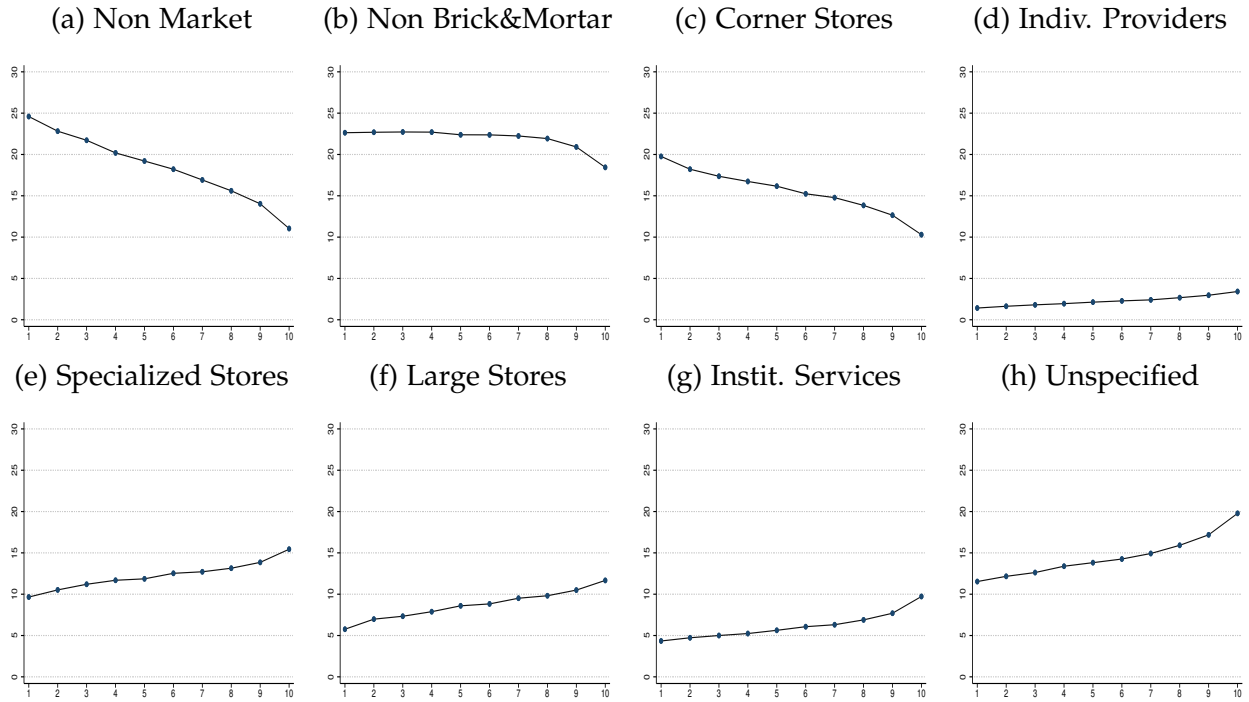
This figure is the equivalent of figure 3 for the two alternative assignments of store types to formality. The Probability scenario uses the observed probability of VAT payment by store type in the Mexican census as the formality probability of each store type across countries (see Figure 1). The robustness scenario differs from the central scenario by assigning specialized stores to the informal sector, in addition to maintaining corner stores, non brick and mortar and self-production in the informal sector. Panel (a) and (c) show informal budget shares, on log per capita GDP. Panel (b) and (d) show the the slope of informality Engel curves on log per capita GDP. The bars correspond to the 95% confidence interval of the slope coefficient. GDP per capita is in constant 2010 USD (Source: World Bank WDI).

Figure A3: Unspecified Places of Purchase by Good



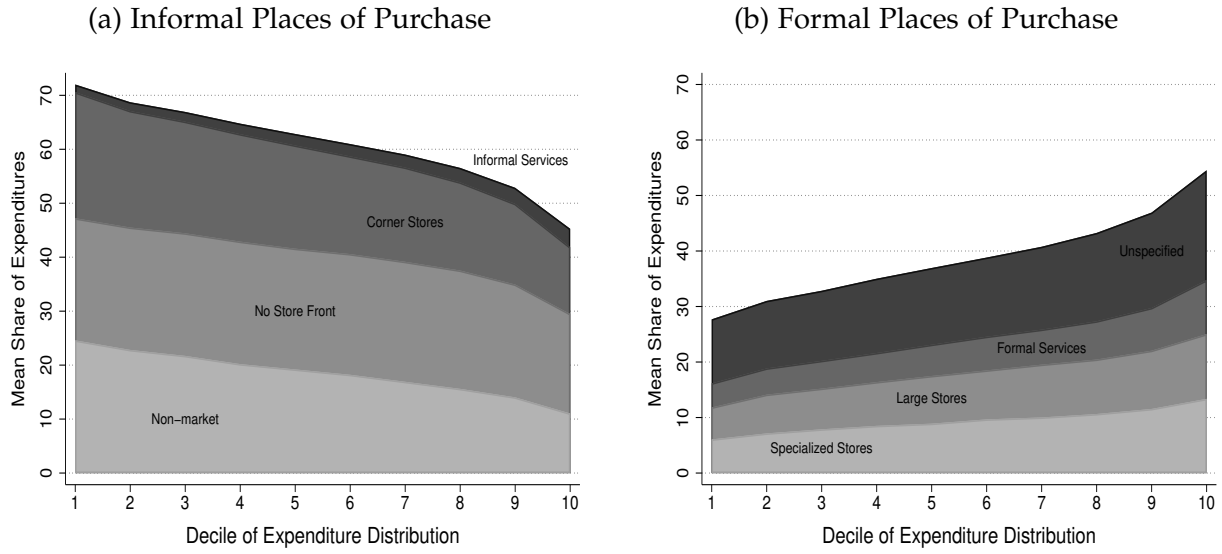
This figure shows the share of expenditures with an unspecified place of purchase by goods categories (COICOP-2 digit level) on average, across the 31 countries of the core sample, discussed in section 2.2. Panel (a) shows this share as a percentage of total expenditures and Panel (b) as a percentage of each goods' total expenditure.

Figure A4: Average Expenditure of Each Decile by Place of Purchase



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

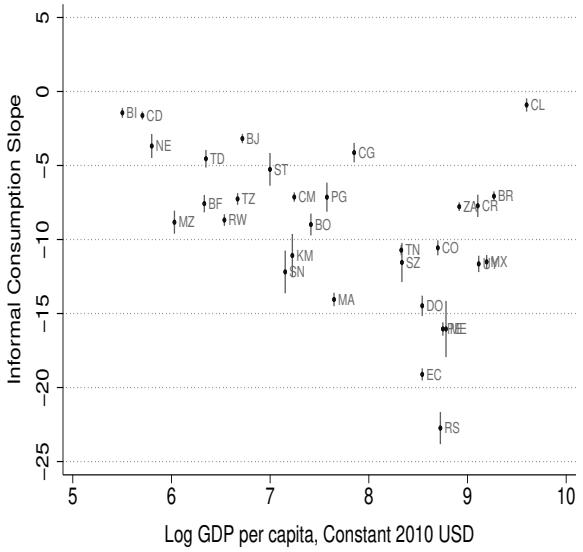
Figure A5: Average Expenditure of Each Decile By Formality Assignment



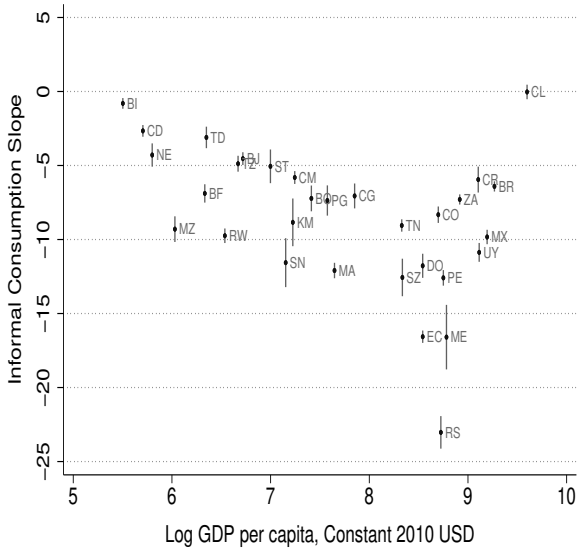
This figure shows the average expenditure of each decile across countries by type of retailer, following the retailer taxonomy described in section 2.2. Panel (a) shows the places of purchase classified as informal and Panel (b) shows the places of purchase classified as formal in the central scenario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

Figure A6: Informality Engel Curve Slopes Controlling for Geography

(a) Control: Rural Location



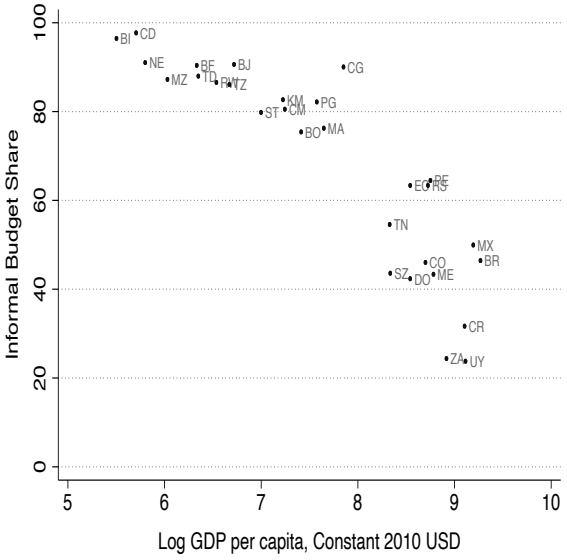
(b) Control: Survey Block



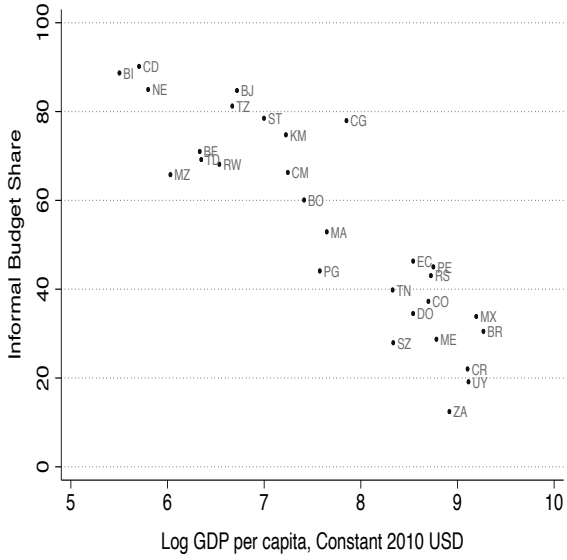
This figure shows each countries Informality Engel curves' on their GDP per capita, when controlling for geographical variables. Panel (a) controls for a rural/urban dummy variable and panel (b) controls for survey enumeration blocks.

Figure A7: Rural vs Urban Informal Consumption

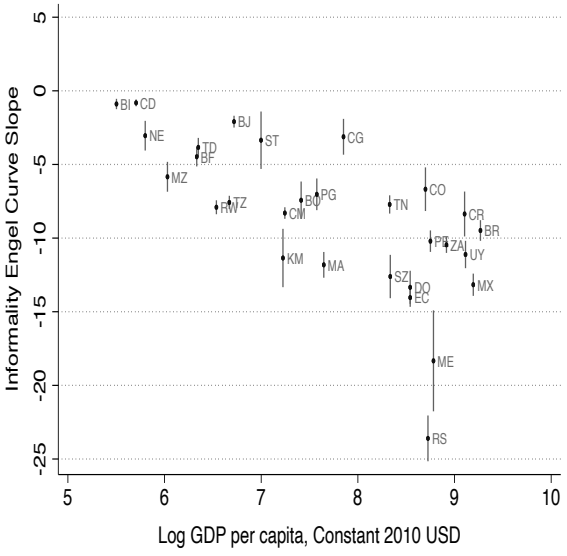
(a) Rural: Budget Share



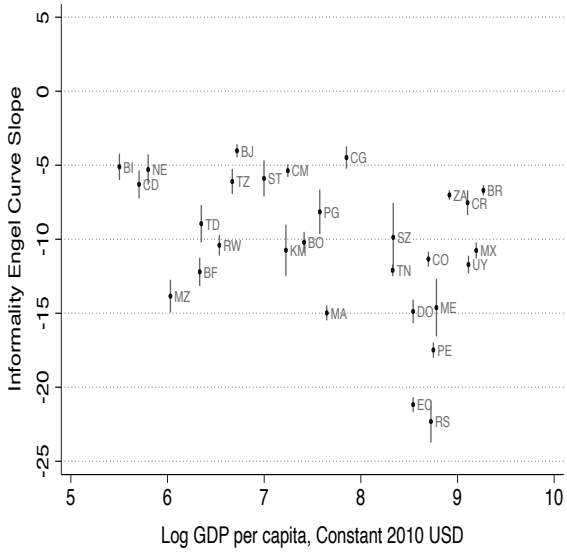
(b) Urban: Budget Share



(c) Rural: IEC slope



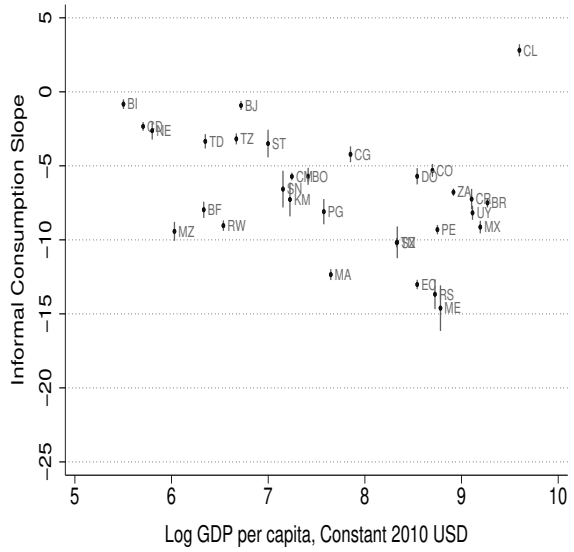
(d) Urban: IEC slope



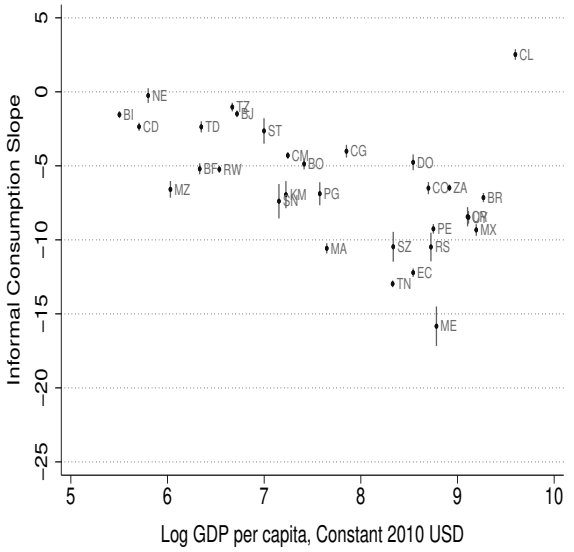
This figure plots informality levels and the slopes of the informality Engel curves for households located in rural regions (graphs a and c) and urban regions (graphs b and d). It only contains 29 countries instead of 31, since the expenditure surveys in Chile and Senegal concern urban population only.

Figure A8: Informality Engel Curve Slopes Controlling for Goods Composition

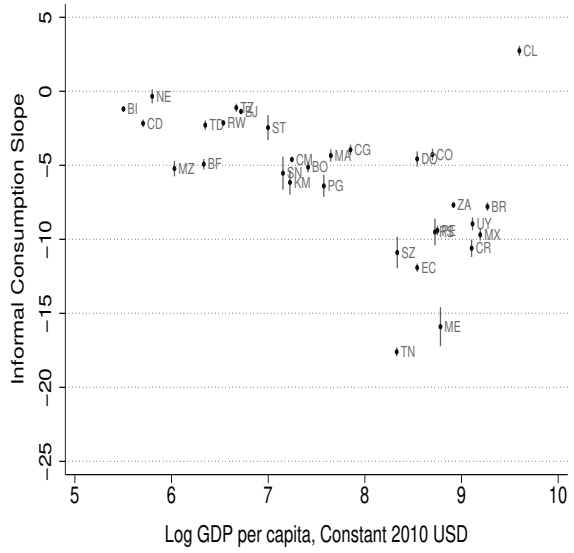
(a) Control: Food Products



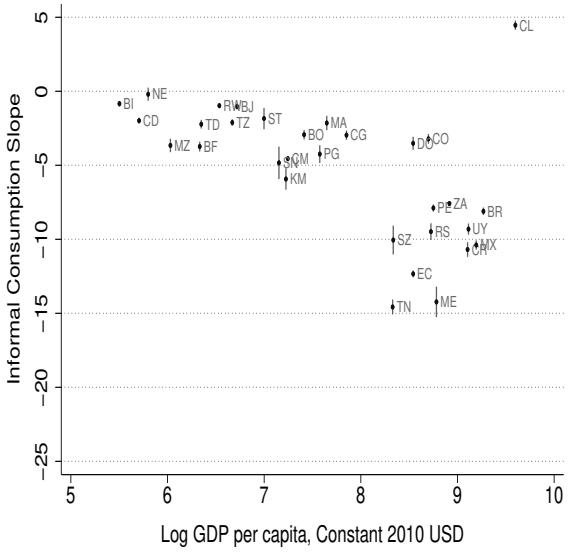
(b) Control: COICOP2



(c) Control: COICOP3

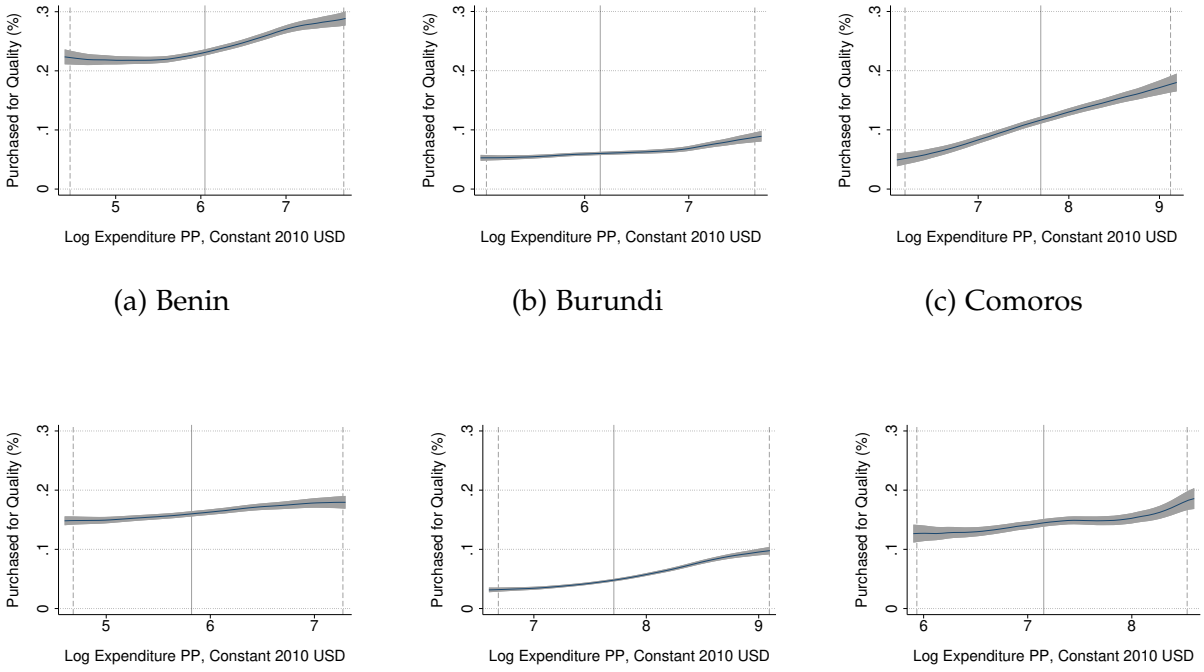


(d) Control: COICOP4



This figure shows the informality Engel curves' slopes across countries when controlling for increasingly narrow products (and household controls). Panel (a) only controls for food products, panel (b) controls for the 12 COICOP2 good categories, panel (c) controls for the 47 COICOP3 categories and panel (d) controls for the 117 COICOP4 categories.

Figure A9: Share of Purchases where Store is Chosen for its Quality by Income



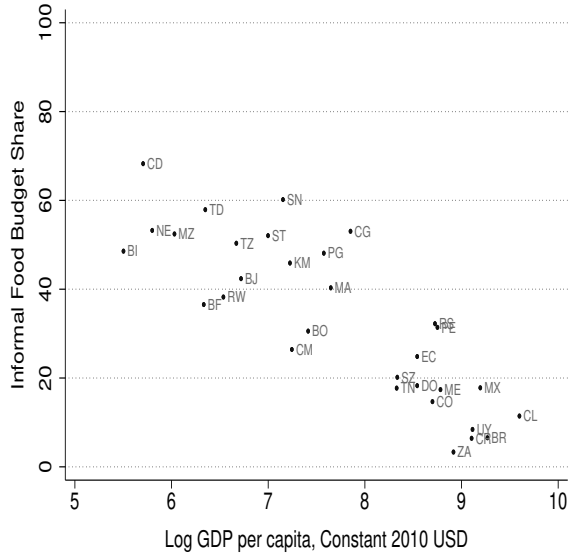
(a) Benin (b) Burundi (c) Comoros

(d) Dem. Rep. of Congo (e) Morocco (f) Rep. of Congo

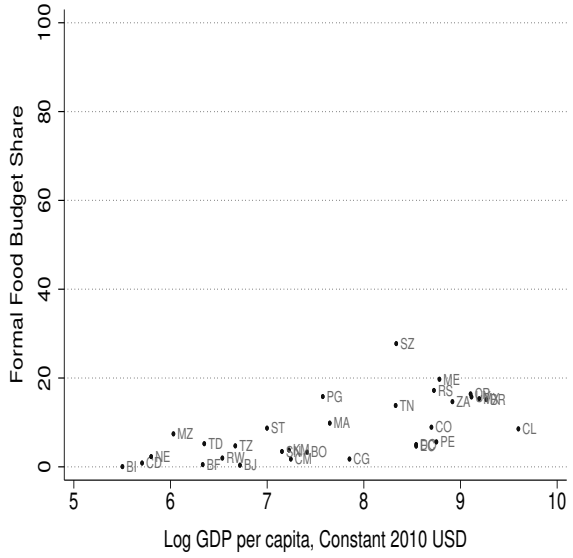
Local polynomial fit of the share of households buying any product for its quality on household's total expenditure per person (log). Each panel corresponds to one of the six countries, for which the expenditure survey asks respondents why they chose this place of purchase for each expenditure. The solid vertical line corresponds to the median household total expenditure, while the dotted lines correspond to the 5th and 95th percentile.

Figure A10: Budget Shares

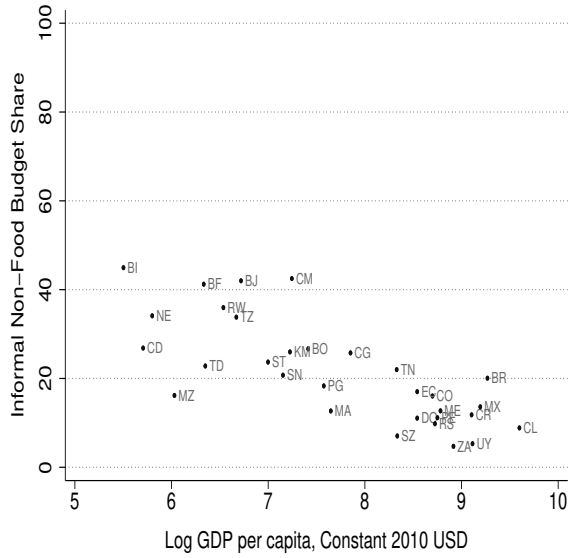
(a) Informal Food



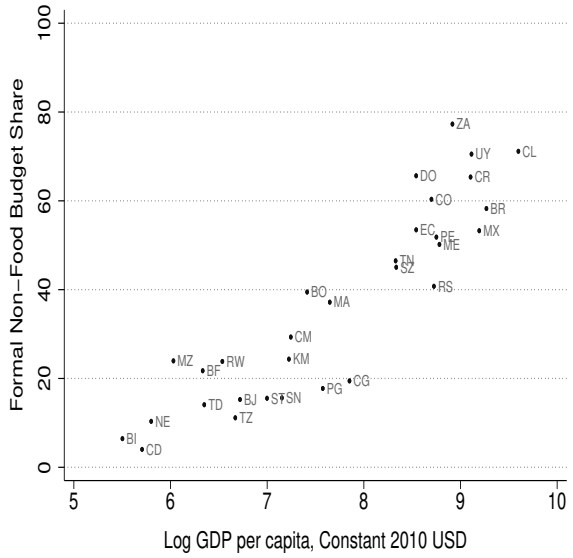
(b) Formal Food



(c) Informal Non-Food

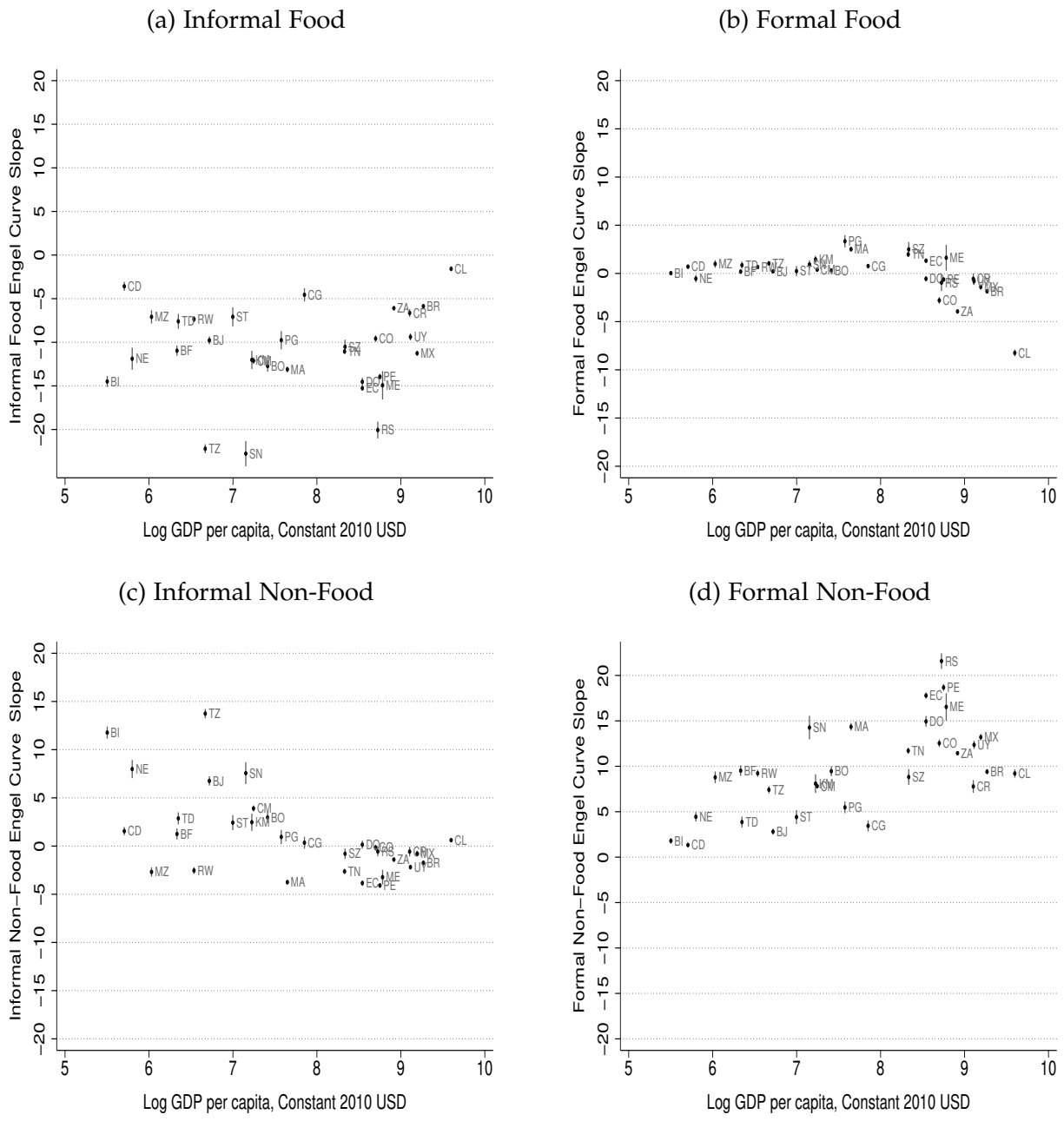


(d) Formal Non-Food



Panel (a) shows country informal food consumption as a share of total consumption, plotted against log per capita GDP. Panel (b), (c) and (d) are constructed similarly to Panel (a), respectively showing formal food consumption, informal non-food consumption and formal non-food consumption.

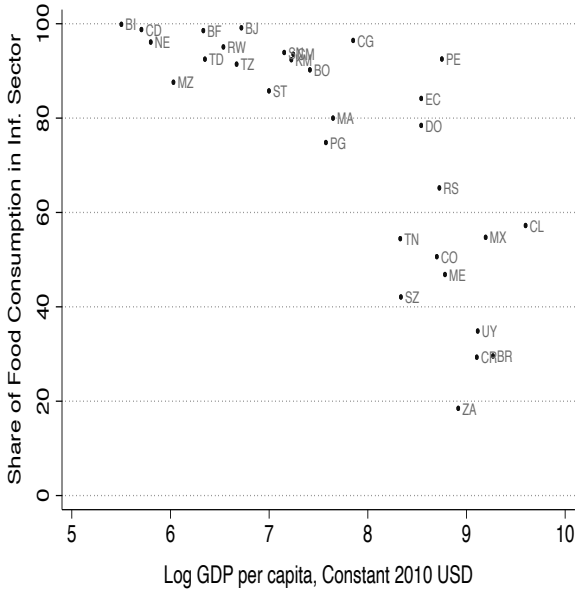
Figure A11: Engel Curve Slopes



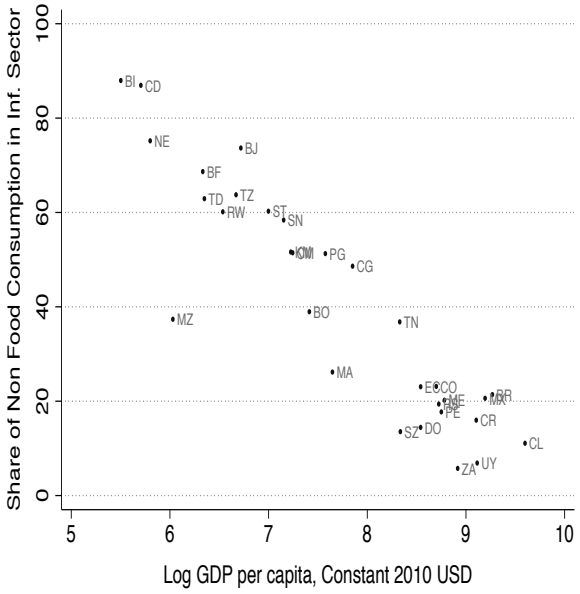
Panel (a) shows the country-specific slope of informal food consumption with respect to log household expenditure, plotted against log per capita GDP. The slope measures the drop in informal food consumption for a doubling of households' income. Panel (b), (c) and (d) are constructed similarly to Panel (a), respectively showing formal food consumption, informal non-food consumption and formal non-food consumption.

Figure A12: Share of Informal Consumption for Food and Non Food Goods

(a) Informal Food



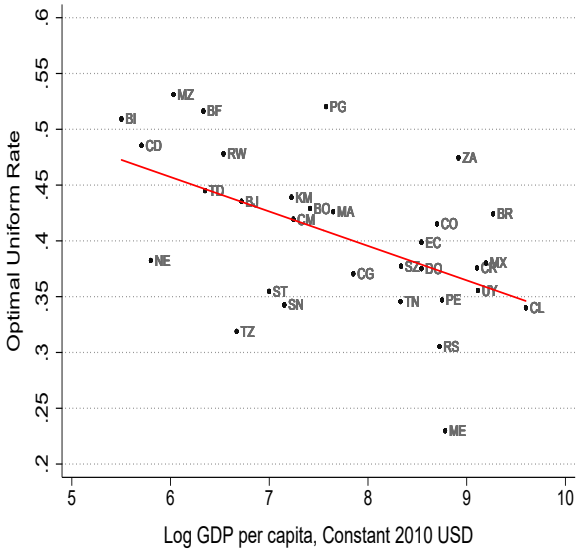
(b) Informal Non-Food



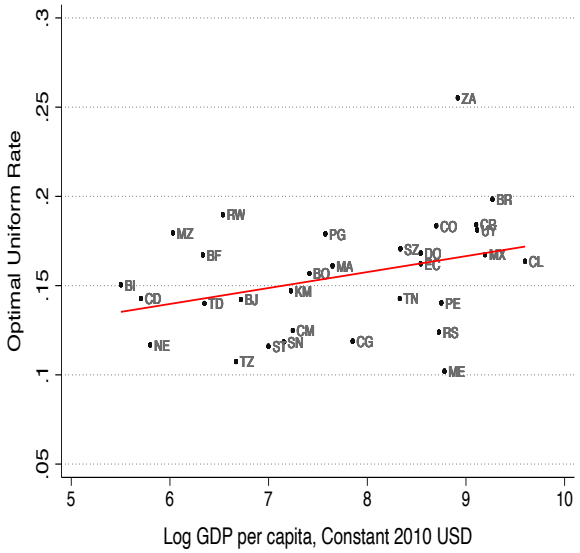
Panel (a) shows the share of food consumption which occurs in the informal sector. Panel (b) shows the share of non food consumption which occurs in the informal sector. For each good, this is constructed by taking total informal consumption of the good and dividing it by the total consumption of that good.

Figure A13: Optimal Uniform Tax on Formal Varieties

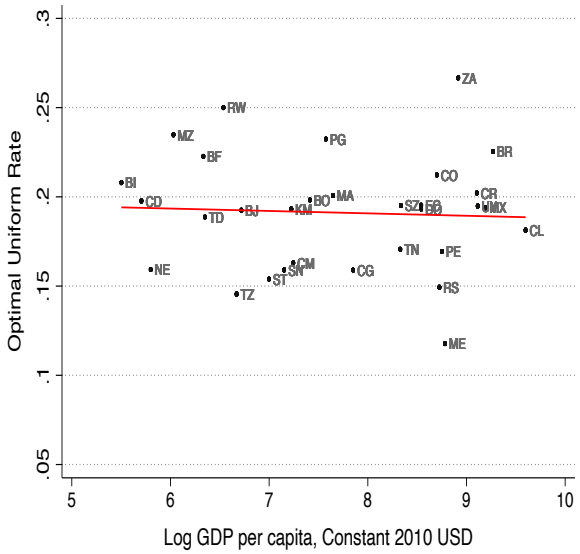
(a) No efficiency consideration



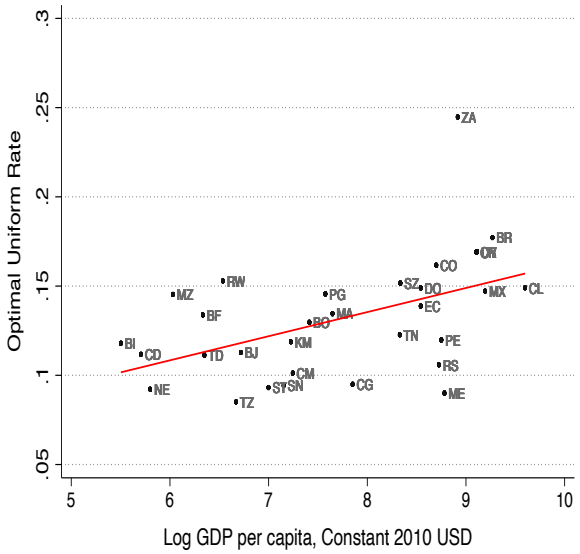
(b) Cross price elasticity = 1.5



(c) Cross price elasticity = 1



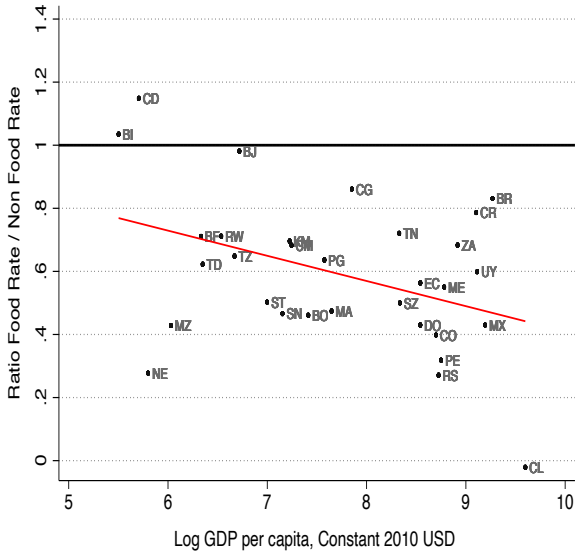
(d) Cross price elasticity = 2



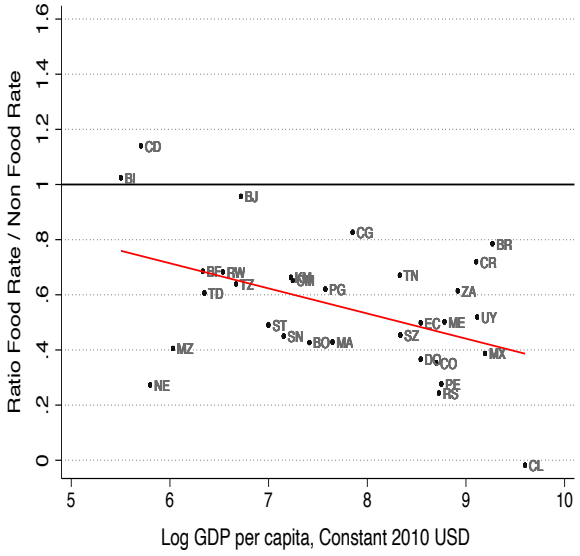
The panels plot the optimal uniform rate for each country calibrated using expression (6) as a function of each country's log GDP per capita. In panel (a) we hold efficiency considerations constant by assuming that uncompensated elasticities of demand are equal across all countries, in panels (b), (c) and (d) we allow price elasticities to vary across goods and countries by calibrating values using expression (4) above. We vary the value of the cross-variety compensated elasticity $\tilde{\epsilon}^C$ across the three panels: it is 1.5 (our baseline value) in panel (b), 1 in panel (c) and 2 in panel (d).

Figure A14: Optimal Rate Differentiation: Robustness

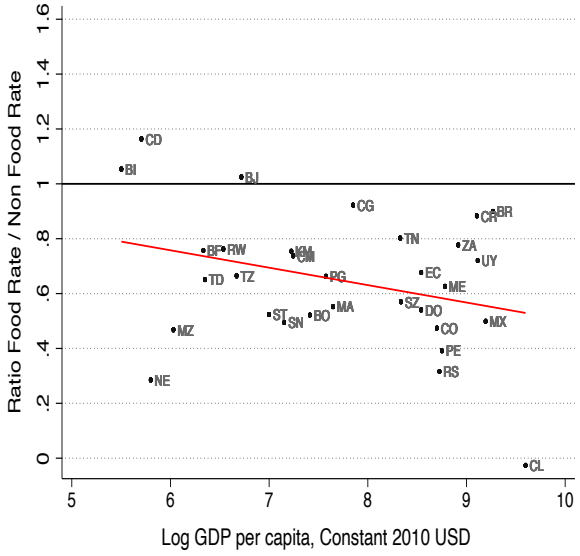
(a) Only Formal Varieties Taxed, Cross price elasticity = 1.5



(b) Only Formal Varieties Taxed, Cross price elasticity = 1

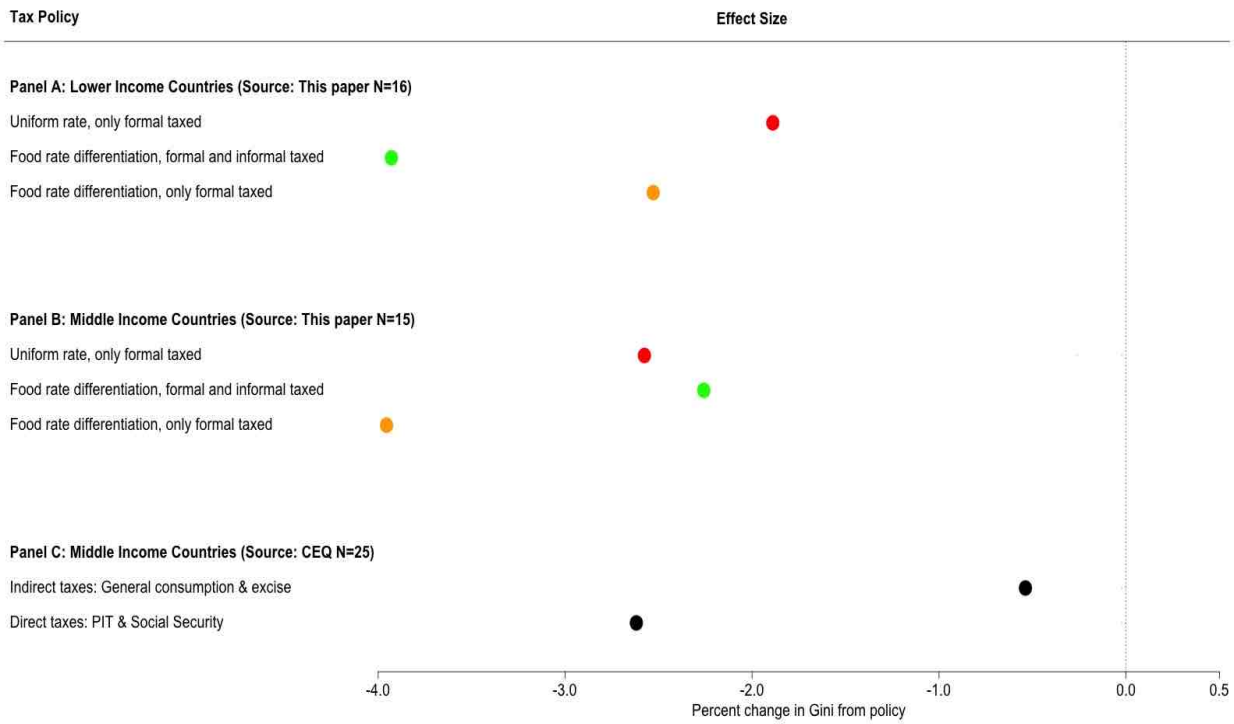


(c) Only Formal Varieties Taxed, Cross price elasticity = 2



The panels plot for each country the ratio of the calibrated optimal rate on food products relative to the optimal rate on non-food products, as a function of that country's log GDP per capita. Optimal rates are calibrated using expression (9). A value equal to 1 indicates that both optimal rates are set at the same level, a lower value indicates that it is optimal to subsidize food products relative to non-food products. In all panels we assume only formal varieties can be taxed and allow price elasticities to vary across goods and countries by calibrating values using expression (4) above. We vary the value of the cross-variety compensated elasticity $\tilde{\epsilon}^C$ across the three panels: it is 1.5 (our baseline value) in panel (a), 1 in panel (b) and 2 in panel (c).

Figure A15: Change in Gini by Income Groups



This figure shows the average percent change in the Gini coefficient per different group of countries for the three scenarios. Panel A shows the drop in Gini for lower income countries (i.e. the countries with a GDP per capita inferior to 2100 $ in constant US 2010.) of this paper core sample. Panel B shows the drop in Gini for middle income countries (i.e. the countries with a GDP per capita superior to 2100 $ in constant US 2010.) of this paper core sample. For Panel A and B the red dot represents the scenario where a uniform rate is implemented but only the formal sector is taxed. The green dot represents the scenario where only non-food items, but both sectors are taxed. The orange dot represents the scenario where only non-food items, and only the formal sector are taxed.Panel C show the drop in Gini using data from the Commitment to Equity Institute (CEQ). This sample gathers both lower and middle income countries.

Table A1: IEC Slopes by Country

Country	Main		Geography			Product Codes			All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Benin	3.31	3.61	3.18	4.54	0.92	1.49	1.36	1.03	1.26
	(0.15)	(0.16)	(0.16)	(0.22)	(0.16)	(0.11)	(0.10)	(0.10)	(0.15)
Bolivia	9.77	11.43	8.99	7.22	5.71	4.87	5.13	2.93	2.74
	(0.29)	(0.33)	(0.38)	(0.44)	(0.29)	(0.19)	(0.18)	(0.16)	(0.25)
Brazil	7.60	7.98	7.07	6.41	7.50	7.15	7.79	8.11	6.64
	(0.15)	(0.17)	(0.17)	(0.18)	(0.16)	(0.16)	(0.15)	(0.13)	(0.14)
Burkina Faso	9.71	10.56	7.58	6.89	7.97	5.20	4.92	3.73	2.39
	(0.30)	(0.32)	(0.30)	(0.32)	(0.28)	(0.19)	(0.18)	(0.17)	(0.19)
Burundi	1.89	2.35	1.44	0.80	0.83	1.54	1.20	0.84	0.33
	(0.16)	(0.17)	(0.17)	(0.18)	(0.17)	(0.12)	(0.10)	(0.10)	(0.12)
Cameroon	8.21	9.35	7.13	5.81	5.72	4.30	4.61	4.55	2.88
	(0.13)	(0.14)	(0.16)	(0.22)	(0.13)	(0.12)	(0.10)	(0.09)	(0.13)
Chad	5.72	6.21	4.54	3.10	3.35	2.37	2.29	2.23	0.90
	(0.29)	(0.30)	(0.30)	(0.37)	(0.25)	(0.19)	(0.16)	(0.15)	(0.22)
Chile	0.97	0.91	0.91	0.03	0.00	0.00	0.00	0.00	0.00
	(0.21)	(0.23)	(0.23)	(0.25)	(0.21)	(0.18)	(0.17)	(0.16)	(0.18)
Colombia	9.76	10.52	10.56	8.32	5.31	6.51	4.28	3.22	3.37
	(0.23)	(0.25)	(0.26)	(0.28)	(0.22)	(0.21)	(0.20)	(0.17)	(0.19)
Comoros	9.54	11.65	11.08	8.84	7.28	6.95	6.16	5.93	4.42
	(0.58)	(0.71)	(0.74)	(0.82)	(0.58)	(0.47)	(0.42)	(0.37)	(0.56)
CongoDRC	2.04	2.76	1.62	2.66	2.34	2.36	2.17	1.99	1.51
	(0.15)	(0.17)	(0.14)	(0.20)	(0.15)	(0.14)	(0.13)	(0.10)	(0.16)
Congo Rep	4.21	5.62	4.13	7.06	4.22	4.01	3.96	2.97	2.78
	(0.32)	(0.33)	(0.33)	(0.43)	(0.27)	(0.22)	(0.18)	(0.16)	(0.25)
Costa Rica	7.22	8.60	7.72	5.95	7.25	8.44	10.60	10.69	8.84
	(0.35)	(0.37)	(0.38)	(0.45)	(0.35)	(0.33)	(0.30)	(0.25)	(0.30)
Dominican Rep	14.39	14.89	14.48	11.78	5.70	4.76	4.57	3.52	2.36
	(0.31)	(0.35)	(0.35)	(0.42)	(0.28)	(0.27)	(0.26)	(0.23)	(0.25)
Ecuador	19.11	20.90	19.11	16.57	13.02	12.22	11.92	12.34	9.46
	(0.18)	(0.19)	(0.21)	(0.21)	(0.16)	(0.15)	(0.14)	(0.12)	(0.13)
Eswatini	11.64	12.38	11.55	12.56	10.17	10.47	10.89	10.05	9.88
	(0.51)	(0.62)	(0.67)	(0.65)	(0.55)	(0.51)	(0.54)	(0.50)	(0.51)
Mexico	12.01	13.57	11.51	9.83	9.14	9.33	9.70	10.39	7.09
	(0.20)	(0.23)	(0.24)	(0.25)	(0.22)	(0.20)	(0.20)	(0.16)	(0.19)
Montenegro	18.16	18.88	16.04	16.60	14.61	15.83	15.91	14.23	12.52
	(0.84)	(0.94)	(0.97)	(1.11)	(0.79)	(0.68)	(0.67)	(0.53)	(0.59)
Morocco	16.85	18.11	14.05	12.09	12.35	10.57	4.34	2.14	0.00
	(0.21)	(0.22)	(0.23)	(0.27)	(0.19)	(0.18)	(0.21)	(0.25)	(0.28)
Mozambique	9.67	11.28	8.83	9.30	9.43	6.59	5.23	3.66	2.74
	(0.37)	(0.39)	(0.40)	(0.44)	(0.33)	(0.29)	(0.27)	(0.23)	(0.29)
Niger	3.90	4.66	3.68	4.29	2.62	0.25	0.34	0.20	0.61
	(0.34)	(0.37)	(0.41)	(0.40)	(0.31)	(0.25)	(0.24)	(0.23)	(0.26)
Papua New Guinea	8.59	9.35	7.14	7.36	8.10	6.88	6.40	4.24	3.06
	(0.49)	(0.49)	(0.50)	(0.52)	(0.43)	(0.40)	(0.38)	(0.30)	(0.32)
Peru	18.04	18.92	16.04	12.59	9.31	9.26	9.41	7.89	4.15
	(0.20)	(0.22)	(0.23)	(0.27)	(0.16)	(0.16)	(0.16)	(0.12)	(0.16)
Rwanda	9.90	10.61	8.68	9.75	9.04	5.23	2.14	0.97	0.09
	(0.19)	(0.20)	(0.20)	(0.25)	(0.18)	(0.12)	(0.08)	(0.08)	(0.09)
Sao Tome	4.66	5.23	5.26	5.06	3.50	2.64	2.46	1.84	1.84
	(0.48)	(0.56)	(0.57)	(0.58)	(0.48)	(0.44)	(0.43)	(0.37)	(0.38)
Senegal	15.20	12.19	12.19	11.56	6.57	7.39	5.53	4.83	4.47
	(0.67)	(0.74)	(0.74)	(0.84)	(0.63)	(0.59)	(0.57)	(0.56)	(0.65)
Serbia	20.91	24.24	22.74	23.03	13.67	10.48	9.50	9.48	8.47
	(0.58)	(0.58)	(0.56)	(0.56)	(0.51)	(0.49)	(0.47)	(0.29)	(0.29)
South Africa	7.50	8.69	7.78	7.29	6.78	6.49	7.69	7.59	6.92
	(0.12)	(0.14)	(0.15)	(0.18)	(0.13)	(0.12)	(0.10)	(0.08)	(0.11)
Tanzania	8.47	7.52	7.26	4.88	3.18	1.03	1.11	2.11	0.80
	(0.20)	(0.20)	(0.20)	(0.27)	(0.19)	(0.15)	(0.14)	(0.11)	(0.15)
Tunisia	13.68	12.98	10.72	9.05	10.18	12.97	17.60	14.57	13.03
	(0.15)	(0.17)	(0.18)	(0.21)	(0.16)	(0.12)	(0.15)	(0.26)	(0.29)
Uruguay	11.57	11.73	11.65	10.87	8.18	8.48	8.96	9.31	8.36
	(0.25)	(0.27)	(0.28)	(0.32)	(0.24)	(0.22)	(0.22)	(0.19)	(0.21)
All Countries (Mean)	9.8	10.6	9.2	8.5	6.9	6.3	6.1	5.4	4.3
Household Characteristics		X	X	X	X	X	X	X	X
Urban/Rural			X						
Survey Blocks				X					X
Food Products					X			X	
COICOP 2-dig						X			
COICOP 3-dig							X		
COICOP 4-dig								X	X

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

Table A2: IEC Slopes: Probabilistic Formality Assignment

Specification:	**Main**		**Geography**		**Product Codes**				**All**
Avg. of 31 Countries	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Slope	7.8	8.3	7.1	6.4	5.0	4.5	4.4	3.8	3.0
Confidence Interval	[7.3,8.4]	[7.7,8.9]	[6.4,7.7]	[5.7,7.1]	[4.4,5.5]	[4.0,4.9]	[3.9,4.7]	[3.4,4.1]	[2.4,3.3]
# of p-values < 0.05	30	30	30	30	30	29	29	30	26
R^2 adjusted	0.15	0.17	0.22	0.39	0.42	0.54	0.54	0.54	0.58
Household Characteristics		X	X	X	X	X	X	X	X
Urban/Rural			X						
Survey Blocks				X					X
Food Products					X				
COICOP 2-dig						X			
COICOP 3-dig							X		
COICOP 4-dig								X	X

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

Table A3: IEC Slopes: Robustness Formality Assignment

Specification:	**Main**		**Geography**		**Product Codes**				**All**
Avg. of 31 Countries	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Slope	7.0	7.4	6.3	5.8	4.3	3.8	3.8	3.3	2.7
Confidence Interval	[6.4,7.6]	[6.7,8.1]	[5.6,7.0]	[4.8,6.4]	[3.5,4.7]	[3.1,4.1]	[3.1,4.1]	[2.7,3.5]	[2.0,3.0]
# of p-values < 0.05	30	30	29	28	28	28	27	25	25
R^2 adjusted	0.12	0.13	0.17	0.34	0.38	0.52	0.53	0.54	0.57
Household Characteristics		X	X	X	X	X	X	X	X
Urban/Rural			X						
Survey Blocks				X					X
Food Products					X				
COICOP 2-dig						X			
COICOP 3-dig							X		
COICOP 4-dig								X	X

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

Table A4: Main Reason for Choosing Place of Purchase

	Outcome: Share of purchases (in %)								
	Benin			Burundi			Comoros		
Reason	All Stores	Informal	Formal	All Stores	Informal	Formal	All Stores	Informal	Formal
Access	39.0	39.3	29.9	49.8	49.9	41.5	36.2	38.6	16.4
Price	26.4	26.8	11.6	27.6	27.8	14.8	31.1	31.7	26.1
Quality	24.3	23.5	51.4	6.4	5.7	41.0	12.4	9.0	39.8
Store Attributes	7.4	7.6	3.3	3.7	3.8	0.8	13.4	14.3	6.0
Other	2.9	2.9	3.9	12.6	12.8	1.9	7.0	6.4	11.7

	Dem. Rep of Congo			Morocco			Rep. of Congo		
Reason	All Stores	Informal	Formal	All Stores	Informal	Formal	All Stores	Informal	Formal
Access	28.7	28.9	16.1	58.5	58.7	57.3	36.8	37.5	26.8
Price	34.3	34.4	27.2	20.1	22.5	6.4	32.4	33.3	20.0
Quality	16.6	16.3	46.5	6.3	3.9	19.7	14.3	12.2	45.0
Store Attributes	7.8	7.8	7.6	1.7	0.6	7.7	7.2	7.4	4.3
Other	12.6	12.7	2.7	13.5	14.3	8.9	9.3	9.7	3.8

The table reports the frequencies across all purchases by reason of choosing a place of purchase, and shows the average for the six countries in the core sample which ask this question. These countries are Benin, Burundi, Comoros, Congo Rep., Morocco and RD Congo. In all surveys seven reasons are listed which we classify into five categories as follows: access is defined as "The retailer is closer or more convenient" and "The good or service cannot be found elsewhere", price as "The good or services are cheaper", quality as "The goods or services are of better quality", store attributes as "The retailer offers credit" and "The retailer is welcoming or is a friend" and other as "Others reasons". Note that Morocco has a few additional small categories, which pertain to attributes of retailer. The table lists the frequency for all purchases of goods and excludes services, which are less comparable along these dimensions, although their inclusion does not impact the results.

Table A5: Unit Values Across Places of Purchase

	Outcome: % dif. in formal sector unit values				# Purchases	# FE
Country	(1)	(2)	(3)	(4)	(5)	(6)
Benin	5.25	1.10	3.38	-0.39	262,280	5,065
	(7.10)	(5.66)	(7.53)	(6.19)		
Bolivia	4.08	3.53	4.69	3.86	120,971	1,549
	(1.40)	(1.12)	(1.40)	(1.15)		
Brazil	-0.11	-0.20	0.14	0.01	704,639	9,437
	(0.37)	(0.35)	(0.38)	(0.35)		
Burundi	2.53	4.39	4.81	5.23	250,139	2,454
	(4.65)	(4.73)	(4.39)	(4.23)		
Chad	-4.36	-3.21	-4.36	-3.21	380,462	1,968
	(1.80)	(1.77)	(1.80)	(1.77)		
Colombia	-0.33	-0.04	-0.30	-0.06	778,203	7,861
	(0.55)	(0.30)	(0.55)	(0.30)		
Comoros	22.56	14.93	21.81	14.49	113,228	1,142
	(5.01)	(3.64)	(4.98)	(3.64)		
CongoDRC	4.62	0.87	9.77	5.89	865,754	5,556
	(16.79)	(12.88)	(17.47)	(14.15)		
Congo Rep	27.84	23.70	27.12	23.01	208,557	1,182
	(5.88)	(4.67)	(6.03)	(4.78)		
Costa Rica	3.04	2.37	1.93	1.58	122,467	1,593
	(2.40)	(2.11)	(2.17)	(1.93)		
Dominican Rep	18.86	13.64	18.94	13.73	340,303	4,416
	(1.69)	(1.01)	(1.68)	(1.00)		
Ecuador	2.29	1.86	2.23	1.82	1,030,387	12,104
	(0.63)	(0.63)	(0.63)	(0.62)		
Eswatini	3.09	2.38	1.31	1.06	89,209	852
	(2.10)	(1.79)	(1.89)	(1.46)		
Mexico	1.10	1.00	1.10	1.00	446,417	6,195
	(1.16)	(1.02)	(1.16)	(1.02)		
Montenegro	10.36	9.57	7.13	6.45	138,446	867
	(3.70)	(3.25)	(3.08)	(2.85)		
Morocco	7.10	5.43	6.88	5.22	743,979	3,598
	(0.87)	(0.70)	(0.92)	(0.75)		
Peru	14.70	13.29	14.69	13.29	1,300,408	10,721
	(2.74)	(2.46)	(2.74)	(2.46)		
Sao Tome	6.81	4.87	6.69	4.86	215,527	2,946
	(1.39)	(1.37)	(1.39)	(1.34)		
Serbia	2.39	2.03	2.86	2.49	503,344	9,332
	(0.49)	(0.46)	(0.51)	(0.48)		
Tanzania	2.11	1.59	2.80	2.21	1,169,193	13,771
	(0.73)	(0.68)	(0.59)	(0.55)		
Avg. of 20 Countries	6.70	5.16	6.68	5.13		
Confidence Interval	[0.7,12.7]	[0.2,10.1]	[0.7,12.7]	[0.1,10.1]		
# of p-values < 0.05	12	12	11	11		
Winsorization [5,95]		X		X		
Self Consumption			X	X		

This table shows the percentage difference in unit values in the formal sector compared to the informal sector. The sample is restricted to food purchases, for which units and unit values are detailed, in the 20 core sample countries with such data. Formally it runs the following specification: $ln(unit\ value)_{ipmu} = \beta\ Formal_{ipmu} + \mu_{pmu} + \epsilon_{ipmu}$, where $ln(unit\ value)_{ipmu}$ is the unit value reported by household i, for product p, in location m, in units u, and $Formal_{ipmu}$ equals one if the product is purchased in a formal store. We add fixed effects at the product * location * unit of measurement. Standard errors are clustered at the location level.

Table A6: Ratio Top over Bottom Quintile of Effective Tax Rates

Tax policy	Baseline Assignment	Baseline + VAT on Input	Baseline + Distributional Savings	Probabilistic Assignment	Robust Assignment
Uniform rate, only formal taxed	2.06	1.62	1.89	1.76	1.95
Food exempt, formal and informal taxed	1.67	1.67	1.54	1.63	1.67
Food exempt, only formal taxed	2.36	2.18	2.17	2.11	2.19

This table shows the progressivity of consumption tax policies, measured as the effective tax rate of the richest household quintile over that of the poorest quintile. The numbers are averages for the 31 countries in the core sample. The rows correspond to the three tax policies considered: (1) a uniform rate on all goods in a context where only formal goods can be taxed, (2) a tax exemption on food in a context where both formal and informal goods can be taxed, and (3) an exemption on food, in a context where only formal goods can be taxed. The columns correspond to the different assumptions on our data. Column (1) corresponds to the central informality assignment of retailers. Column (2) models a 10% pass-through of taxes onto prices in the informal sector, following the share of formal inputs in informal firms in Mexico's census. Column (3) allows for savings rate, which increases linearly from 0 for the bottom decile to 15% for the top decile, following evidence form consumer finance surveys. Columns (4) and (5) use the alternative assignment of store types to formality, by using either the probability of formality by store type from Mexico's census, or by assigning specialized stores to the informal sector instead of the formal sector.

Table A7: Percent Change in Gini from Optimal Tax Policy

	(1) Baseline Assignment	(2) Baseline + VAT on Input	(3) Baseline + Distributional Savings	(4) Probabilistic Assignment	(5) Robust Assignment	(6) Cross-variety Elasticity	(7)
Tax policy						$\tilde{\epsilon}^C = 1$	$\tilde{\epsilon}^C = 2$
Uniform rate, only formal taxed	-2.30	-2.02	-1.24	-1.71	-1.54	-2.82	-1.95
Food rate differentiation, formal & informal taxed	-3.16	-3.16	-0.96	-3.16	-3.16	-3.16	-3.16
Food rate differentiation, only formal taxed	-3.26	-3.11	-1.88	-2.42	-2.03	-3.84	-2.85
Full rate differentiation, formal & informal taxed (12 goods)	-4.81	-4.81	-2.40	-4.60	-4.81	-4.81	-4.81
Full rate differentiation, only formal taxed (12 goods)	-4.19	-4.01	-3.11	-3.22	-2.83	-4.73	-3.82

This table shows the redistributive impact of different consumption tax policies under different hypothesis, as presented in section 6. Our metric for redistribution is the percent change in Gini from the pre-tax income distribution to the net-of-tax distribution. We take the average across the 31 countries in the core sample. The rows correspond to the tax policy scenarios considered. (1) a uniform rate on all goods in a context where only formal goods can be taxed, (2) optimal tax rates on food and non food goods in a context where both formal and informal goods can be taxed, (3) optimal tax rates on food and non food goods, in a context where only formal goods can be taxed, (4) optimal tax rates for each of the 12 COICOP-2 digit level goods in a context where both formal and informal goods can be taxed, and (5) optimal tax rates for each of the 12 COICOP-2 digit level goods in a context where only formal goods can be taxed. The columns correspond to the different assumptions on our data. The baseline in column (1) corresponds to the central assignment of retailers to informality status, and to a value of the cross-variety elasticity of 1.5. Column (2) adds to the baseline scenario a 10% pass-through of taxes onto prices in the informal sector, following the share of formal inputs in informal firms in Mexico's census. Column (3) adds the baseline scenario a distributional savings rate which ranges from 0 for the bottom decile to 15% for the top decile, following evidence form consumer finance surveys. Columns (4) and (5) use the alternative assignment of store types to formality. Finally columns (6) and (7) keep the central store assignment to informality but vary the value of the elasticity of substitution between the formal and informal variety of goods.

B Data Appendix

All statistical codes to replicate the paper are available at https://github.com/pierrebachas/Informality_Taxes_Redistribution. This includes the cleaning files for the micro data of each country's survey, as well as all files generating the tables and figures of the paper.

B.1 Core Sample

Inclusion Criteria

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

1. The household expenditure survey is nationally representative and dates from the 21st century.
2. The expenditure module(s) in the survey is structured as an open consumption diary, rather than a pre-fill diaries for a limited set of products.
3. The expenditure survey includes a variable for the place of purchase (data on where each item was purchased). The place of purchases are detailed enough for us to apply our taxonomy of store types, as further outlined in section B.2.
4. The place of purchase variable rarely contains missing values, particularly for food, clothing and household goods product categories (see Figure A3).

Data Sources and Coverage

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

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

Table B1: Regional Survey Representation

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

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

Table B2: Discarded Household Expenditure Surveys

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

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

Table B3: Household Expenditure Surveys

Country name	Survey	Year	Source	# HH	# items/HH	Exp/HH	Urban	HH Size	# PoP	# Modules	Product Code	Comments
						Cst. 2010 USD						
Benin	EMICOV	2015	World Bank	19872	31.9	261	48.2%	4.3	12	22	COICOP	
Bolivia	ECH	2004	Stat. Office	9149	49.4	585	60.7%	4.2	24	3	COICOP	
Brazil	POF	2009	Stat. Office	56049	48	3892	84.4%	3.3	33	8	Country-specific	
Burkina Faso	EICVM	2009	Stat. Office	8404	161.6	563	29.3%	6.7	45	1	COICOP	
Burundi	ECVM	2014	World Bank	6681	90.2	242	9.0%	4.8	13	23	COICOP	
Cameroon	ECAM	2014	World Bank	10303	95.8	1889	44.5%	4.6	17	1	COICOP	
Chad	ECOSIT	2003	World Bank	6747	92	356	10.9%	5.9	17	18	Country-specific	
Chile	EPF	2017	Stat. Office	15237	129.2	6872	100.0%	3.3	22	1	COICOP	No self-production, Only urban
Colombia	ENIG	2007	Stat. Office	42733	79.6	1850	82.4%	3.8	24	5	COICOP	
Comoros	EDMC	2013	Stat. Office	3139	83.5	1809	49.1%	5	12	19	COICOP	
Congo DRC	E123	2005	World Bank	12098	106.9	198	16.0%	5.3	13	1	COICOP	
Congo Rep	ECOM	2005	World Bank	5002	84.8	641	63.8%	5.1	17	1	COICOP	
Costa Rica	ENIGH	2014	Stat. Office	5705	67.5	5256	73.2%	3.4	41	1	COICOP	
Dominican Rep	ENIGH	2007	Stat. Office	8363	89.1	2396	67.6%	3.7	88	3	COICOP	
Ecuador	ENIGHUR	2012	World Bank	39617	88.6	1923	68.0%	3.9	75	7	COICOP	
Eswatini	HIES	2010	World Bank	3167	43.9	1283	37.4%	4.5	13	2	COICOP	
Mexico	ENIGH	2014	Stat. Office	19459	57.4	2272	64.5%	3.8	19	1	COICOP	
Montenegro	HBS	2009	World Bank	1223	148.9	3731	62.7%	3	7	3	COICOP	Cant separate categories 3 & 4
Morocco	ENCDM	2001	World Bank	14243	87.5	1679	61.6%	5.9	47	17	COICOP	
Mozambique	IOF	2009	World Bank	10809	48.7	363	28.9%	4.7	6	6	COICOP	
Niger	ENCBM	2007	World Bank	4000	221.2	325	17.2%	6.4	15	6	COICOP	
Papua NG	HIES	2010	World Bank	3811	111.2	1002	11.3%	5.1	6	1	COICOP	
Peru	ENAHO	2017	Stat. Office	43545	78.5	2609	76.8%	3.9	41	8	Country-specific	
Rwanda	EICV	2014	World Bank	14419	53.6	417	17.1%	4.6	11	8	COICOP	Pre-filled items
SaoTome	IOF	2010	World Bank	3145	105.9	705	68.1%	3.8	21	3	COICOP	
Senegal	EDMC	2008	World Bank	1443	517.8	640	100.0%	7.7	41	1	COICOP	Only urban
Serbia	HBS	2015	World Bank	6531	106	1888	61.9%	2.8	9	2	COICOP	
South Africa	IES	2011	U. of Cape Town	25325	44.2	3557	67.3%	3.8	6	1	COICOP	Cant separate categories 3 & 4
Tanzania	HBS	2012	World Bank	10186	317.8	478	21.9%	5	13	2	COICOP	Cant separate categories 3 & 4
Tunisia	ENBCNV	2010	Stat. Office	11281	139.1	1732	67.6%	4.3	9	1	COICOP	Cant separate categories 3 & 4
Uruguay	ENIGH	2005	Stat. Office	7042	77.5	2855	84.9%	3	39	1	COICOP	

The column '# PoP' refers to the number of different places of purchase in the country classification.

Consumption Module Structure

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

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

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

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

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

B.2 Categories of Places of Purchase

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

- **Goods**

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

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

(3) Market consumption, corner and convenience shops

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

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

- **Services**

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

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

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

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

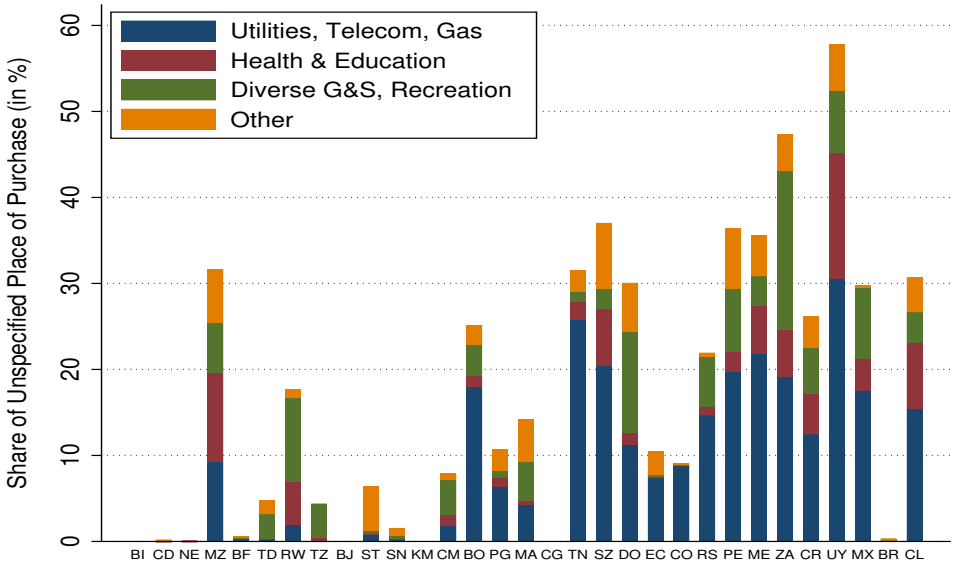
- **Unspecified**

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

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

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

Figure B1: Average Share of Unspecified Category by COICOP



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

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

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

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

B.3 Global Consumption Database

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

We obtained access to the Global Consumption Database microdata, in order to compare food expenditure in our core sample of 31 countries to the 79 low and middle income countries available in the GCD. From our sample of 31 countries, 21 countries overlap with the GCD and usually have the exact same survey as an original source. With this enhanced dataset, which represents 51% of the world population,[56] we measure food consumption as a share of total consumption and the slope of the food Engel curve. First we note that for the 21 overlapping countries we find Engel curve slopes within 5% of the GDC estimate. Second we compare food expenditure patterns in our core sample 31 countries to the 58 countries which only appears in the GCD. We find remarkably similar food expenditure shares and food Engel curve slopes, as a function of development, which supports that our core sample informality measures could be informative to the entire population located in developing countries, with the caveat that informality patterns might still differ geographically.

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

Table B4: Country-Specific Places of Purchase

BENIN

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

BOLIVIA

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

BRAZIL

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

BURKINA FASO

Formal	0.6	magasin de gros a petits prix	5 large stores
	0.7	grands magasin	5 large stores
	0.5	quincallerie (petite taille)	4 specialized shops
	1.3	atelier, service reparation	4 specialized shops
	1.7	station service (lubrifiants)	4 specialized shops
	1.0	pharmacie	4 specialized shops
	1.0	clinique, laboratoire medical public	6 institutions
	0.7	ecole, lycee, universite publics	6 institutions
	1.2	ecole, lycees, universite privas	6 institutions
	0.6	cabine telephone privee	6 institutions
	2.0	telephone, eau, electricite	6 institutions
	1.8	bar, cafe, restaurant, hotel	8 entertainment
Informal	14.3	boutique de quartier	3 corner shops
	0.9	kiosque ou echoppe quartier	2 no store front
	40.5	marche	2 no store front
	1.0	marchant ambulants	2 no store front
	10.1	bien ou service autoproduit	1 non-market
	12.4	menage	1 non-market
	0.5	cadeau recu en nature ou en espace	1 non-market
	2.1	autres service prives	7 service from individual
	1.4	service de transport prive	7 service from individual
Unspec.	0.3	other	99 n.a./other

BURUNDI

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

CAMEROON

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

CHAD

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

CHILE

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

COLOMBIA

Formal	4.7	Supermercados de barrio	5 large stores
	10.1	Almacenes o supermercados de cadena y tien	5 large stores
	2.1	Plazas de mercado y galerías	5 large stores
	0.6	Hipermercados	5 large stores
	11.1	Establecimientos especializados en la vent	4 specialized shops
	1.7	Farmacias y droguerías	10 pharmacies
	1.0	Televentas y ventas por catálogo	6 institutions
	5.1	Restaurantes	8 entertainment
Informal	0.9	Graneros	3 corner shops
	13.5	Tiendas de barrio	3 corner shops
	1.7	Vendedores ambulantes o ventas callejeras	2 no store front
	1.1	Persona particular	1 non-market
	0.9	Transfers, from household	1 non-market
	10.3	Self production	1 non-market
	1.0	Cafeterías y establecimientos de comidas	9 informal entertainment
Unspec.	32.9	Other	99 n.a./other

COMOROS

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

CONGO DRC

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

COSTA RICA

Assigned	%	Original name	Classification
Formal	1.2	tienda por departamentos	5 large stores
	17.1	supermercado	5 large stores
	11.3	local especializado	4 specialized shops
	4.3	gasolinera y estación de servicio	4 specialized shops
	1.1	carnicería / pescadería	4 specialized shops
	1.0	salones de estética o belleza	4 specialized shops
	3.4	almacén de electrodomésticos y de tecnol	4 specialized shops
	3.4	tienda de ropa / zapatería / perfumería	4 specialized shops
	1.9	laboratorio / clínica / centro médico	6 institutions
	1.1	en el exterior	6 institutions
	3.9	restaurante / soda / cafetería / helader	8 entertainment
	1.7	comedor en lugar de trabajo	8 entertainment
Informal	6.2	pulpería o minisuper	3 corner shops
	2.4	vendedor ambulante o a domicilio	2 no store front
	0.5	feria del agricultor	2 no store front
	0.8	local de artículos usados	2 no store front
	9.1	recibido o comprado a otros hogares	1 non-market
	0.8	retiro del negocio	1 non-market
Unspec.	25.1	other	99 n.a./other

DOMINICAN REPUBLIC

Formal	3.8	tienda por departamentos	5 large stores
	3.6	supermercados	5 large stores
	0.7	tienda de respuestos de vehiculos	4 specialized shops
	1.1	taller de mecanica en general, desabulladu	4 specialized shops
	0.6	puesto de rifa de aguante y loteria electr	4 specialized shops
	1.2	tienda de ropa	4 specialized shops
	1.0	ferreterias	4 specialized shops
	0.8	carniceria	4 specialized shops
	1.2	tienda de electrodomesticos	4 specialized shops
	0.5	peluqueria	4 specialized shops
	1.4	salon de belleza	4 specialized shops
	2.3	farmacias	4 specialized shops
	1.2	compañia de teléfonos	6 institutions
	1.9	envasadora de gas	6 institutions
	1.8	comedor popular	6 institutions
	2.4	clinica	6 institutions
	1.3	hospitales	6 institutions
	1.7	corporación de electricidad	6 institutions
	3.5	estación de gasolina	6 institutions
	1.5	colegio	6 institutions
	0.5	restaurante	8 entertainment
Informal	20.3	colmado	3 corner shops
	0.7	almacen de provisiones	3 corner shops
	0.6	picapollo	2 no store front
	0.5	puesto de pollo	2 no store front
	3.2	vendedora ambulante	2 no store front
	1.2	mercados	2 no store front
	0.9	puestos de venta	2 no store front
	1.9	autoproduction	1 non-market
	1.6	cafeteria	9 informal entertainment
Unspec.	29.5	other	99 n.a./other

ECUADOR

Formal	4.0	supermercados de cadena	5 large stores
	1.2	hipermercados	5 large stores
	0.5	repuestos de automotores	4 specialized shops
	0.5	tercena/carnicera	4 specialized shops
	1.3	librerias y papelerias	4 specialized shops
	1.4	otros sitios de compra especializados	4 specialized shops
	1.0	salas de belleza	4 specialized shops
	0.5	computadoras y accesorios	4 specialized shops
	1.1	gasolineras	4 specialized shops
	2.1	electrodomesticos y accesorios	4 specialized shops
	4.1	ropa de todo tipo	4 specialized shops
	1.9	calzado de todo tipo	4 specialized shops
	1.2	panaderas	4 specialized shops
	1.2	mecanicas automotrices	4 specialized shops
	0.8	muebles y enceres	4 specialized shops
	5.1	boticas y farmacias	4 specialized shops
	2.2	establecimientos privados de salud	6 institutions
	4.7	establecimientos educativos	6 institutions
	0.5	instituciones publicas	6 institutions
	2.2	transporte de pasajeros	6 institutions
	1.2	venta por cat·logo o television	6 institutions
	0.6	aseguradoras	6 institutions
	1.4	servicios profesionales (abogados, arqu)	6 institutions
	2.3	restaurantes, salones	8 entertainment
Informal	12.8	tiendas de barrio	3 corner shops
	1.5	bodegas, distribuidores	3 corner shops
	10.4	mercados	2 no store front
	1.1	ferias libres	2 no store front
	2.0	vendedores ambulantes	2 no store front
	11.1	productos autoconsumo, autosuministro	1 non-market
	0.9	personas particulares	7 service from individual
Unspec.	12.9	other	99 n.a./other

ESWATINI

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

MEXICO

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

MONTENEGRO

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

MOROCCO

Formal	0.7	Gas stations (benzine)	4 specialized shops
	0.5	Small Bookshop, kiosk	4 specialized shops
	0.8	Modern clothing shop	4 specialized shops
	3.1	Butcher or retail chicken seller	4 specialized shops
	1.2	Craftsman's shop (hairdresser, tailor)	4 specialized shops
	1.8	Pharmacy	10 pharmacies
	35.5	Public and semi-public agencies	6 institutions
	1.5	Regular transportation means (bus, train,	6 institutions
	1.7	Public administration	6 institutions
	0.7	Public baths, shower, swimming pool	6 institutions
	4.1	Private education institution	6 institutions
	1.3	Medical care in private institution	6 institutions
Informal	1.9	Grocer's	3 corner shops
	9.7	Neighborhood or village grocer	3 corner shops
	2.6	Neighborhood market	2 no store front
	1.3	Itinerant merchant selling on sidewalks	2 no store front
	10.7	Weekly market	2 no store front
	0.5	City market or central market	2 no store front
	3.6	Self production/consumption	1 non-market
	0.7	Cafe, non-standing restaurant	9 informal entertainment
Unspec.	11.2	Other	99 n.a./other

MOZAMBIQUE

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

NIGER

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

Assigned	%	Original name	Classification
		PAPUA NEW GUINEA	
Formal	34.5	Supermarket	5 large stores
Informal	9.4	Small shop, canteen, tuck shop	3 corner shops
	10.5	Local market	2 no store front
	3.8	Street vendor	2 no store front
	14.2	Home production	1 non-market
	10.2	Gift	1 non-market
Unspec.	17.6	Other	99 n.a./other
		PERU	
Formal	3.4	Supermercado	5 large stores
	1.1	Bodega (por mayor)	5 large stores
	0.8	Panadería	4 specialized shops
	0.6	Peluquería	4 specialized shops
	0.9	Librería	4 specialized shops
	0.5	Tienda especializada al por mayor	4 specialized shops
	5.8	Tienda especializada al por menor	4 specialized shops
	3.6	Farmacia	10 pharmacies
	3.4	Empresas de Transporte formales	6 institutions
	1.6	Centro de estudios	6 institutions
	1.3	Grifos de empresas	6 institutions
	0.5	Talleres formales	6 institutions
	0.6	Clínica particular	6 institutions
	1.0	Restaurantes y/ó bares	8 entertainment
Informal	14.6	Bodega (por menor)	3 corner shops
	24.0	Mercado (por menor)	2 no store front
	2.7	Feria	2 no store front
	3.3	Mercado (por mayor)	2 no store front
	5.0	Ambulante	2 no store front
	0.5	Empresas de Transporte informales	7 service from individual
Unspec.	22.3	Other	99 n.a./other
		CONGO REPUBLIC	
Formal	1.0	Grands magasins	5 large stores
	7.0	Autres commerces modernes	4 specialized shops
	3.9	Secteur transports	6 institutions
	2.5	Cliniques, laboratoires médicaux et écol	6 institutions
	5.8	Prestataires de services publics	6 institutions
	3.9	Hotels, restaurants, bars, cafes	8 entertainment
Informal	3.4	Epiceries modernes	3 corner shops
	8.4	Echoppes sur marches et sur bord de route	2 no store front
	6.2	Marchands ambulants	2 no store front
	42.8	Marches	2 no store front
	3.9	Ménages	1 non-market
	4.5	Produit autoconsommes	1 non-market
	0.5	Cadeau recu	1 non-market
	5.5	Prestataires de services individuels	7 service from individual
Unspec.	0.0	Other	99 n.a./other
		RWANDA	
Formal	0.6	Supermarket/big shop	5 large stores
	4.6	Specialized shop	4 specialized shops
	2.4	Bar/restaurant	8 entertainment
Informal	13.5	Small shop/boutique	3 corner shops
	1.7	Individual	2 no store front
	0.8	Mobile seller	2 no store front
	12.5	Market	2 no store front
	11.5	Self production	1 non-market
	26.5	From a household	1 non-market
	13.1	Service provider	7 service from individual
Unspec.	12.7	Other	99 n.a./other
		SAO TOME	
Formal	5.2	Lojas modernas	5 large stores
	5.4	Grandes Lojas	5 large stores
	1.3	Outros comercios modernos	4 specialized shops
	0.5	Clinicas laboratorios medicos Hospitais	6 institutions
	0.8	Sector de transportes	6 institutions
	4.3	Prestates de servicios publicos	6 institutions
	0.9	Hotels, restaurantes, bares, cafes	8 entertainment
Informal	33.9	Quiosque / Quitanda	3 corner shops
	24.0	Mercado	2 no store front
	7.8	Vendedor Ambulante	2 no store front
	0.5	Prendas Recebidas	1 non-market
	0.9	Campo, mato	1 non-market
	1.9	Auto Consumo	1 non-market
	0.6	Autoabastecimento	1 non-market
	3.7	Prestates de servicios individuais	7 service from individual
	1.6	Candongueiro	7 service from individual
Unspec.	6.4	Other	99 n.a./other

Assigned	%	Original name	Classification
		SENEGAL	
Formal	2.0	Station service (carburants, lubrifiants,e	4 specialized shops
	1.8	Boulangerie, pâtisserie	4 specialized shops
	1.2	Service de transport public	6 institutions
	11.6	Bar, café, restaurant, hôtel	8 entertainment
Informal	40.3	Boutique de quartier	3 corner shops
	0.6	Marchand Ambulant	2 no store front
	10.8	Kiosque ou échoppe au quartier	2 no store front
	16.3	Marchés	2 no store front
	3.5	Cadeau reçu en nature	1 non-market
	1.6	Bien ou service autoproduit	1 non-market
	1.2	Autres services privés	7 service from individual
	5.1	Service de transport privé	7 service from individual
Unspec.	1.4	Other	99 n.a./other
		SERBIE	
Formal	8.9	Hypermarket	5 large stores
	23.8	Specialized shop	4 specialized shops
	2.9	Discounted shop	4 specialized shops
Informal	29.6	Minimarket	3 corner shops
	4.8	Market/open	2 no store front
	1.8	Gray economy	2 no store front
	5.3	Own production/Own business	1 non-market
	2.2	Gifts/received transfers	1 non-market
Unspec.	20.7	Other	99 n.a./other
		SOUTH AFRICA	
Formal	38.6	Chain store	5 large stores
	11.2	Other retailer	4 specialized shops
Informal	0.9	Street trading	2 no store front
	2.7	Other	2 no store front
	0.6	From a household	1 non-market
Unspec.	45.7	Other	99 n.a./other
		TANZANIA	
Formal	0.5	Duka kubwa(Department stores)	5 large stores
	37.3	Shop	4 specialized shops
Informal	2.4	Street vendor	2 no store front
	22.8	Market	2 no store front
	4.7	Other household	1 non-market
	0.5	Gift from other household	1 non-market
	16.0	Produced by household	1 non-market
	1.8	Gift or free	1 non-market
Unspec.	13.7	Other	99 n.a./other
		TUNISIA	
Formal	1.3	Hyper, supermarche	5 large stores
	67.8	Boutique privee	4 specialized shops
Informal	1.2	Point de vente marche	2 no store front
	4.5	Ambulant	2 no store front
	1.6	Cadeau	1 non-market
	1.3	Auto production	1 non-market
Unspec.	22.2	Other	99 n.a./other
		URUGUAY	
Formal	11.7	autoservicio, cadena de supermercados	5 large stores
	1.0	shopping o galeria	5 large stores
	0.5	barraca, ferreteria, vidrieria	4 specialized shops
	1.3	casa de electrodomesticos, telefonos	4 specialized shops
	0.7	verduleria, puesto, fruteria	4 specialized shops
	0.9	zapateria, marroquineria, talabarteria	4 specialized shops
	2.3	merceria, tienda	4 specialized shops
	1.5	panaderia, confiteria	4 specialized shops
	2.6	carniceria, polleria, pescaderia	4 specialized shops
	0.7	farmacia, perfumeria, panalera	4 specialized shops
	0.8	fuera del pais	6 institutions
	0.5	cantina, trabajo, colegio	8 entertainment
	0.9	restaurante, parrillada	8 entertainment
Informal	7.8	almacen	3 corner shops
	0.1	almacen de ramos generales	3 corner shops
	1.0	vendedor ambulante, puesto callejero, carr	2 no store front
	0.7	quiosco, salon	2 no store front
	1.5	feria vecinal	2 no store front
	0.5	distribuidor o repartidor a domicilio	1 non-market
	0.8	bar, pizzeria	9 informal entertainment
Unspec.	59.6	other	99 n.a./other

C Theory Appendix

C.1 Proof of expression (4)

Under our assumption that $p_{j1} \approx p_{j0}, \forall j$ we can write the uncompensated elasticity of product j as a function of the uncompensated elasticities of varieties $j1$ and $j0$ and the cross-variety price elasticities in the following way:

$$\epsilon_j = \epsilon_{j1}(1-\alpha_j) + \epsilon_{j0}\alpha_j + \epsilon_{j1,0}(1-\alpha_j) + \epsilon_{j0,1}\alpha_j \tag{10}$$

where $\alpha_j = \frac{p_0 x_0}{px}$ is the share of informal consumption in total consumption of the product and $\epsilon_{j0,1}$ it the elasticity of demand for the informal variety with respect to the price of the formal variety.

Writing ϵ_j^C the compensated price elasticity of product j the Slutsky equation is $\epsilon_j = \epsilon_j^C - \eta_j s_j$. Using this and the equalities $\eta_j = \eta_{j1}(1-\alpha_j) + \eta_{j0}\alpha_j$ and $s_j = s_{j1} + s_{j0}$ we obtain:

$$\epsilon_j^C = \epsilon_{j1}^C(1-\alpha_j) + \epsilon_{j0}^C\alpha_j + \epsilon_{j1,0}^C(1-\alpha_j) + \epsilon_{j0,1}^C\alpha_j \tag{11}$$

Slutsky symmetry implies $\epsilon_{j1,0}^C(1-\alpha_j) = \epsilon_{j0,1}^C\alpha_j$. Using our assumptions of equal compensated cross-variety elasticity across products ($\epsilon_{j0,1}^C = \tilde{\epsilon}^C, \forall j$) , equal compensated own-price elasticity across varieties within products ($\epsilon_{j1}^C = \epsilon_{j0}^C, \forall j$) and equal compensated own-price elasticities across products ($\epsilon_j^C = \epsilon^C, \forall j$), and re-arranging, we obtain:

$$\epsilon_{j1}^C = \epsilon^C - 2\tilde{\epsilon}^C\alpha_j \tag{12}$$

To obtain an expression for the compensated price elasticity ϵ_{j1}, the parameter of interest, we use the Slutsky equation again and obtain:

$$\epsilon_{j1} = \epsilon^C - 2\tilde{\epsilon}^C\alpha_j - \eta_{j1}s_{j1} \tag{13}$$

C.2 Proof of Proposition 1

In what follows we assume that all product and variety Engel curves are linear with respect to log household income. Taking a first-order Taylor approximations around $y^i = \bar{y}$ and assuming $s_j^i(\bar{y}) = s_j$ we can write $s_j^i = s_j + \beta_j(\phi^i - 1)$ where β_j is the slope of the EC for product j. We can then write the tax rate on a product j when all varieties are taxed as:

$$\tau_j^* = \frac{\int_i (\bar{g} - g^i)\phi^i (1 + \beta_j \frac{\phi^i - 1}{s_j})}{-g\epsilon_j} \tag{14}$$

The change in the optimal rate over the development path, holding efficiency considerations constant ($\partial \epsilon_j = 0$), can therefore be written as:

$$\partial \tau_j^* = \frac{\beta_j}{s_j}\left(\frac{\partial \beta_j}{\beta_j} - \frac{\partial s_j}{s_j}\right)\frac{\int_i (\bar{g} - g^i)\phi^i(\phi^i - 1)}{-g\epsilon_j} \tag{15}$$

where $\frac{\int_i (\bar{g} - g^i)\phi^i(\phi^i - 1)}{-g\epsilon_j} > 0$.

Similarly we can write the tax rate on a product j when only formal varieties are taxed as:

$$\tau_j^{**} = \frac{\int_i (\bar{g} - g^i)\phi^i (1 + \beta_{j1} \frac{\phi^i - 1}{s_{j1}})}{-g\epsilon_{j1}} \tag{16}$$

and

$$\partial \tau_j^{**} = \frac{\beta_{j1}}{s_{j1}}\left(\frac{\partial \beta_{j1}}{\beta_{j1}} - \frac{\partial s_{j1}}{s_{j1}}\right)\frac{\int_i (\bar{g} - g^i)\phi^i(\phi^i - 1)}{-g\epsilon_{j1}} \tag{17}$$

The first part of proposition 1 states that the redistribution gain from taxing all formal varieties uniformly is decreasing over the development path, ie that equity considerations push the optimal uniform rate τ_1^* downwards with development. Applying the above to the case of τ_1^*, we find that that, holding efficiency considerations constant, $\partial \tau_1^* < 0$ if the following condition holds:

$$\frac{\partial \beta_1}{\beta_1} < \frac{\partial s_1}{s_1} \tag{18}$$

The negative slope of the Informality Engel Curves implies $\beta_1 > 0$. Equity considerations therefore push the optimal uniform rate down over the development path as long as the formal aggregate budget share s_1 increases faster than the slope of the Engel curve for all varieties β_1, which is minus the slope of the Informality Engel Curve depicted in the paper. This proves the first part of the proposition.

To prove the second part of the proposition, which relates to how the efficiency cost of taxing all formal varieties changes over the development path, start from expression (4) in the paper for a 'product' consisting of all formal varieties. Writing η_1 this product's income effect, s_1 it's budget share, α the share of all formal varieties in total consumption, we obtain:

$$\epsilon_1 = \epsilon^C - 2\tilde{\epsilon}^C\alpha - \eta_1 s_1 = \epsilon^C - 2\tilde{\epsilon}^C\alpha - \eta_1(1-\alpha) \tag{19}$$

where the last expression is obtained by using $s_1 = (1-\alpha)$. The change in ϵ_1 over the development path, under our assumptions, can therefore be written as:

$$\partial\epsilon_1 = \partial\alpha(-2\tilde{\epsilon}^C + \eta_1) \tag{20}$$

As shown in the paper the size of the informal sector falls with development, so $\partial\alpha < 0$. The term is therefore positive as long as $\tilde{\epsilon}^C > \frac{\eta_1}{2}$. When this condition is met the price elasticity of demand for formal varieties increases over the development path, so the efficiency cost of taxing these varieties falls.

C.3 Proof of proposition 2

We start by proving the first part of the proposition, which states under which conditions the redistribution gain from taxing food less than non-food products increases over the development path when all varieties can be taxed. This implies that, absent efficiency considerations, the optimal rate on food τ_F^* falls over development relative to the optimal rate on non-food τ_N^*. Using expression (15) above we obtain the following condition for $\partial\tau_N^* > \partial\tau_F^*$:

$$\frac{\beta_N}{s_N}\left(\frac{\partial\beta_N}{\beta_N} - \frac{\partial s_N}{s_N}\right) > \frac{\beta_F}{s_F}\left(\frac{\partial\beta_F}{\beta_F} - \frac{\partial s_F}{s_F}\right) \tag{21}$$

Re-arranging and using the fact that $\beta_N = -\beta_F$ and $s_F = 1 - s_N$ when all varieties can be taxed, we obtain

$$\frac{\partial\beta_N}{\beta_N} > \frac{\partial s_N}{s_N}\frac{1-2s_N}{s_N(1-s_N)} \tag{22}$$

Where, as shown in the paper, we have $\partial s_N > 0$ and $\beta_N > 0$. There are two cases of interest, depending on which of s_N or s_F is highest. When households spend on aggregate more on non-food than on food products ($s_N > 0.5$), as is the case in most countries in our sample, the right-hand-side of the expression is negative, so that the condition holds as long as β_N (minus the slope of the food Engel curve) does not fall too much over the development path. This is the case described in the first part of proposition 2 in the text. Note however that if $s_N < 0.5$ the condition can still hold as long as $\partial\beta_N$ is positive and s_N does not increase too much relative to β_N. This case is less empirically relevant, but note that it can be explained using the intuition developed in the paper. All else equal (in particular, holding the slope of Engel curves constant), the redistribution potential of taxing food and non-food at different rates is minimized when

food and non-food are consumed in same proportions in the aggregate. An increase in the slope of the non-food Engel curve (or, equivalently, a steepening of the food Engel curve), all else equal, increases this redistribution potential. The redistribution potential will thus fall in a context in which s_N starts below 0.5 and increases, unless the slope of the Engel curve increases enough to compensate for the increase in s_N.

The change in the efficiency cost of taxing food less than non-food products over the development path, discussed in the second part of proposition 2, is governed by the relative values of $\partial\epsilon_N$ and $\partial\epsilon_F$. The uncompensated price elasticity of demand for product j is given by:

$$\epsilon_j = \epsilon^C - \eta_j s_j \tag{23}$$

Under our assumptions the change in this elasticity can be written as:

$$\partial\epsilon_j = -\eta_j \partial s_j \tag{24}$$

As shown in the paper over the development path $\partial s_N = -\partial s_F > 0$, which implies $\partial\epsilon_N < \partial\epsilon_F$. The efficiency cost of taxing non-food products therefore increases relative to that of taxing food products: efficiency considerations push the optimal rate on food up relative to that non non-food products over the development path.

C.4 Proof of Proposition 3

To prove the first part of proposition 3, we use expression (17) above to write the change in optimal rates on food and non food when only formal varieties can be taxed, $\partial\tau_N^{**}$and $\partial\tau_F^{**}$. The condition for $\partial\tau_N^{**} > \partial\tau_F^{**}$ can be written as:

$$\frac{\beta_{1N}}{s_{1N}}\left(\frac{\partial\beta_{1N}}{\beta_{1N}} - \frac{\partial s_{1N}}{s_{1N}}\right) > \frac{\beta_{1F}}{s_{1F}}\left(\frac{\partial\beta_{1F}}{\beta_{1F}} - \frac{\partial s_{1F}}{s_{1F}}\right) \tag{25}$$

Re-arranging, we obtain:

$$\frac{\partial\beta_{1N}}{s_{1N}} - \frac{\partial\beta_{1F}}{s_{1F}} + \frac{\partial s_{1F}}{s_{1F}}\frac{\beta_{1F}}{s_{1F}} - \frac{\partial s_{1N}}{s_{1N}}\frac{\beta_{1N}}{s_{1N}} > 0 \tag{26}$$

This expression will hold as long as the non-food formal Engel curve slope increases more (or decreases less) than the food formal Engel curve slope ($\frac{\partial\beta_{1N}}{s_{1N}} > \frac{\partial\beta_{1F}}{s_{1F}}$) and the non-food formal budget share doesn't increase too fast relative to the food formal budget share.

The second part of the proposition states under what conditions the efficiency cost of taxing non-food products increases relative to that of taxing food products in a world

in which only formal varieties can be taxed, ie $\partial\epsilon_{F1} > \partial\epsilon_{N1}$. Under our assumptions the change in the price elasticity of the formal variety of product j over the development path is given by:

$$\partial\epsilon_{j1} = -2\tilde{\epsilon}^C \partial\alpha_j - \partial s_{j1} \tag{27}$$

The condition $\partial\epsilon_{F1} > \partial\epsilon_{N1}$ is satisfied when:

$$2\tilde{\epsilon}^C(\partial\alpha_F - \partial\alpha_N) < \partial s_{N1} - \partial s_{F1} \tag{28}$$

This condition holds as long as, over the development path, the informal share of food consumption falls faster than that of non-food consumption ($\partial\alpha_F < \partial\alpha_N$) and the aggregate budget share of formal food varieties does not increase too fast relative to that of non-food varieties.

C.5 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:

$$q_{jl} = \frac{\epsilon_{jl}}{\epsilon_{jl} - 1} \frac{w}{\phi_{jl}} \tag{29}$$

This implies the following expression for consumer prices:

$$p_{j1} = (1 + t_j) \frac{\epsilon_{j1}}{\epsilon_{j1} - 1} \frac{w}{\phi_{j1}} \tag{30}$$

and

$$p_{j1} = \frac{\epsilon_{j0}}{\epsilon_{j0} - 1} \frac{w}{\phi_{j0}} \tag{31}$$

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

C.6 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:

$$x_{jl} = \left(\sum_k \alpha_{jlk} x_{jlk}^{\frac{\rho-1}{\rho}} \right)^{\frac{\rho}{\rho-1}} \tag{32}$$

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:

$$p_{jl} = (1 + t_j f_{jl}) \frac{P_{jl}}{\phi_{jl}} \frac{\epsilon_{jl}}{\epsilon_{jl} - 1} \tag{33}$$

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:

$$P_{jl} = \left(\sum_k \alpha_{jlk}^{\rho} p_{jlk}^{1-\rho} \right)^{1/(1-\rho)} \tag{34}$$

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:

$$p_{jlk} = (1 + t_k f_k (1 - f_{jl})) w \frac{\rho}{\rho - 1} \tag{35}$$

Combining expressions (33), (34) and (35), 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:

$$\frac{\partial p_{j1}}{\partial t_j} \frac{1 + t_j}{p_{j1}} = 1 \tag{36}$$

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

$$\frac{\partial p_{j0}}{\partial t_j} \frac{1 + t_j}{p_{j0}} = s_{j0F} \tag{37}$$

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

$$s_{j0F} = \sum_k f_k \alpha_{j0k}^{\rho} P_{j0}^{\rho-1} p_{j0k}^{1-\rho} \tag{38}$$

D Calibration Appendix

This sub-section explains how we calibrate tax rates under the three optimal policy scenarios defined in expressions (6), (8) and (9). Table 4 summarizes our choice of calibration parameters.

First, we calibrate several parameters directly from our data: we use the observed budget shares described in Section 3, total household expenditure to proxy for household income and the the observed informal shares of consumption for each good and country. We relax our theoretical assumptions that Engel curves are log-linear and that economic development does not affect within-country inequality, using instead the observed budget shares and income distributions in each country. Note that our model calls for using budget shares observed under a counterfactual 'no tax or transfers' scenario. We do not attempt to adjust observed budget shares to take into account the fact that they are affected by current tax systems as this would require an in-depth understanding of the tax and transfer system in each country in our sample which is beyond the scope of this paper.

We similarly use our data to obtain estimates of income elasticities for all goods and varieties.To obtain an estimate of the income elasticity of demand for the formal variety, η_{j1} we use our estimates of the slope of the formal Engel curve for good j, β_{j1}, and the expression $\eta_{j1} = 1 + \frac{\beta_{j1}}{s_{j1}}$. We similarly obtain income elasticities η_j using $\eta_j = 1 + \frac{\beta_j}{s_j}$.

Second, we use existing literature to calibrate the remaining parameters.There are no estimates of the cross-price compensated elasticity of demand between formal and informal varieties $\tilde{\epsilon}^C \alpha_j$ so we use estimates of the elasticity of substitution in consumption across stores of different types available in the literature. The cross-price elasticity

is related to this elasticity of substitution σ in a CES utility function by the expression $\epsilon^C_{0,1} = \sigma s_0$ where s_0 is the share of informal consumption of total consumption of the good. Faber and Fally (2017) estimate an elasticity of substitution between large and small stores in the US of 2.2, Atkin et al. (2018b) estimate the elasticity of substitution between foreign and domestic supermarkets and find estimates in the 2-4 range. We therefore use 3 as our baseline of σ. For an average value of s_0 of 0.5 this yields a baseline value of $\tilde{\epsilon}^C \alpha_j$ of 1.5, we consider the range 1-2 as a robustness check. We set a value of -0.7 for the own-price compensated elasticity of goods. Together, these parameters yield values for the own-price uncompensated elasticity of goods (calibrated using expression (33) that are in the $[-2, -0.5]$ range, in line with estimates from the literature (see for example Deaton et al., 1994).

Finally, we specify government preferences by setting the same social welfare weight for households in a given decile of the household expenditure distribution in each country. Our specification implies that governments place ten times more weight on income received by households in the poorest decile than in the richest decile. In all countries the richest decile is assigned a weight g_i equal to 1, the second richest decile a weight equal to 2, the third a weight equal to 3, and so on, until the poorest decile, which is assigned a weight equal to 10.