Pontifícia Universidade Católica
do Rio de Janeiro

Jonas Gouveia de Azevedo Maia

Informality and Consumption of Formal Goods

Dissertação de Mestrado

Thesis presented to the Programa de Pós–graduação em Economia, do Departamento de Economia da PUC–Rio in partial fulfillment of the requirements for the degree of Mestre em Economia.

Advisor : Prof. Yvan Pierre Becard
Co-advisor: Prof. Gustavo Maurício Gonzaga

Rio de Janeiro
April 2022

Pontifícia Universidade Católica
do Rio de Janeiro

Jonas Gouveia de Azevedo Maia

Informality and Consumption of Formal Goods

Thesis presented to the Programa de Pós–graduação em Economia da PUC–Rio in partial fulfillment of the requirements for the degree of Mestre em Economia. Approved by the Examination Committee:

Prof. Yvan Pierre Becard
Advisor
Departamento de Economia – PUC–Rio

Prof. Gustavo Maurício Gonzaga
Co-advisor
Departamento de Economia – PUC–Rio

Prof. Juliano Junqueira Assunção
Departamento de Economia – PUC–Rio

Prof. Cézar Augusto Ramos Santos
FGV – EPGE

Rio de Janeiro, April the 25th, 2022

Jonas Gouveia de Azevedo Maia

Bachelor in Economics, University of Brasília (UnB), 2019

Bibliographic data

Maia, Jonas Gouveia de Azevedo

Informality and Consumption of Formal Goods / Jonas Gouveia de Azevedo Maia; advisor: Yvan Pierre Becard; co-advisor: Gustavo Maurício Gonzaga. – 2022.

53 f: il. color. ; 30 cm

Dissertação (mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Economia, 2022.

Inclui bibliografia

1. Economia – Teses. 2. Informalidade. 3. Consumo. 4. Mercado de Trabalho. 5. Crescimento de Renda. 6. Brasil. I. Becard, Yvan Pierre. II. Gonzaga, Gustavo Maurício. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Economia. IV. Título.

CDD: 331.1

To my parents, for their invaluable support
and indispensable love.

Acknowledgments

I would like to thank my advisors, Prof. Yvan Becard and Prof. Gustavo Gonzaga, for their support and guidance that were key to this achievement. I also thank Prof. Juliano Assunção and Prof. Cézar Santos for the sharp and fruitful comments, and Prof. Nathalie Gimenes for kindly helping with the econometric challenges. Any mistakes are my own.

I thank my family for being a safe harbor in tough moments and always pushing me further. I thank the support of my kind-hearted and patient friends from PUC-Rio, specially Manuela, Maria, Marina, Victor, André, and Bárbara, who would carefully help me with multiple hard questions and hear long complaining audios, advising me wisely whenever I needed. My experience in PUC-Rio has been very pleasant and joyful thanks to them and all the others colleagues.

A special thanks goes to Roberto Hsu for helping with the literature. Also to my friends Pablo and Lia for being by my side in my ups and down. The whole economics department at PUC-Rio has excelled in teaching quality and administrative support, so I am grateful for the unique opportunity of attending the Masters program here.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001 and the Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq.

Abstract

Maia, Jonas Gouveia de Azevedo; Becard, Yvan Pierre (Advisor); Gonzaga, Gustavo Maurício (Co-Advisor). **Informality and Consumption of Formal Goods**. Rio de Janeiro, 2022. 53p. Dissertação de Mestrado – Departamento de Economia, Pontifícia Universidade Católica do Rio de Janeiro.

As economies develop and grow, their informal sector shrinks. The literature emphasizes a number of supply-side causes (higher costs of informality for larger and capital-intensive firms, improved state enforcement capacity, higher levels of education) to explain this phenomenon. This thesis contributes to the debate by proposing a new, demand-side explanation. We argue that the rise in formality can be explained, in part, by a rise in demand for formal goods and services from households whose income is growing. Using Brazilian household expenditure survey data, we document that in the cross-section, higher-earning households consume a larger fraction of formal goods (7 percentage points as income doubles). We also show that, over time, formal consumption increases together with income. We attempt to provide a causal estimate by analysing exogenous increases in the minimum wage. Last, we propose a theoretical discussion on the type of preferences consistent with this observed behavior.

Keywords

Informality; Consumption; Labor Market; Income Growth; Brazil.

Resumo

Maia, Jonas Gouveia de Azevedo; Becard, Yvan Pierre; Gonzaga, Gustavo Maurício. **Informalidade e Consumo de Bens Formais**. Rio de Janeiro, 2022. 53p. Dissertação de Mestrado – Departamento de Economia, Pontifícia Universidade Católica do Rio de Janeiro.

À medida em que uma economia se desenvolve e cresce, seu setor informal encolhe. A literatura enfatiza um conjunto de causas ligadas ao lado da oferta (maior custo de operar informalmente para firmas maiores e intensivas em capital, maior capacidade fiscalizadora do estado, e maiores níveis educacionais) para explicar esse fenômeno. Esta tese contribui para o debate propondo uma nova explicação, olhando para o lado da demanda. Argumenta-se que o aumento nos níveis de formalização pode ser explicado em parte pelo aumento da demanda por bens formais por parte das famílias cuja renda está crescendo. Usando dados de três aplicações da Pesquisa de Orçamento Familiares (POF), documentamos que no cross-section famílias de maior renda consomem uma maior proporção de bens formais (7 pontos percentuais a mais quando a renda dobra). Também mostramos que, ao longo do tempo, o consumo de bens formais aumenta com a renda. Buscamos também prover uma estimativa causal por meio da análise de aumentos exógenos do salário mínimo. Por fim, propomos uma discussão téorica acerca do tipo de preferências consistentes com o comportamento observado.

Palavras-chave

Informalidade; Consumo; Mercado de Trabalho; Crescimento de Renda; Brasil.

Table of contents

List of figures

List of tables

List of Abreviations

BF - Bolsa Família (cash transfer program)

LPM - Linear Probability Models

IBGE - Instituto Brasileiro de Geografia e Estatística (Brazilian statistics bureau)

PME - Pesquisa Mensal de Emprego (Brazilian employment survey)

PNAD - Pesquisa Nacional por Amostra de Domicílios (household sample survey)

POF - Pesquisa de Orçamento Familiares (Brazilian households consumption survey)

The mind and the world are opposites, and vision arises where they meet. When your mind doesn't stir inside, the world doesn't arise outside. When the world and the mind are both transparent, this is true vision. And such understanding is true understanding.

Bodhidharma, *Wake-up Sermon.*

1
Introduction

There is a strong positive correlation between per capita income and the size of the formal economy across different countries (La Porta and Shleifer, 2014; Perry et al., 2007). This pattern can also be observed across states of the same country (Gerard and Gonzaga, 2021). Over time, as countries develop, the size of their informal sector shrinks. For instance, Brazil experienced a 12% increase in formality rates from 2001 to 2013, as seen in Figure 1.1[1]. This period was also marked by high increases in GDP, accumulating a total real growth of 54.5% (36.2% per capita). The concomitant occurrence of both events raises questions of whether this correlation hides causal mechanisms involved. The traditional literature discusses only *supply side* causality, in the sense that more productive, larger, and capital-intensive firms self allocate to the formal sector. Established facts that support this are that the costs of operating informally increase in size, formal firms grow more than their informal counterparts, more productive firms employ more workers, and informal firms face higher capital costs (Ulyssea, 2010; Haanwinckel and Soares, 2020).

This project proposes a new causal mechanism candidate focusing on the *demand side*. We argue that, as households become richer, they consume proportionally more formal goods. In turn, as formal firms face a higher increase in demand when income grows, they will experience faster growth and hire more workers than informal firms, leading more firms to enter the formal sector. This mechanism boosts formalization as the informal sector shrinks. We evaluate the impact of income increases on households' allocation choices both in the cross-section and in different periods, employing analytical tools to establish this causal mechanism through which higher income causes higher greater demand for formal goods and hence higher formality rates.

To study our hypothesis empirically, we focus on the mentioned case

[1]We use data from the National Household Sample Survey (PNAD). This survey was not conducted in 2000 and 2010 due to the realization of the national census in both years, and in 1994 due to reorganization. Multiple sources document a similar increase. For example, Maurizio and Monsalvo (2021) documents a bottom of formality of 64% in 2004 to a peak of 76% in 2015 in urban areas using the Brazilian Monthly Employment Survey (PME); Fairris and Jonasson (2020) state a 10 p.p. increase of formality rates from 2000 to 2010 from census data; Haanwinckel and Soares (2020) present an increase from 71.9% in 2003 to 82.7% in 2012 for the private sector.

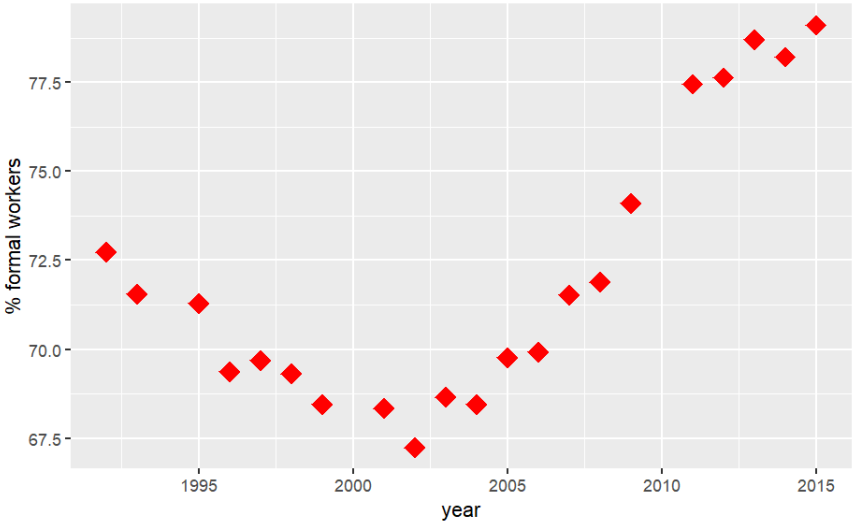

Figure 1.1: Formal employment as share of total

of Brazil. Using the Brazilian consumption survey (*Pesquisa de Orçamento Familiar* - POF) in different periods, we study within each period the distribution of formal share per household across different income percentiles — measuring the income elasticities and obtaining regression coefficients —, and across periods the evolution of the total share of formal goods consumption. In order to obtain concrete measures, we apply the recent methodology proposed by Bachas et al. (2020) using the three feasible samples of the Brazilian consumption survey: the 2002-2003, 2008-2009, and 2017-2018 surveys[2]. In this sort of survey, households do not declare whether the consumption good came from a formal or informal source — and it is likely that consumer often does not know. To circumvent this, their methodology proposes a proxy for the formalization status of a good: the place of its purchase, as discussed in Chapter 2.

In our study of the cross-section relations, we find that the share of formal goods' consumption is higher for wealthier households. In fact, our regression results state that a 100% increase in per capita income is associated with a 7.18 p.p. higher share in the 2017-18 sample, 9.13 p.p in the 2008-09 sample, and 8.8 in the 2002-03 sample. The coefficients are still large and significant when we control for observable variables. Furthermore, we document that between the 2002-2003 and 2008-2009 surveys the share of formal consumption increased by 1.4 p.p., whereas between the 2008-2009 and the 2017-2018 surveys there was a drop in this share at our benchmark specification. Using only a narrow basket of goods, comprised only of food goods, to calculate the difference in shares yields that the increase between the first two surveys was 5 times higher, and

[2]In POF, each household is accompanied for 12 months.

between 2008-2009 and 2017-2018 it was actually positive at 2.19 p.p.

Moreover, we investigate a candidate for exogenous income shock, the minimum wage increase. We use a propensity score matching estimator to compare households that earn the minimum wage in 2002-2003 (which we denote as the control group) with the most similar household earning the minimum wage in 2008-2009 (the treated group). We find that our treated group in fact consumed a greater share of formal goods than the control group. We predict that, in response to the 37% increase in minimum wage during the 2002 to 2008 period, households must have increased the formal share of consumption by 0.6 p.p.

Theoretically, our statement that the ratio of consumption of formal and informal goods is not constant to income variation implies that preferences are non-homothetic. In Chapter 3, we study the implications of non-homotheticity in the choice of a representative utility function, as we argue that an important corollary of our mechanism is that common forms of the utility function such as CES cannot nest our findings and should not be used in models (e.g. search-and-match and DSGE) that aim to explain the formalization process. We also propose micro-founded arguments for the shift in consumption. This phenomenon can be thought of under a multitude of explanations. In general, goods in the formal sector may be preferred because regulation on superior types of goods makes them only available in the formal sector; high productive firms self allocate to the formal sector (separating equilibrium); similarly, formalization may work as a signal of quality, allowing for government inspections, brand visibility, wages payment commitment, etc.

This thesis relates to three strands of the literature: first, the traditional labor economics studies supporting *supply-side* correlations between formalization and economic growth; second, the development literature which acknowledges differences in sectoral composition in the path of development, and third, the literature on the causes of the formalization experience in Brazil during the 2000s. The first strand was synthesized earlier, and highlights the value-added of this paper to the literature: papers discussing *demand-side* causality are hard to find and, to our knowledge, no paper states the role of non-homothetic consumption, tries to quantify the contribution of such a channel empirically or to extend the argument for the short run, and no paper calls attention for this factor to the Brazilian experience.

We relate to the second strand to the extent that non-homothetic preferences are known to development literature, accounting for differences in sectoral composition of production (e.g. industries, service, and agricultural

shares)[3]. In fact, La Porta and Shleifer (2014) claim that the presence of a large informal sector in low-income countries is related to demand constraint for modern production technologies, supporting a "dual view" through which formal and informal firms serve different customers. Şirin Saraçoğlu (2008) theoretically separates goods by sector (formal, informal, and agriculture) and introduces demand shift as households get wealthier in a development process. She calibrates the model for Turkey in 2000 using national accounts and simulates it to a steady state.

Last, our third strand concerns the overall causes of the size of the informal sector. Disregarding correlations with income, many factors are mentioned as influencing the high level of informality in developing countries. Examples are high entry costs in the formal sector (Auriol and Warlters, 2005), enforcement capacity (Almeida and Carneiro, 2012), tax liabilities (Rocha et al., 2018), education, minimum wage and cash transfer program (Haanwinckel and Soares, 2020; Fairris and Jonasson, 2020), and cost of financial intermediation (D'Erasmo, 2013). From a holistic point of view, there are two papers worth mentioning with distinguished approaches to quantify the relative importance of each event in causing formalization in Brazil. Haanwinckel and Soares (2020) propose a search-and-match model of the labor market accounting for different skill levels for workers and different productivity for firms. In turn, Fairris and Jonasson (2020) use an empirical approach to decompose the change in informality in the variation of their determinants, including also trade liberalization and industry composition (we discuss their paper further in Appendix A.1 as we try to identify which production sector contributed the most to the formalization process, concluding in favor of the retail sector). The fact that they try to capture every factor contributing to a bigger formal sector and don't account for the income effect on households' consumption allocation was a key motivation for this thesis.

The next Chapter (2) details our methodology and empirical analysis for Brazil, then Chapter 3 discusses hypotheses for non-homethetic preferences over formal and informal goods and also their implications for the utility function format, and the last Chapter 4 concludes.

[3]One example of advocacy for non-homothetic preferences is the Engel's Law, which states that the percentage of income spent on food decreases as households get richer.

2
Empirical analysis

Bachas et al. (2020) propose an innovative methodology for computing informal consumption based on Consumption Surveys. To quantify the correlation between income and share of goods bought from the informal sector, they use the **type of store** in which purchase occurs as a proxy for formal or informal establishments. This approach can be justifiable upon premises of, for instance, the status of each kind of place, the typical size, the structure etc.[1] Altogether they join evidence of downward slopes for 32 countries, including Brazil in 2008 and 2009[2]. Also using the Brazilian consumption survey (POF), we extend their analyses by comparing the shares of informal consumption for households in three periods: the first encompasses the survey conducted in 2002 and 2003 (shortened to 2002-03), the second is the same 2008 and 2009 (2008-09) survey, and the third consists of the 2017 and 2018 (2017-18) survey. We discuss the adaptations necessary from the mentioned paper in Appendix A.2.

2.1
Data

The data consists of the three consumption surveys (POFs) together with the classification of each store type. Earlier surveys, such as the 1995-96 survey, had very few designations of places of purchase, impeding concrete or comparable analyses. The number of places of purchase registered by POF is 624 in the 2002-03 survey, 774 in the 2008-09 survey, and 794 in the 2017-18 survey. Table A.1 in Appendix A.2 decomposes aggregate consumption across households for broad types of stores. Following the classification in Bachas et al. (2020), the idea is straightforward: places of purchase that denote large stores and specialized stores are denoted formal — in line with the correlation of firm size and formalization status mentioned in the Introduction (Chapter 1) — and on the other hand, corner stores, non-brick and mortar front, from farms or

[1] This approach is even more reasonable when we think that consumers themselves must proxy for formalization status of the place of purchase based on this kind of information, as they lack the power to enforce the establishment to reveal its type.

[2] Their goal is to discuss to which extent consumption taxes in fact increase inequality in developing countries, finding that, once informal consumption is accounted for, consumption taxes are actually redistributive.

persons, etc. are informal establishments. In the table, classification *1* denotes informal stores, whereas *0* denotes formal stores.

A shortcoming of this methodology which we innovatively circumvent is the fact that a large share of purchases comes from non-reported places of purchase. We believe this point is not addressed properly by Bachas et al. (2020), for most of those purchases were cleaned during datasets compatibilization. According to our estimates, those purchases comprise 39.17% of the total in the 2002-03 survey, 35.17% in the 2008-09 survey, and 34.38% in 2017-18. We approach this challenge in the benchmark by excluding those observations from the formal consumption share calculation, but subsection *A.3* in the Appendix presents robustness checks by ways of including those expenditures using a complement to the methodology: we use the specific type of good (among the more than 10,000 products cataloged by IBGE) to account for informality[3]. Among the unspecified, there were household services: this type of expense can typically be divided into formal or informal according to whether or not the employer contributes to the public pension system, so we are able to include them in our analysis.

2.2
Formal consumption in the cross-section

We begin presenting our results by computing the mean formal goods share of total consumption expenditures by income percentile in the three samples. Figure 2.1 displays each percentile according to the log of the mean income on the horizontal axis, and the respective mean formal share on the y-axis. We can see clearly a positive correlation, as higher percentiles tend to consume, on average, relatively more formal goods. We also observe that the trend is very close to linear, suggesting that OLS models can provide a good description of the relationship. Furthermore, the dots nearly overlap when we look at a given income level, meaning that income is a strong predictor of consumption formality shares across surveys. Nevertheless, it is noticeable that for lower income levels the green dots seem to be slightly above the others, and for higher incomes they are below, suggesting that the slope of the relationship between formal consumption share and the log of income is flatter in the 2017-18 survey, while in 2002-03 (red dots) and 2008-09 (blue dots) the relationship is similar. In Figure 2.2 we present the same information, but in the horizontal axis we compute the income percentile from 1st to 100th.

[3]Although we show this methodology is particularly useful in this limiting scenario, it is less appropriate than the original methodology when it is applicable since by using the type of good we ignore that the same good may be available in both sectors.

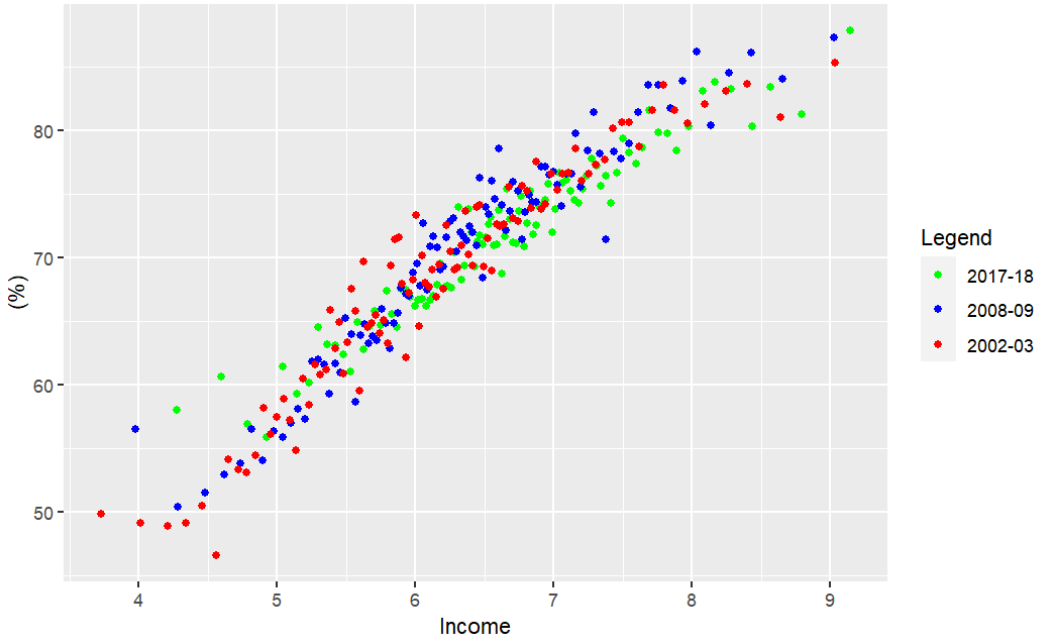

Figure 2.1: Percentage of formal consumption (y-axis) by log of per capita household income (x-axis). Dots are mean income percentile values. Last percentile omitted for better visualization.

To establish this correlation at the household level, we present the simple OLS regressions of the share of formal goods on the log of per capita households' income in the three surveys, as shown on the left side of Table 2.1. Figure 2.3 plots the regression fit for all the years by income percentile. We see that the formal share increases by an average of 7.18 p.p. in the 2017-18 sample, 9.13 p.p. in the 2008-09 sample, and 8.8 p.p. in the 2002-03 sample when per capita income doubles.

Nevertheless, the discussed *demand-side* channel may have had positive feedback from labor market dynamics. Also, increasing availability of formal products (derived from other causes mentioned in the literature review) may have increased formal products' supply, exacerbating general equilibrium effects. In the right side of Table 2.1, we control for other variables available at POF that potentially cause consumption formality choices: apart from per capita income as reported by households, we add the number of household members ("household size"), mean years of study (for household members with more than 16 years old, the bottom of the working-age population), the mean age of household members, the existence of paving in the street and access to sewage. Those last two are a potential proxy for the availability of formal stores nearby. We identify that the correlation between income and share of informal consumption is still large (coefficients of 2.716 for 2017-18, 4.084 for 2008-09, and 3.67 for 2002-03). All multiple regressions shown here include dummies for

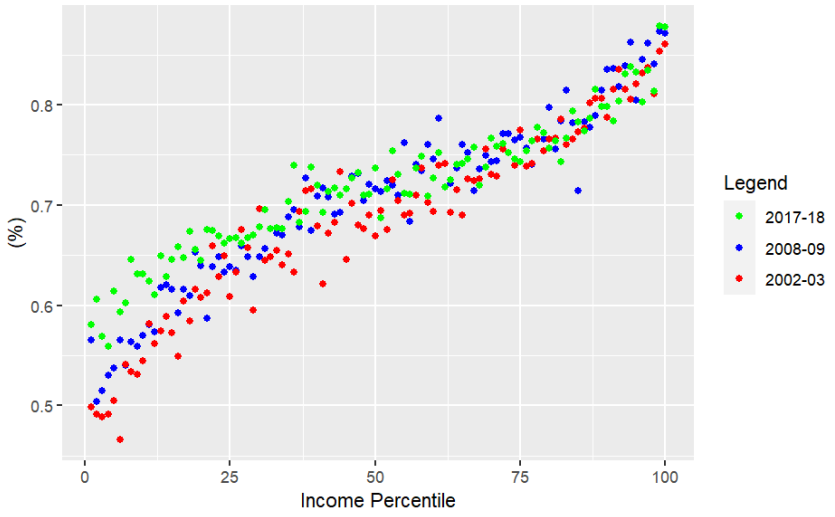

Figure 2.2: Mean formal consumption (y-axis) as share of total by income percentile (x-axis)

states, not shown for conciseness.

Table 2.1: Simple and multiple OLS regressions.

	Dependent variable:					
	Share of Formal					
	[2017-18]	[2008-09]	[2002-03]	[2017-18]	[2008-09]	[2002-03]
log of income	7.175***	9.129***	8.8***	2.716***	4.084***	3.668***
	(0.139)	(0.131)	(0.142)	(0.192)	(0.201)	(0.237)
household size				1.690***	1.395***	1.180***
				(0.103)	(0.092)	(0.101)
mean years of study				1.156***	1.308***	1.172***
				(0.045)	(0.048)	(0.061)
mean age				0.177***	0.120***	0.121***
				(0.011)	(0.011)	(0.014)
sewage				2.474***	2.676***	3.356***
				(0.315)	(0.341)	(0.424)
paving				3.024***	3.843***	3.037***
				(0.329)	(0.325)	(0.379)
States Dummies	NO	NO	NO	YES	YES	YES
Constant	19.01***	10.149***	14.382***	28.816***	21.580***	24.797***
	(1.013)	(0.849)	(0.839)	(1.457)	(1.470)	(1.473)
Observations	58,025	55,258	46,502	58,025	55,258	46,502
R^2	0.086	0.15	0.157	0.196	0.265	0.261
Adjusted R^2	0.086	0.15	0.157	0.195	0.264	0.261
Residual Std. Error	787.878	718.692	706.761	739.148	668.418	663.937
F Statistic	5,448.033***	9,922.244***	9,056.991***	441.244***	621.643***	513.069***
Note:			*p<0.1; **p<0.05; ***p<0.01			

Dependent variable is the share of expenditures on formal goods.

We estimate income elasticities for formal and informal goods in every survey. For informal goods in 2002-03, an increase in income of 1% would cause an increase of only 0.66% in informal goods expenditures, whereas formal goods expenditures would increase by 1.22%. The numbers are similar for 2008-09: 0.65% and 1.17%; and for 2017-18: 0.65% and 1.18%, meaning that informal goods are necessity goods, but formal goods are luxury goods. These results substantiate the analysis conducted in Chapter 3.

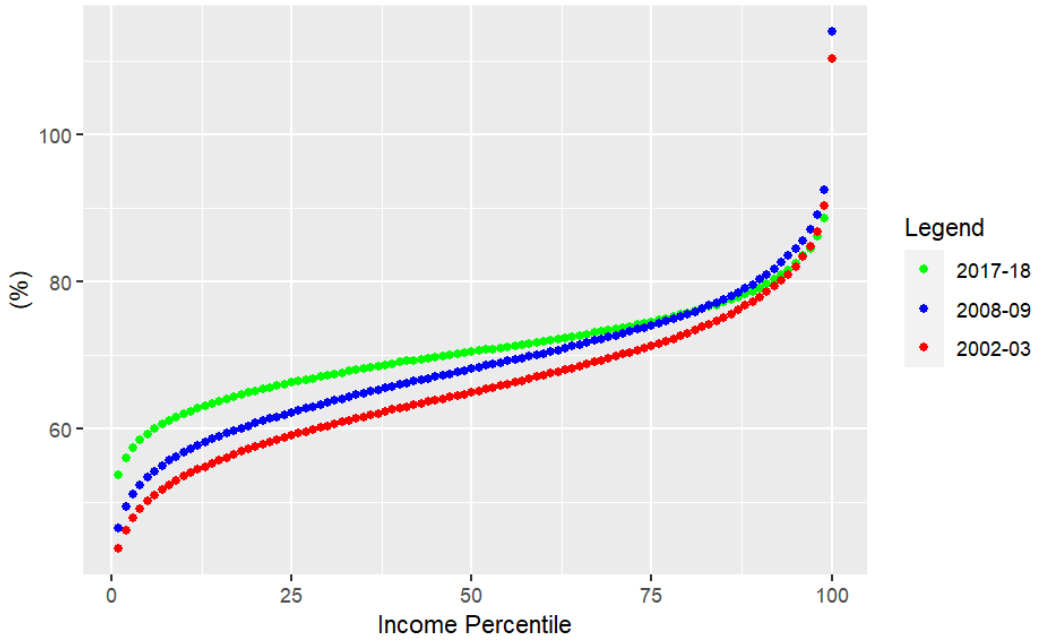

Figure 2.3: Formal share of consumption according to OLS projection by percentile

2.3 Formal consumption over time

As seen in Figure 2.2, the red dots usually lie below the blue dots, suggesting consumption became more formal in 2008 and 2009 as compared to 2002 and 2003, while the green dots are above the others for the lowest income percentile, but tend to concentrate in their middle for the higher percentiles. In fact, among all the households, total formal consumption expenditures corresponded to 74.2% of total expenditures in 2002-03, compared to 75.6% in 2008-09 — an increase of 1.4 p.p —, and 75.1% in 2017 — a fall of 0.5 p.p. compared to the last survey, representing the bigger weight of the highest income percentiles in total consumption.

To convey if those movements are robust, we can reinterpret the exercise in the following manner: instead of thinking of *one observation* comprised of the total formal expenditures divided by total expenditures, we can think of one realization of formal share per household, and the aggregate share would be nothing more than the mean of the shares weighted by each household total expenditures. We use this understanding to generate a confidence interval based on a two-step bootstrap approach. First, we sample the households of each survey (with replacement) 200,000 times, using the total expenditure as **probabilities**, in such a fashion that the distribution of each sample resembles the original distribution but weight-corrected, so the simple mean of this new distribution gives us a direct estimate of the desired parameter (the mean

formal consumption in each survey). We compute the simple mean of the new distribution and save the distance from the mean share found using the original distribution. In the second step, we repeat step one 5,000 times and find a distribution of the distances between the bootstrapped distributions' means and the original mean, from which we can compute the confidence interval directly. This yields that, for the 2002-03 survey the confidence interval of 95% lies within ±0.09 p.p. of the mean share stated above (74.2%), the same is true for the 2008-09 survey (whose mean share is 75.6%), and for the 2017-18 survey this confidence interval lies within ±0.1 p.p of the mean share (75.1%). These mean that the changes (of respectively +1.4 and −0.5) occur far beyond the confidence interval of each estimate, suggesting they are strongly significant.

If we take the coefficients of the simple OLS regression in Table 2.1 as a structural estimate, for the case we are more concerned with predicting consumption shares based on income, along with the given growth of 36.24% in real per capita GDP between 2001 and 2013, then consumption should have grown around 3.03 p.p. in this period (using the mean of the 2002-03, 2008-09, and 2017-18 coefficients as the relevant parameter). Notice that if a fall in consumption informality share corresponds one-to-one with a fall in labor informality, this statistic suggests the mechanism could account for up to 25% of the corresponding fall in the period. However, for reasons beyond the scope of this work, per capita GDP growth does not seem to change pairwise with the household income when we measure the latter using POF: while the 2002-2008 period experienced a real per capita GDP growth of 17%, the income growth between the two samples was 11.14%; yet much more stunning is the discrepancy between the real per capita GDP growth of mere 3% between 2008-2017 and the sample counterpart of 23.3%. See Figure A.5 in Appendix A.7 for a comparison of income per percentile in both periods as measured by the surveys, corrected by the IPCA price index. Nevertheless, even taking the 11.14% measure of income increase between 2002-2008, this income growth could account for 1 p.p. of the total 1.4 p.p increase in formal consumption reported in the period (using the mean of 2002-03 and 2008-09 coefficients). For the 2008-2017 period, using the 3% GDP per capita growth measure and the mean of the 2008-09 and 2017-18 regression coefficients, the contribution of income growth in formality consumption share should have been a growth of 0,24 p.p., and using the measure of 23.3% suggests the increase should have been 1.9 p.p. Instead, so far the evidence suggests a small decrease in consumption formality rate for the 2008-2017 period. Ahead we take a different approach to compute the share of formal consumption, which presents evidence favorable to an **increase** in such rate for the period.

If we consider the effect of income on formal goods consumption once the observable variables are controlled for, we can say this mechanism alone accounts for 0.43 p.p. (31%) of the 1.4 p.p. total variation in formal consumption share *observed* between 2002-03 and 2008-09, using the average of the partial coefficients of those samples. The prediction for 2017-18 in comparison to 2008-09 likewise should be of a 0.8 increase instead of the observed fall of 0.5 p.p. Comparing the labor formality share and our measure of GDP growth as expected income growth between 2001-2013 years, the projection using the mean of all coefficients suggests an increase in consumption formality of 1.26 p.p. against a labor formalization rate increase of 12 p.p.

2.4 Formal consumption by type of good

Instead of analysing the whole contingent of consumption goods, another interesting approach is to narrow our basket and address only a kind of good to which consumers are more confident to specify a place of purchase, and there is more discerning regarding formality or informality of such places. We decided to focus on food as meeting those criteria. Considering only this kind of goods, the share of food bought from formal places was 59.37% in 2002-03, versus 66.92% in 2008-09, a much larger increase than observed for overall goods. This suggests we may be underestimating the total increase in formal goods consumption. In turn, the 2017-18 share was 69.11%, in line with the prediction that emerged from the observed growth in income[4]. Figure 2.4 below depicts the mean share of formal consumption for food over the total by income percentile (left figure) and the OLS projection of this share on income (right figure). Appendix subsection *A.4* presents the table from multiple OLS estimation.

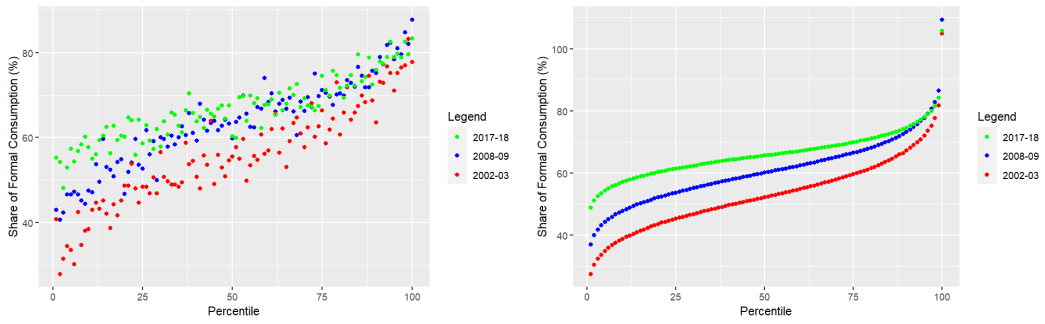

Figure 2.4: Mean share of formal consumption for food goods per income percentile (Left) and OLS projection of the share of formal consumption for food goods by income percentile (Right).

[4]Remember that using the observed income growth along with the mean of the 2008-09 and 2017-18 simple OLS regressions' coefficients suggested the increase should have been 1.9 between these years.

We can go further on this approach and detail the analysis for all the goods classifications used by Bachas et al. (2020) for the 2008-09 survey. The compatibilization of these to the 2002-03 and 2017-18 was largely straightforward, but doesn't exclude minor adaptations. Table A.7 in Appendix A.5 below summarize our findings.

Two movements happening together that can explain the observed shift in formal consumption share are observed in Table A.7: both external composition effects, in the sense that as income grows, households may consume relatively less of a type of good that is more often bought from the informal sector and buy more of types of good usually bought from formal firms; and the internal composition effect, in that the same type of good can be purchased more often from formal firms. For the list of types used, the second effect seems to be much more important, while the first is at most subtle. As shown, "Alimentation" is a clear example of the second type of movement, which has highlighted importance since they account for, on average, 27.2% of total expenditures over the surveys. Other categories that have also formalized through the years include "home maintenance" (10.02 p.p. from 2002-03 to 2017-18), "household goods repair" (25.71 p.p.), "appliances" (6.84 p.p.), "smoking" (18.62 p.p.), "recreation and sports" (15.04 p.p.), "products for personal use" (4.32 p.p.), "dental treatment" (3.47 p.p.) and "cleaning" (10.19 p.p.).

Referring to the cross-section results, among the 40 types of goods that had a relevant level of consumption coming from the informal sector in the three surveys (condition necessary to obtain meaningful coefficients in OLS regressions), only one had **negative** and significant results in all the surveys when controlling for observable variables: "home maintenance", which include from small repairs (floor, roof, wall painting, electricity, etc.) to external providers of household services (cleaner, maid, gardener, governess, janitor, etc.). Regardless, as argued, the across periods movement of this kind of good reveals they formalized over the years. Other 3 categories observed significant but negative results in 2 of the surveys — while the 3rd coefficient was insignificant —, which were: "real estate (maintenance)", "medical appointment", and "domestic gas", the first of which also include items for small repairs, and the second have low magnitude (coefficients are higher than -0.5) as the share of formal consumption of this kind is as high as 98%.

Despite these few exceptions, most commonly the types of good have high, significant and, **positive** coefficients observed in the cross-section even in the multiple regression, as is the case of "appliencies" (coefficients ranging from 2.91 to 7.19), "furniture and household goods" (4.72 to 6.65), "smoking" (4.17 to 7.25), "products for personal use" (from 2 to 4.52), "school articles" (from

2.28 to 2.55), the categories of clothes and shoewear (from 3.85 to 6.14 in all the 4 categories), "fabrics" (from 2.45 to 3.45), "jewelry and bijou" (6.55 to 6.74), "cellphones and accesories" (2.78 to 6.61), "vehicle acquisiton" (from 5.85 to 9.85), "cleaning" (3.67 to 5.74), "perfume" (2.51 to 5.1), "hair products"(2.39 to 3.56), "soap" (2.77 to 5.72), and "toys and games" (3.64 to 4.9), while almost all the others either have significant and positive but small coefficients or insignificant coefficients.

2.5 Formal consumption following minimum wage increases

To address causality, we should look for exogenous income shocks. This task is particularly difficult since the surveys are cross-sectional, present great time gaps, and are only representative on the national scale. In what follows, we evaluate a major country-wide policy that had a notable impact on households' income: the minimum wage increase. Also, in the Appendix section A.6, we evaluate the possibility of using *Bolsa Família* (BF), but we are not able to present a robust estimation for its effect and discuss the causes.

Our attempt focuses on the sharp increase in minimum wage during the 2002-2008 period, of no less than 37% in real terms. There is an operational difficulty that prevents us from including households that received the minimum wage for less than a year before the survey application. This is because the minimum wage varied annually, and thus, in the period concerning a single POF, there are different values to be targeted. This implies we must know for which month and year the worker is reporting the wage in the survey. In the 2008-09 survey, it is possible to deduce this information using the week of the research period in which the household was surveyed, plus the variable "last month worked". For 2002-03, this last variable is not reported. To circumvent this issue, knowing that the reference period for which work-related questions are answered consists of the preceding 12 months, we restrict the sample to those workers who actually worked all the 12 months, since the last month their wage is reported is the month before the survey application.

The strategy will consist of focusing on the households whose only wage earner receives an amount near the minimum wage (we allow for a 10% deviation to account for approximate/round answers). Table 2.2 proposes a Propensity Score Match to match minimum wage-earning families in 2008 (treatment group) to equivalent families in 2002 (control group). Our variable of interest is ("log of income growth"), which we compute by first taking the ratio of the real income between matched families, and then taking the log of this ratio. We add controls for the family member that earns the minimum wage — denoted by

"(worker)", including worker's years of study, age, sex (1 is female), and race (1 is non-white). The same household-level variables of previous estimations are included.

Table 2.2: Propensity score matching analysis of minimum wage earning families between 2002-03 and 2008-09.

	Dependent variable:
	Share of Formal
log of income growth	1.889**
	(0.819)
years of study (worker)	0.363
	(0.301)
household size (worker)	1.440**
	(0.572)
age (worker)	−0.100
	(0.070)
sex (worker)	3.989***
	(1.500)
race (worker)	−2.294
	(1.459)
mean years of study	0.912***
	(0.351)
mean age	0.317***
	(0.066)
paving	3.936**
	(1.544)
sewage	3.449**
	(1.747)
States Dummies	YES
Constant	34.078***
	(6.087)
Observations	3,256
R^2	0.183
Adjusted R^2	0.174
Residual Std. Error	678.552 (df = 3219)
F Statistic	20.049*** (df = 36; 3219)
Note:	*p<0.1; **p<0.05; ***p<0.01

The coefficient of "log income growth" can be compared to previous estimates[5]. Take for example the coefficients of the regression of the share of formal consumption on the log of income, which ranges between 2.716 to 4.084 when all controls are included. A coefficient of 2.716 means that when income doubles, the formal share will increase by 2.716 p.p. In our estimation,

[5]In fact, it has basically the same interpretation for small growth rates, since $\log(1+x) \cong x$, so a $x\%$ increase has approximately $x\%$ times the coefficient impact.

we use the log of the ratio of the incomes as the variable of interest. A 100% increase would mean that the ratio of the income in 2008 to income in 2002 was equal to 2, and in turn $\log(2) \cong 0.7$. Therefore, a 100% increase would imply a $0.7 * 1.915 = 1.34$ p.p. increase in the share of formal consumption. The observed growth of 37% then would represent, given $\log(1.37) \cong 0.315$, a 0.6 p.p increase in consumption formality during the period plausibly attributed to exogenous minimum wage growth. Remember that the total increase in the period was computed as 1.4 p.p, but with an observed overall increase of income measured of only 11.14%; if we take the parameter calculated for minimum-wage earning families to be structural for the whole economy, the causal effect would be estimated to be 0.2 p.p. or roughly 14.3% of the observed change.

3
Theoretical background

In this chapter, we discuss candidate explanations for the hypothesized non-homothetic preferences and their implications in terms of modeling.

Our empirical analysis shows that, following an increase in income, households increase their consumption of formal goods by more than one-for-one while they increase their consumption of informal goods by less than one-for-one. Because of that, we observe that the share of formal goods over total increases with income, and the calculated income elasticities are different from 1. This has important implications in terms of modeling since it suggests that the traditionally used homothetic preferences over formal and informal consumption goods might not best describe consumer behavior in developing countries.

In what follows, we first refer to the literature to present 3 candidate explanations for an income elasticity for formal goods greater than 1 (meaning they are luxury goods). Next, we establish that future articles trying to capture the causality relation studied here must use non-homothetic functions as a premise.

3.1
Why are formal goods luxury goods?

As claimed by La Porta and Shleifer (2014), informal firms produce "low-quality products for low-income costumers using little capital and adding little value" (p. 1). Also, Bachas et al. (2020) identify empirically a quality-price trade-off in which "formal stores offer high-quality varieties at higher prices" (p. 14). We refer to the literature and find 3 candidate explanations behind the notion that the good provided by formal firms is preferable and, in particular, when the same good is available in both sectors, the formal good is perceived as having better quality. These are regulation, the cost structure, and signaling.

The first explanation relies on the distinction between formal and informal firms. According to Perry et al. (2007), a traditional approach to defining informality declares as informal the agents that operate in the economy at the margins of current laws and regulations. Likewise, informal firms are defined as those *evading* state's norms, including those not complying with tax laws,

conducting illegal activities, competing unfairly, not paying workers' rights, etc. (Perry et al., 2007). In this sense, once regulation is at the origin of the split between sectors, it should be key to understanding the preferences over the goods. For instance, some products are just too risky to be bought without certification of safe origin, quality certification, and appropriate conduct. Whenever enforcement is hard to avoid, these goods will only be available in the formal sector, thus generating preferences based on availability. This is the case of zoos, airplane travels[1], drugstores, schools, hospitals, banking, petrol stations, private pension, insurance companies, etc.

The second explanation will bring back our discussion on *supply-side* arguments for the correlation between income growth and formal sector size. Based on our literature review in the Introduction (Chapter 1), firms with better product quality self-allocate to the formal sector for endogenous reasons. Suppose two firms producing the same product, all else equal but one with better good quality; this firm is thought to grow more and be more productive, and larger firms tend to formalize (Haanwinckel and Soares, 2020), as avoiding inspection becomes more costly and there are gains in the sense of accessing financial markets and lower capital cost. In this causal direction, an endogenous separating equilibrium is formed solely by the cost structure. Also, access to capital and its intrinsic technology must account for a better quality of formal firms (as argued by La Porta and Shleifer, 2014).

A third justification is that inspections may provide a signal that the variety of the same good offered by formal firms has a quality above minimum standings. Firms formalize for signaling. A way to see this is in the *market for lemons* perspective. In a continuum of unlimited suppliers of a good, quality must vary in a non-degenerated distribution; allowing for sufficient variance yields that risk-averse agents will fear lower realizations of quality — which could even represent negative utility in certain cases — and reduce consumption of this product on behalf of less variant substitute goods. Formalization in this scenery could bound the distribution at lower quality realizations — products have a certification of origin and are inspected by government regulatory agencies —, increasing demand for the good. Also in the fashion of signaling good quality, firms can differentiate themselves by establishing a brand and conducting advertising, a condition that increases firm visibility and plausibly is avoided by informal firms to prevent inspections. Another possibility is that consumers associate firms' commitment to pay wages and benefits regularly —

[1]In fact, according to our data the air transport sector (number 62) has the largest labor formality rate in both 2000 and 2010, close to 90%. Formal firms can hire informal workers in Ulyssea (2018) intensive margin of firm formality definition, but we are concerned with the extensive margin of firm formality only.

imposed by labor laws — with a better service from staff, in a *efficiency wages* context.

3.2 Preferences are non-homothetic

Our theory implies that informal goods should be *necessity goods*, whereas formal goods would be *luxury goods*. This proposition has implications for the form the household preferences upon formal and informal goods can be modeled, as they can't be represented by the commonly used homothetic preferences. We use this section to discuss this corollary of the work, assessing the fit of types of utility functions to the theory.

The definition that "as the consumer gets more income, he consumes more of both goods but proportionally more of one good (the luxury good) than of the other (the necessary good)" (Varian, 1992, p. 117) means that the earlier has an income elasticity of demand bigger than 1, and the latter lesser than 1[2]. As proved in Appendix A.8, homothetic preferences have the characteristic that when income increases by $x\%$, the consumption of each good (formal and informal) will both increase by the same $x\%$, and thus by definition both goods have income elasticity equal to 1. As a corollary, the share of expenditure on each good is going to be the same no matter the income. Our theory defends formal goods are *luxury* goods, as we have seen in Chapter 2 that when income doubles the share of formal goods increases by no less than 7 p.p. This means that these common forms of utility representation cannot nest the theory defended in this thesis when we separate formal and informal goods.

The upper graph in Figure 3.1 depicts the case of a preference over two goods x_i and x_j which is homothetic. Homothetic functions are extensively used in economic theory, even when the goal is to study the informal sector, for instance, Ulyssea (2010) uses a CES utility function for constructing the consumption good based on intermediary formal and informal goods. It is immediate to show that preferences depicted by Constant Elasticity of Substitution (CES) utility functions are homothetic (see Appendix A.8).

Then, we turn to 2 non-homothetic preferences to check if they can fit our theory. First, we propose that a classic example of non-homothetic utility functions, the quasilinear utility functions, cannot nest our empirical findings. As shown in Appendix A.8, for this utility to be compatible with the data, we

[2]The theory doesn't eliminate the possibility that, if the difference in quality is particularly pronounced and there are substitutes available in the formal sector, the household actually buys less quantity of informal goods when income increases, substituting a share of these goods previously bought, for which we would define them as *inferior goods*, and formal goods as opposingly *normal goods* (with income-elasticity respectively less than 0 and more than 0). In chapter 2, our empirical estimation suggests both goods are *normal goods*.

should have observed income elasticities of 0 for Informal goods (instead of the calculated elasticity of nearly 0.65 in all three surveys).

Attempting to propose an utility function that can reassemble the data, we turn next to the Stone-Geary form, proposed by Geary (1950) and used in similar context by Şirin Saraçoğlu (2008). In the case of two goods x_i and x_j, let $\gamma_i, \gamma_j, \lambda_i, \lambda_j \in \mathbb{R}$, the utility assumes the form:

$$U(x_i, x_j) = (x_i - \gamma_i)^{\lambda_i}(x_j - \gamma_j)^{\lambda_j}$$

where, according to Geary (1950), we can assume $\lambda_i + \lambda_j = 1$ without loss of generality. Denote x_i as the Informal good and x_j as the Formal goods; in our particular case, it is convenient to set $\gamma_j = 0$, while we must have $0 < \gamma_i < x_i$. We show in Appendix A.8 that these assumptions are enough to generate an income elasticity of demand of less than one (but larger than zero) for informal goods and more than one for formal goods, being thus a non-homothetic preference that can nest our empirically documented facts. The example illustrated in the bottom graph of Figure 3.1 shows that both the consumption of formal and informal goods increase, but the first increase proportionally more.

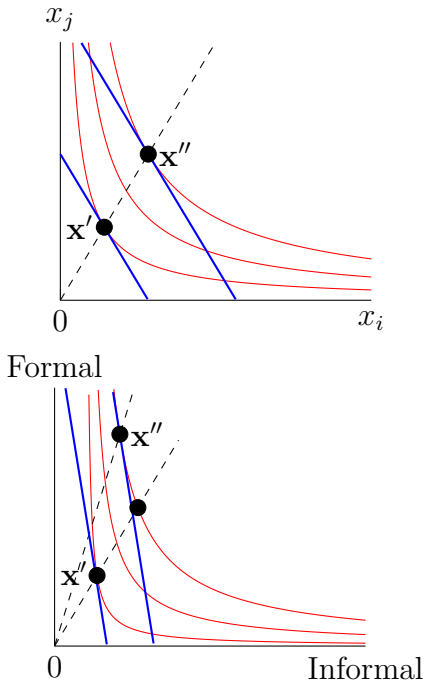

Figure 3.1: Example of a homothetic utility function (above) and of a Stone-Geary utility function (below).

4
Conclusion

This paper revisits the discussion around the positive correlation between development and the size of the formal sector. We have provided evidence favorable to a hypothesis that has been undocumented for the context: demand for formal goods increases proportionally more than demand for informal goods when income grows. This result implies that formal goods are luxury goods, while informal goods are necessity goods, and in turn preferences are non-homothetic. We argue that this effect propels formalization as countries develop.

Our empirical approach consisted of studying the formalization process that occurred in Brazil in the 2000s, which witnessed year-after-year drops in informality labor share for a total drop of 12 p.p. between 2001 and 2013, while the (geometric) average real growth rate in this period was 3.4%. We study three applications of the Brazilian household expenditure surveys (POF) in the years 2002-2003, 2008-2009, and 2017-2018 in order to measure how households' consumption choice varies with income, both in the cross-section and across surveys.

In Chapter 2, we find that our interest variable is highly correlated with the log of per capita income in the cross-section analysis of each survey, and regression coefficients are still large when we add controls for observable variables. We support the existence of consumption shifting from informal to formal goods during the 2000s in Brazil, but our results for the variation between 2008-09 to 2017-18 differ upon specifications. Computing *ceteris paribus* contributions of income growth on total consumption share in turn revealed to be difficult once expected income growth, as given by GDP, contrasts with observed growth. We conclude the analysis by evaluating the increase in the minimum wage as a candidate for an exogenous income shock. The propensity score matching estimator using the log of the ratio of incomes between matched households points to a significant role of the income increase in affecting positively the share of formal goods.

On the theoretical aspects, Chapter 3 proposes micro-founded arguments for a stronger preference over formal goods based on the literature, from which we highlight 3 key elements: regulation, the cost structure, and signaling. On

top of that, we also discuss the implications of our findings for the form of the utility function that better represents the demand shift from informal to formal goods as income grows, stating that the commonly used homothetic preferences cannot nest the observed effects.

Additional contributions to the paper can come in different forms. One of them is proposing new exogenous shocks for better estimation of the causal impact, allowing for comparison with our result. Also, further work may expand the study of how formal consumption changes over time for other countries and for more years. Another form is proposing a search-in-match model that includes non-homothetic preferences, using our numerical estimations as targets and parameters. This last exercise can be particularly useful for calculating the contribution of this new mechanism when interactions with traditional mechanisms are accounted for, detailing feedback effects, and further exploring the link between income shifts and formalization growth.

Bibliography

Almeida, R. and P. Carneiro (2012, July). Enforcement of labor regulation and informality. *American Economic Journal: Applied Economics 4*(3), 64–89.

Auriol, E. and M. Warlters (2005). Taxation base in developing countries. *Journal of Public Economics 89*(4), 625–646. Cornell - ISPE Conference on Public Finance and Development.

Bachas, P., L. Gadenne, and A. Jensen (2020, June). Informality, consumption taxes and redistribution. Working Paper 27429, National Bureau of Economic Research.

D'Erasmo, P. (2013). Access to credit and the size of the formal sector in brazil.

Fairris, D. and E. Jonasson (2020, 10). Determinants of changing informal employment in brazil, 2000-2010. *Journal of Development Economics 1*, 1–43.

Geary, R. C. (1950). A note on "a constant-utility index of the cost of living". *The Review of Economic Studies 18*(1), 65–66.

Gerard, F. and G. Gonzaga (2021, August). Informal labor and the efficiency cost of social programs: Evidence from unemployment insurance in brazil. *American Economic Journal: Economic Policy 13*(3), 167–206.

Haanwinckel, D. and R. R. Soares (2020). Workforce composition, productivity, and labor regulations in a compensating differentials theory of informality. Working Papers 2020-45, Becker Friedman Institute for Research In Economics.

La Porta, R. and A. Shleifer (2014, September). Informality and development. *Journal of Economic Perspectives 28*(3), 109–26.

Mas-Colell, A., M. D. Whinston, J. R. Green, et al. (1995). *Microeconomic theory*, Volume 1. Oxford university press New York.

Maurizio, R. and A. P. Monsalvo (2021, January). Informality, labour transitions, and the livelihoods of workers in latin america. *2021*.

Perry, G. E., W. Maloney, O. S. Arias, P. Fajnzylber, A. D. Mason, and J. Saavedra (2007). *Informality: Exit and Exclusion*. The World Bank Group.

Rocha, R., G. Ulyssea, and L. Rachter (2018). Do lower taxes reduce informality? evidence from brazil. *Journal of Development Economics 134*, 28–49.

Ulyssea, G. (2010). Regulation of entry, labor market institutions and the informal sector. *Journal of Development Economics 91*(1), 87–99.

Ulyssea, G. (2018, August). Firms, informality, and development: Theory and evidence from brazil. *American Economic Review 108*(8), 2015–47.

Varian, H. R. (1992). *Microeconomic Analysis* (3rd ed.). W. W. Norton & Company.

Şirin Saraçoğlu, D. (2008). The informal sector and tax on employment: A dynamic general equilibrium investigation. *Journal of Economic Dynamics and Control 32*, 529–549.

A
Appendix

A.1
Retail sector

Methodologically, Fairris and Jonasson (2020) estimate Linear Probability Models (LPM) for 2000 and 2010 and decompose informality rates change into changes in the means of explanatory variables and estimated regression coefficients using Blinder-Oaxaca decomposition[1]. One key feature of their exercise is that they account for the majority of traditional factors influencing informality in the literature, including: rising rates of law enforcement, rising educational levels, *Bolsa Família*[2], increases in minimum wage, tax payment simplification, availability of credit, trade liberalization, and turning out to be most important, industry composition.

Fairris and Jonasson (2020) call attention to industry composition as a less explored factor. According to their estimation, it is the changing way in which industry translates into informality (i.e., changes in the estimated coefficients) that accounts for the largest portion of the decline in informality. They divide the labor force into 16 industry categories, of which Retail Trade (21% of the total in 2010) and Real-estate services (9% of the total in 2010) were the sectors that grew more in the share of the labor force (2 p.p., page 18). Of all the categories, Retail-Trade is also the only with a significant coefficient (p-value $< 5\%$) in both periods, with a well-identified decrease in the way a worker occupied in this section is more likely to be informal (1.35 p.p. in 2000 vs 0.415 p.p in 2010, page 22).

According to their decomposition, these changes in Retail Trade alone would cause informality to decrease more than 3 times the observed value, but this is not the only sector with large effects: it is surpassed only by Manufacture, which would cause a ceteris paribus decrease in labor informality of roughly 4 times more. Other sectors such as Utilities and Real-estate services also would cause a decrease of

[1]Let IS_t be the probability of worker be employed in the informal sector in year t and let X_t be explanatory variables (here both work-level and municipal-level variables). Thus, LPM provides $\overline{IS}_t = \overline{X}_t\hat{\beta}_t$ where bar indicates mean values. The decomposition follows from:

$$\overline{IS}_1 - \overline{IS}_0 = \left[\left(\bar{X}_1 - \bar{X}_0\right)\hat{\beta}_0\right] + \left[\left(\hat{\beta}_1 - \hat{\beta}_0\right)\bar{X}_1\right]$$

[2]*Bolsa Família* was one of the world's largest and most successful cash transfer program, lasting from 2004 to 2021.

more than 2 times more; the other sectors typically have large effects but lesser than 1.5 times more.

We are motivated by their findings to investigate further the role of each sector in the formalization process. As was previously discussed in Chapter 3, consumers should derive greater utility from buying from formal firms due to signaling of good quality. If this particular effect is prominent enough, then retail market formalization should be leading in the economy, as it is the sector whose type matters first from this channel[3].

To empirically access the trends of the informality of each sector, we as Fairris and Jonasson (2020) use the census from 2000 and 2010, this time matching the sector definition of each year at the least comprehensive level. In this fashion, we identify 40 sectors (identified in the 1:100 interval). Comparing the formality rates between the two years (Figure A.1), all but 3 sectors became more formal[4].

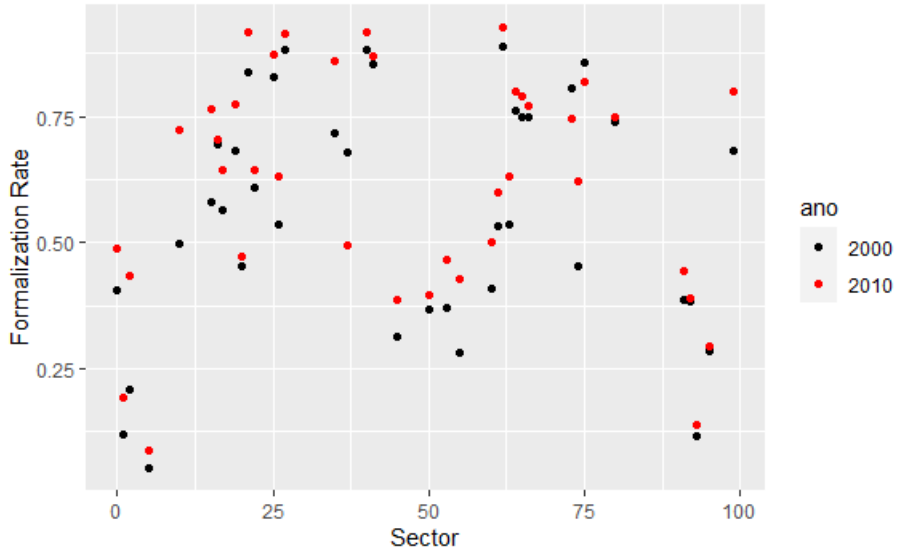

Figure A.1: Share of formal employment by sector.

It is difficult to access which sector contributed the most to the formalization process since the counterfactual is hard to estimate. The Retail Trade sector is sector 53, which wasn't the sector that had the greatest percentage points increase in the rate of formalization, ranking 9^{th} in this metric with an increase of 11 p.p.[5], but other factors lead us to conclude it was indeed key for the whole process, particularly due to its size. The Retail Trade sector represented 14% of all jobs in

[3]Particularly, the establishment type in Ulyssea (2018)'s extensive margin should matter most, because consumers have more difficulty assessing the formality status of the workers and its implication for the product quality.

[4]The sector 37 had a fall in the formality rate of 18 p.p.: that is the recycling sector, which expanded largely during the years compared.

[5]The sectors of extractive industries and forest exploration increased formalization rate as much as 25 p.p.

2000, the second-largest, and 16% in 2010[6], by then the largest of the considered sectors. When we weight percentage point growth by the share of employees, either by 2000 or 2010 ratios, this sector accounts for the largest position, suggesting it was the sector that contributed the most to the formalization process.

A.2
Consumption decomposition

In the table A.1 below, we present the categorization that yielded the formal/informal classification. To facilitate the classification, Bachas et al. (2020) focus on the subgroups proposed by the Brazilian Institute of Geography and Statistics (IBGE)[7] — which amount to 94, 165, and 162 subgroups for 2002-03, 2008-09 and 2017-18 surveys respectively — and assign them to 30 categories. We use the same method, essentially using the same classification as them when applicable, given the 2008-09 survey as the benchmark.

The divergences are as follows: first, we classify two subcategories not classified by the authors, categories "006" and "042", representing respectively small markets and health institutions. The second modification, reflecting the innovation of this work to account for the unspecified type of purchase, is that we proxy for formal and informal domestic services according to whether or not the household contributed to the worker's public pension, given that these expenditures were previously classified as unknown. The third modification was made necessary by the inclusion of the 2017-18 survey in the analysis. We realized the place of purchase denominated "Mercado" (portuguese for Market), originally accounted for as Grocery Store — thus informal —, grew largely in its share of total consumption, and from the Brazilian experience, those often refers to supermarkets and should be accounted as formal. This divergence is meaningless for the 2002-03 to 2008-09 comparison because this classification wasn't available for 2002-03, and in 2008-09 it corresponded to merely 0.02% of total consumption. The impact is focused on the 2017-18 survey, where it represented 2.6% of the total. The last and most discrepant change was that we disagreed with the authors in a specific subgroup, the subgroup "080", comprising "Escritórios e Administradoras em Geral" (portuguese for Offices and Management Entities), which was originally assigned as Private Services — and thus informal —, but given that services like advocacy, in particular, are provided by law offices and are formal, we decided to allocate this subgroup to Specialized Shop, that is formal. In particular, this subgroup accounted for 2.6% of consumption expenditures in

[6]This 2 p.p. increase makes section 53 the one that increased its labor share of the total the most, with a similar increase when we consider each sector's share of only formal employment. Within the period, the sector gained 3.2 million formal workers and only 1.2 informal workers.

[7]The institution that conducts the surveys in Brazil.

2017-18, 1.7% in 2008-09, and 2.1% in 2002-03.

Table A.1: Decomposition of expenditures by type of store (total and percent) for each consumption survey.

Category	Total 2017	% 2017	Total 2008	% 2008	Total 2002	% 2002	Classifcation
bank	1734.79	0.07%	13841.99	0.93%	12273.19	1.09%	0
bar-cafe	37442.52	1.51%	26590.43	1.79%	18420.90	1.64%	1
communication company	4091.76	0.16%	5615.095	0.38%	3095.93	0.28%	0
department store	13874.74	0.56%	9436.75	0.64%	5911.89	0.53%	0
education institution	26134.99	1.05%	36094.87	2.44%	27629.19	2.46%	0
fair	17184.02	0.69%	10766.14	0.73%	7638.53	0.68%	1
from farm	5663.39	0.23%	4438.949	0.30%	1239.39	0.11%	1
grocery store	45648.05	1.84%	18057.75	1.22%	10964.62	0.97%	1
health institution	48344.21	1.95%	18826.3	1.27%	40320.35	3.58%	0
hotel	565.80	0.02%	792.9531	0.05%	478.07	0.04%	0
internet	28011.65	1.13%	6040.042	0.41%	3151.52	0.28%	0
lottery	274.69	0.01%	6795.408	0.46%	4753.47	0.42%	0
market	534.68	0.02%	1151.839	0.08%	9692.35	0.86%	1
non-profit	2174.00	0.09%	1181.948	0.08%	2144.63	0.19%	0
own production	2975.73	0.12%	1852.838	0.13%	1138.61	0.10%	1
own production_other hh	6221.28	0.25%	4312.787	0.29%	5686.46	0.51%	1
pharmacy	107999.72	4.35%	47459.92	3.20%	48613.17	4.32%	0
private service	208391.71	8.40%	107006.2	7.22%	64722.78	5.75%	1
public administration	5638.54	0.23%	5128.942	0.35%	7952.41	0.71%	0
public health	19554.99	0.79%	9200.524	0.62%	16109.45	1.43%	0
real estate agent	180.25	0.01%	1545.051	0.10%	10630.78	0.94%	0
recreation events	12613.74	0.51%	6914.121	0.47%	5420.78	0.48%	1
restaurant	69857.71	2.82%	41276.54	2.79%	17475.85	1.55%	0
formal domestic services	1623.51	0.07%	973.165	0.07%	4576.20	0.41%	0
informal domestic services	4543.99	0.18%	1995.792	0.13%	10660.24	0.95%	1
small market	21592.83	0.87%	28257.15	1.91%	22890.56	2.03%	1
small shop	22660.44	0.91%	8829.494	0.60%	5981.95	0.53%	1
specialized shop	476375.05	19.20%	295256.1	19.93%	185431.16	16.48%	0
street seller	19779.42	0.80%	14348.29	0.97%	12039.10	1.07%	1
supermarket	283733.98	11.44%	139636.7	9.42%	77612.62	6.90%	0
unspecified	852732.37	34.38%	521094.3	35.17%	440736.45	39.17%	-
vehicle	132391.56	5.34%	86922.99	5.87%	39833.40	3.54%	0

Total expenditure values are given in millions of current *reais* (R$). 1 signs for informal and 0 for formal.

A.3
Unspecified type of purchase

In this section, we discuss the implications of ignoring consumption from unspecified sources to calculate the share of formal good consumption. Table A.2 makes clear that the share of consumption expenditure from unspecified places of purchase ("Share of Unspecified") increases with the log of per capita household income ("log of income"). This rise concerns that our analysis is biased by the exclusion of this component. To address the issue, in Table A.3, we present the multiple OLS regression for two extreme scenarios: including all unspecified as formal goods and including them all as informal goods.

Although in both cases the positive correlation persists, classifying all unspecified as informal is the most harmful for our hypothesis. We ease this concern by analysing the kind of goods whose place of purchase are unreported and present the results in table A.4 and A.5. We classify around 98% of purchases from unspecified places, but around 49% in the 2002-03 survey, 28% in the 2008-09 and 19% in the 2017-18 survey escape this new method by being too aggregate kind of goods (e.g. "alimentation") or real estate related[8]. We can see that a great share (more than half) is composed of payment of public goods and services, taxes, financial services, fuel, internet, telephone, and TV signatures, and formal transportation services (taxes and airplanes), thus should rather be deemed as formal.

The shares of formal consumption using the items as proxies are 76.9% in 2002-03, 78% in 2008-09, and 79.3% in 2017-18, again reflecting a circumstance wherein 2017-18 the share of formal consumption is actually higher. Notice that excluding non-identified places of purchase from the sample is equivalent to guessing that their composition in terms of formal and informal is equivalent to the identified sample, an approach that seems conservative compared to the analysis of the extremes of Table A.3.

[8]Different from Bachas et al. (2020), we don't exclude real estate payments in the identified local of purchase sample.

Table A.2: OLS regression between share of consumption from unspecified sources and log of per capita income.

	Dependent variable:		
	Share of Unspecified		
	[2017-18]	[2008-09]	[2002-03]
log of income	3.206***	4.549***	3.973***
	(0.114)	(0.116)	(0.129)
Constant	10.095***	0.277	12.441***
	(0.809)	(0.718)	(0.725)
Observations	58,039	56,091	48,568
R^2	0.034	0.075	0.061
Adjusted R^2	0.034	0.075	0.061
Residual Std. Error	572.631	527.725	542.883
F Statistic	2,059.967***	4,573.326***	3,129.202***
Note:			*p<0.1; **p<0.05; ***p<0.01

Table A.3: Analysis of extreme scenarios for classifying unspecified goods.

	Dependent variable:					
	Share of Formal (Unspecified = Formal)			Share of Formal (Unspecified = Informal)		
	[2017-18]	[2008-09]	[2002-03]	[2017-18]	[2008-09]	[2002-03]
log of income	1.172***	2.889***	2.756***	1.091***	0.843***	0.355***
	(0.099)	(0.105)	(0.104)	(0.112)	(0.120)	(0.118)
mean years of study	1.016***	1.096***	0.777***	0.340***	0.585***	0.802***
	(0.024)	(0.026)	(0.027)	(0.027)	(0.030)	(0.031)
mean age	0.119***	0.077***	0.078***	0.026***	0.011*	−0.009
	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)
paving	3.194***	3.698***	2.542***	1.185***	2.374***	1.509***
	(0.194)	(0.190)	(0.184)	(0.221)	(0.218)	(0.208)
sewage	2.577***	2.461***	2.524***	−0.137	0.895***	1.856***
	(0.189)	(0.191)	(0.195)	(0.215)	(0.219)	(0.221)
States FE	YES	YES	YES	YES	YES	YES
Constant	57.749***	45.723***	50.883***	34.111***	36.306***	33.527***
	(0.922)	(0.945)	(0.935)	(1.048)	(1.081)	(1.061)
Observations	58,038	55,318	46,508	58,038	55,318	46,508
R^2	0.226	0.294	0.285	0.042	0.076	0.097
Adjusted R^2	0.225	0.293	0.285	0.042	0.076	0.096
Residual Std. Error	567.183	539.521	500.856	644.698	617.510	568.334
F Statistic	545.986***	741.233***	597.968***	82.933***	146.776***	160.451***
Note:						*p<0.1; **p<0.05; ***p<0.01

Dependent and control variables are defined the same way as in Table 2.1.

Table A.4: Simple OLS regression (goods from unknown sources are classified by type of good).

	Dependent variable:		
	Share of Formal		
	[2017-18]	[2008-09]	[2002-03]
log of income	5.749***	7.286***	6.905***
	(0.112)	(0.117)	(0.117)
Constant	34.418***	25.779***	34.181***
	(0.827)	(0.760)	(0.694)
Observations	58,039	56,065	48,568
R^2	0.082	0.131	0.156
Adjusted R^2	0.082	0.131	0.156
Residual Std. Error	645.921	621.012	557.679
F Statistic	5,206.320***	8,469.714***	8,957.448***
Note:			*p<0.1; **p<0.05; ***p<0.01

Table A.5: Multiple OLS regression (goods from unknown sources are classified by type of good).

	Dependent variable:		
	Share of Formal (unspecified by kind of good)		
	[2017-18]	[2008-09]	[2002-03]
log of income	1.215***	2.215***	2.383***
	(0.147)	(0.168)	(0.189)
household size	1.196***	1.101***	0.496***
	(0.083)	(0.079)	(0.080)
mean years of study	1.035***	1.138***	0.767***
	(0.036)	(0.040)	(0.046)
mean age	0.182***	0.128***	0.098***
	(0.008)	(0.010)	(0.011)
paving	3.291***	4.182***	2.804***
	(0.267)	(0.276)	(0.283)
sewage	2.586***	2.657***	2.726***
	(0.254)	(0.282)	(0.311)
States Dummies	YES	YES	YES
Constant	47.574***	38.907***	47.657***
	(1.135)	(1.249)	(1.205)
Observations	58,038	55,292	46,508
R^2	0.223	0.274	0.281
Adjusted R^2	0.222	0.273	0.281
Residual Std. Error	594.524 (df = 58005)	567.512 (df = 55259)	516.788 (df = 46475)
F Statistic	519.894*** (df = 32; 58005)	651.043*** (df = 32; 55259)	568.468*** (df = 32; 46475)

Note: *p<0.1; **p<0.05; ***p<0.01

Dependent and control variables are defined the same way as in 2.1.

A.4
Food multiple regression

Table A.6 presents the same exercise as Table 2.1, but looking only to expenditures on food goods. The dependent variable is the share of expenditures on food goods that are formal. Independent variables are the log of household income ("log of income") as reported by households, the number of household members ("household size"), mean years of study (for household members with more than 16 years old, the bottom of the working-age population), the mean age of household members, paving and sewage.

Table A.6: Multiple OLS regression for food goods expenditures only.

	Dependent variable:		
	ShareFormal		
	(1)	(2)	(3)
log of income	1.287***	3.062***	2.754***
	(0.294)	(0.318)	(0.383)
idade_m	0.092***	0.069***	0.013
	(0.017)	(0.018)	(0.023)
household size	1.146***	1.020***	0.321*
	(0.166)	(0.143)	(0.165)
mean age	1.202***	1.399***	1.512***
	(0.072)	(0.078)	(0.097)
paving	5.822***	6.830***	5.666***
	(0.526)	(0.513)	(0.613)
sewage	4.854***	5.472***	6.184***
	(0.502)	(0.549)	(0.673)
State Dummies	YES	YES	YES
Constant	38.363***	16.446***	11.613***
	(2.594)	(2.279)	(2.356)
Observations	52,672	51,967	44,716
R^2	0.143	0.193	0.199
Adjusted R^2	0.142	0.193	0.199
Residual Std. Error	1,139.200 (df = 52639)	1,017.424 (df = 51934)	1,019.379 (df = 44683)
F Statistic	273.891*** (df = 32; 52639)	388.296*** (df = 32; 51934)	347.528*** (df = 32; 44683)

Note: *p<0.1; **p<0.05; ***p<0.01

Dependent and control variables are defined the same way as in 2.1.

A.5
Types of goods composition

In Table A.7, we extend the summary classification of types of goods in Bachas et al. (2020) for the surveys of 2002-03 and 2017-18. For the 2002-03 years, the product list from IBGE had 10,429 goods cataloged; this number was 13,778 in 2008-09 and 13,474 in 2017-18. The original classification included 70 types of goods, but differently from the authors, we excluded all the consumption from non-specified types of purchase, which cleaned 20 types.

Table A.7: Analysis by goods' classification

name_category	% Total 2017	% Total 2008	% Total 2002	% Formal 2017	% Formal 2008	% Formal 2002	Coef Multi 2017	Coef Multi 2008	Coef Multi 2002
TV and internet	0.11%	0.00%	0.00%	98.38%	0.00%	0.00%	-	-	-
Home Maintenance	2.59%	4.72%	4.80%	55.13%	51.77%	45.11%	-3.25***	-3.66***	-1.99***
Household goods repair	2.82%	0.46%	0.44%	95.63%	70.43%	69.92%	1.9***	1.53	1.77
Taxes	0.00%	0.38%	1.67%	0.00%	99.60%	98.90%	-	0.2	1.32***
Real Estate (maintenance)	3.46%	3.36%	2.70%	55.67%	65.75%	67.98%	-2.06**	-2.3***	-0.41
Others	3.16%	1.84%	2.00%	78.40%	82.05%	84.63%	2.17***	2.11***	0.61
Real Estate (acquisition)	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	-	-	-
Appliances	2.87%	3.98%	3.03%	82.57%	85.21%	75.73%	7.19***	4.75***	2.91***
Furniture and household goods	3.07%	3.38%	2.68%	76.89%	83.27%	81.72%	6.65***	4.72***	5.76***
Labor contributions	0.00%	0.00%	0.02%	0.00%	0.00%	99.86%	-	-	-
Alimentation	28.51%	28.92%	24.16%	69.61%	66.92%	59.22%	1.31***	3.06***	2.75***
Smoking	0.82%	0.85%	0.83%	52.66%	40.93%	34.04%	7.25***	6.91***	4.17***
Games and gambling	0.00%	0.42%	0.35%	0.00%	87.00%	86.73%	-	0.25	2.12***
Newspapers Magazines and Books	0.28%	0.51%	0.47%	91.15%	94.30%	93.27%	1.87***	2.38***	1.31*
Recreation and sports	0.98%	0.83%	0.80%	21.09%	12.28%	6.05%	0.54	0.1	0.65
Medication	6.43%	5.45%	7.04%	97.09%	98.11%	98.66%	0.72***	0.64***	0.53***
Treatment material	0.62%	0.44%	1.39%	95.22%	97.68%	95.45%	1.55***	1.63***	2.17***
Products for personal use	4.16%	1.70%	2.31%	82.50%	79.93%	78.18%	2***	2.19***	4.52***
School articles	0.38%	0.39%	0.35%	93.24%	91.18%	89.93%	2.28***	2.51***	2.55***
Men's clothing	1.83%	2.09%	1.70%	84.82%	85.58%	86.16%	4.21***	5.2***	6.12***
Women's clothing	2.30%	2.61%	1.86%	84.62%	84.18%	83.60%	4.38***	4.87***	6.14***
Children's clothing	1.27%	0.99%	0.84%	79.57%	81.46%	82.89%	4.31***	3.85***	4.67***
Fabrics	0.09%	0.10%	0.13%	91.96%	92.91%	88.57%	2.45***	2.55**	3.45***
Shoewear and similar	2.03%	2.44%	1.69%	86.61%	86.28%	89.71%	4.3***	4.27***	5.8***
Health Insurance	4.52%	3.33%	3.80%	99.15%	99.65%	96.19%	0.26	0.39*	1.73***
Medical appointment	1.52%	0.43%	2.30%	98.12%	99.27%	97.95%	-0.5**	-0.24	-0.47**
Varied examinations	0.46%	0.33%	0.80%	99.77%	99.80%	98.28%	0.07	0.02	-0.1
Medical treatment	0.04%	0.21%	1.21%	92.19%	94.20%	92.92%	-5.74	0.01	-0.66
Dental treatment	0.69%	0.53%	2.20%	99.53%	99.46%	96.06%	-0.89	0.22	0.43
Maintenance and accessories	3.67%	3.14%	2.21%	96.94%	97.68%	96.32%	3.12***	0.74	0.85***
Jewelry and Bijou	0.31%	0.40%	0.36%	72.78%	69.43%	63.05%	6.74***	6.63***	6.55***
Celphones and accessories	0.78%	0.56%	1.22%	75.96%	80.88%	93.85%	5.92***	6.61***	2.78***
Other courses and activities	0.33%	1.13%	0.96%	95.45%	95.11%	94.74%	-1.61***	-0.06	0.16
Textbooks and journals	0.27%	0.20%	0.22%	96.69%	95.28%	95.58%	0.06	0.93	2.66***
Regular courses	0.28%	1.23%	1.44%	99.71%	99.92%	99.63%	-0.02	0.25**	0.23
Vehicle acquisition	12.23%	13.24%	8.51%	63.30%	71.51%	67.38%	9.1***	9.85***	5.85***
Cleaning	0.97%	1.09%	0.95%	80.82%	77.91%	70.63%	3.67***	5.5***	5.74***
Mortgage Payment	0.00%	0.12%	1.30%	0.00%	84.64%	85.81%	-	0.14	-2.98*
Monetary rent	0.00%	0.34%	2.24%	0.00%	37.28%	34.90%	-	6.99***	6.61***
Condominium	0.00%	0.14%	1.00%	0.00%	97.86%	86.14%	-	0.7	2.78*
Mobile phone	0.22%	1.91%	0.91%	95.03%	94.30%	97.43%	2.59***	2.31***	0.86**
Domestic gas	1.64%	1.41%	1.55%	13.32%	12.02%	20.48%	-0.86***	-0.39	-1.25**
Water and sewage	0.00%	0.00%	0.07%	0.00%	0.00%	36.56%	-	-	7.18***
Perfume	1.78%	1.50%	0.93%	53.91%	67.20%	73.68%	3.56***	2.51***	5.1***
Hair products	0.84%	0.32%	0.27%	86.36%	85.41%	81.81%	2.39***	2.68***	3.46***
Soap	0.44%	0.20%	0.18%	80.10%	83.46%	76.06%	2.77***	5.7***	5.72***
Surgery	0.50%	0.31%	1.60%	99.49%	98.53%	95.98%	-1.38	1.29	-0.02
Hospitalization	0.06%	0.07%	0.53%	100.00%	99.92%	100.00%	-	0.65	-
Tertiary education courses	0.22%	1.49%	1.62%	99.88%	99.90%	99.53%	0.15*	-0.31*	0.36*
Toys and Games	0.45%	0.47%	0.39%	84.42%	86.85%	87.04%	4.9***	3.9***	3.64***

The first 3 numeric columns stands for the share of total expenditures to which the kind of good corresponds. The next 3 are for the formal share of consumption of each type good. The last columns show the coefficients for a multiple OLS regression with controls identical to Table 2.1.

A.6
Bolsa Família (BF) analsis

We argue that the cash transfer program *Bolsa Família* (BF) in theory represents an opportunity to explore another exogenous income shock, but is **unfeasible** in practice. This program would provide direct transfer for poor families conditional on having children or teenagers, and to extremely poor families unconditionally. It was officially created on January 9th, 2004, post our first analysed POF survey. In theory, it presents a well-settled cut-off determining poor and extremely poor families (households who earn less than R$120 were registered as poor, and those who earn less than R$60 were registered as extremely poor).

The idea was therefore to explore this cut-off. In principle, there are two ways to estimate causal effects in such an environment. First, we could compare the treated group (those below the poverty line) to the group that is slightly above the cut-off in an RDD setting. Second, we could define the group slightly below the cut-off in both surveys as the treatment group (since this is the group that should earn BF benefits in the 2008-09 survey) and the group slightly above as the control group, and study the difference of the difference (DID estimation) between the groups in 2008-09 compared to 2002-03. In both settings, the idea is that poor families will receive an income shock that families above the cut-off will not, so the share of formal consumption should increase in the former compared to the latter[9]. In the RDD, those with income (excluding BF benefits) slightly below R$120, for instance, would have a de facto total income higher than those with income slightly above R$120, for the former will earn BF benefits, and the latter will not. In the DID estimation, for the same reasoning, the treatment group should witness an increase in the share of formal consumption greater than the control group, so the gap between the shares should decrease and maybe even become negative.

Despite the formal income level requirement, Figure A.2 makes clear that many families who receive *Bolsa Família* income are above the cut-off. This scenario indicates low enforcement and plausibly arises from the fact that households with informal labor will tend not to report their true income in order to receive the benefit. Since such work is not registered, it is harder to supervise, but its income is observed by POF because it does not affect benefit eligibility. In fact, if we try to

[9]We acknowledge it is not for granted that the relation between formal consumption share and income holds for the poor. For instance, as discussed in Chapter 3, if households' preference over formal/informal goods follows a quasi-linear utility, then families would consume informal goods up to a level of expenditures, and only after a threshold would they start buying formal goods. To address this concern, we re-estimate the regression using only the first decile of income (not shown), for which we find the correlation holds as strong as in the whole sample. Nevertheless, we still observe that this relation does not hold for the extremely poor, whose income lay below the 3rd percentile, as will be suggested below based on 2.1

calculate the benefit through the income and family composition rules (Table A.8), we find these factors can explain (according to R2) no more than 32% of received income, and the level of the benefit differs significantly from the rule. For instance, being an extremely poor family beneficiary of *Bolsa Família* endows the household with an R$58 benefit, whereas on average those families receive less than half of that value (R$19.42, see coefficient *extreme poverty*).[10] Figure A.3 displays the distribution of the benefits, showing that, instead of a pattern of clusters around the legally proposed transfer amounts, the households are reporting on a continuous support with the mode around R$100, suggesting they might be approximating.

Figure A.2: Histogram of overall income (green) and *Bolsa Família* beneficiaries pre-benefit income (red).

As Table A.8 makes clear, we do not know which characteristics determine entrance in *Bolsa Família*. This prevents us from creating a robust control in a parametric regression, so just adding a dummy for *Bolsa Família* beneficiaries would be correlating with important omitted variables and likely exacerbating the low income of the beneficiaries cohort. In other words, we cannot, by adding controls, make sure that the only difference in a group with a certain level of variables that received the transfer in 2008-09 and the equivalent group in 2002-03 that did not receive the transfer is the transfer. In the next subsection, we present the exercise comparing the group that in theory should be made up of the only eligible candidates in 2008-09 with this group in 2002-03. However, these criteria come far from previewing on large scale the accession to the program, so the imprecise estimation was not unexpected.

[10]The other coefficients refer to being poor and having young children up to 15 years (*child poverty*) and being poor and having teenagers between 16 to 17 years (*teen poverty*), which should yield respectively R$18 and R$30.

Table A.8: Decomposition of *Bolsa Família* income per criteria.

	Dependent variable:
	BF transfer
extreme poverty	19.422***
	(2.090)
child poverty	26.380***
	(0.411)
teen poverty	27.923***
	(2.498)
Observations	9,268
R^2	0.317
Adjusted R^2	0.317
Residual Std. Error	2,188.591 (df = 9265)
F Statistic	1,432.734*** (df = 3; 9265)
Note:	*p<0.1; **p<0.05; ***p<0.01

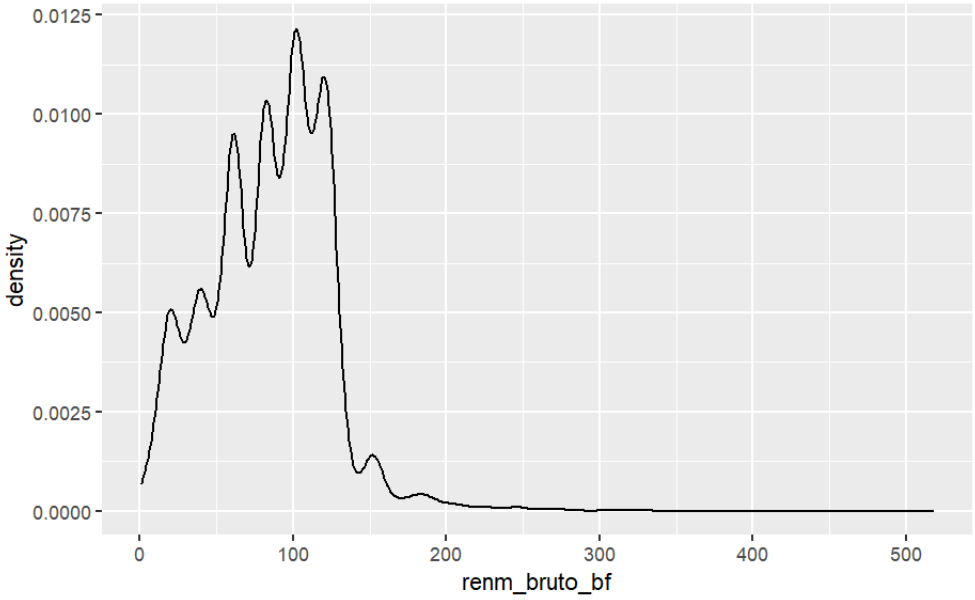

Figure A.3: Distribution of total benefits from *Bolsa Família* received by households.

A.6.1
BF parametric analysis

Table A.9 projects the impact of multiples variables on the percentage share of formal consumption. The focus is the interaction between being a eligible candidate for the transfer, according to the rule ("child poverty" > 0, "teen poverty" > 0 or "extreme poverty" > 0), and the dummy that indicate the later sample ("year 2008" for years 2008-09). If the participation rule were strictly enforced (only candidates could be treated) and all the candidates received the benefits (both conditions that Figure A.2 shows don't hold), then we expected that those characteristics would imply candidates to receive a greater income in 2008-09 when the program is active than the equivalent candidates in 2002-03, since the program hadn't been created yet. The paper predicts that greater income is to be associated with a smaller share of informal consumption.

We choose this approach because in this setting we can control parametrically for all characteristics that define the "treated group", that is those who fall within the criterion for receiving BF benefits and are thus more likely to receive them; for instance, to have "child poverty" > 0 you need to have kids between 0 and 15 years (that are controlled for using "child 0 to 15") and to be poor, which in turn we control by the log of the per capita income deflated to 2008 (log of income as 2008), and in order for "child poverty" > 0 to translate into greater chances to the program, the household must be registered in the 2008-09 survey, thus the interaction term "year 2008" — in turn, to measure the effect of this greater eligibility, we also add the term without interaction to represent the counterpart in the 2002-03 sample.

Note again we can't do the same exercise with the *de facto* beneficiaries because we don't know the characteristics that define this group as opposed to non-treated. For instance, if we designed a dummy for those who earn benefits from the program, we would be also signaling unobservable variables that relate to entering the program as opposed to not entering, which could include proximity to a bank agency, internet access, overall living conditions, city[11], inspection by the prefecture, among others.

Since Table A.8 implies low explanation power of the formal rule and the *de facto* beneficiaries of the program, it is not surprising that the impact is badly estimated. To begin with, our reasoning seems to hold for those who are poor and have kids ((year 2008)*(child poverty)), but we can't conclude that for poor families with teenagers ((year 2008)*(teen poverty)) and, for further surprise, the opposite seems to be true for the extremely poor ((year 2008)*(extreme poverty)), although only within the 10% confidence range. However, when we analyse the percentage of formal consumption by the log of per capita household income as in Figure 2.1, we do see a hook in the very first percentiles, that comprise the extremely poor, in that they seem to buy more informal goods out of a marginal income increase, thus that seems to be a phenomenon specific of that group rather than an opposition to the theory. This graphic is also illustrative in the sense that we can see that the red dots and the blue dots usually overlap, suggesting that, conditional on income, other variables do not strongly affect the share of formal consumption; for the green dots, on the other hand, there does seem to be an exogenous variable that tends to put the share below the other dots for the majority of high percentiles.

[11]Remember that to the number of households is no larger than 52,639, thus being far from having representativeness at the city level.

Table A.9: Impact of *BF* on the target population.

	Dependent variable:
	Share of Formal
log of income as 2008	1.571^{***}
	(0.207)
year 2008	-1.345
	(1.433)
(log income as 2008)*(year 2008)	0.327
	(0.220)
(year 2008)*(child poverty)	0.730^{***}
	(0.266)
(year 2008)*(teen poverty)	-1.381
	(1.014)
(year 2008)*(extreme poverty)	-1.860^{*}
	(1.042)
credit card	4.291^{***}
	(0.244)
sewage	2.846^{***}
	(0.267)
paving	3.037^{***}
	(0.249)
mean years of study	1.206^{***}
	(0.039)
mean age	0.139^{***}
	(0.010)
child 0 to 15	1.594^{***}
	(0.128)
pregnant	0.010
	(0.363)
teens 16 or 17	1.573^{***}
	(0.287)
child poverty	-1.535^{***}
	(0.213)
teen poverty	-2.119^{***}
	(0.779)
extreme poverty	-5.381^{***}
	(0.776)
States Dummies	YES
Constant	38.512^{***}
	(1.364)
Observations	101,644
R^2	0.273
Adjusted R^2	0.273
Residual Std. Error	663.216 (df = 101600)
F Statistic	886.951^{***} (df = 43; 101600)
Note:	*p<0.1; **p<0.05; ***p<0.01

A.7
Distributions histograms and income growth

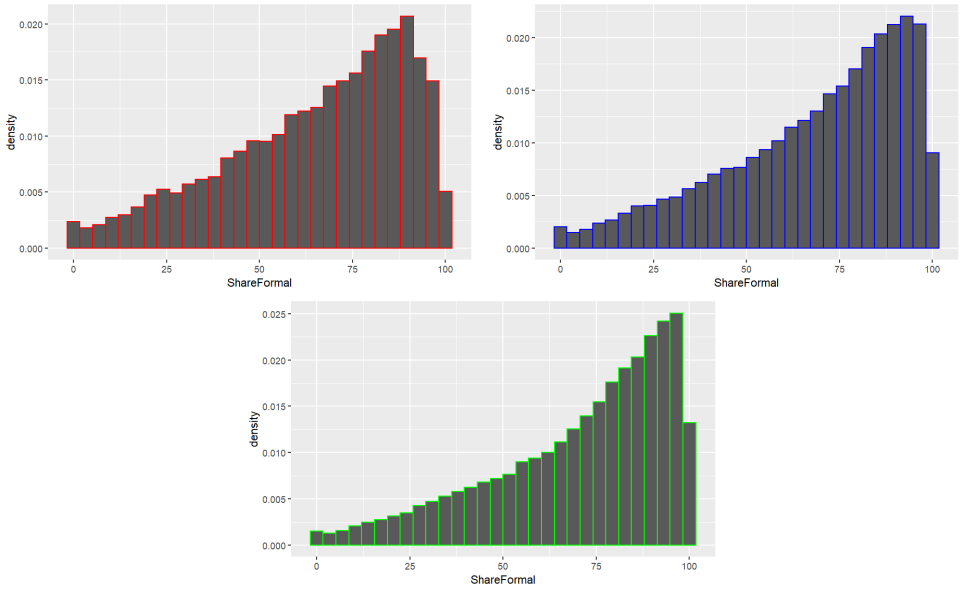

Figure A.4: Histograms with the share of formal goods consumption over total (percentage) by each surveys respectively: 2002-03 (red), 2008-09 (blue) and 2017-18 (green).

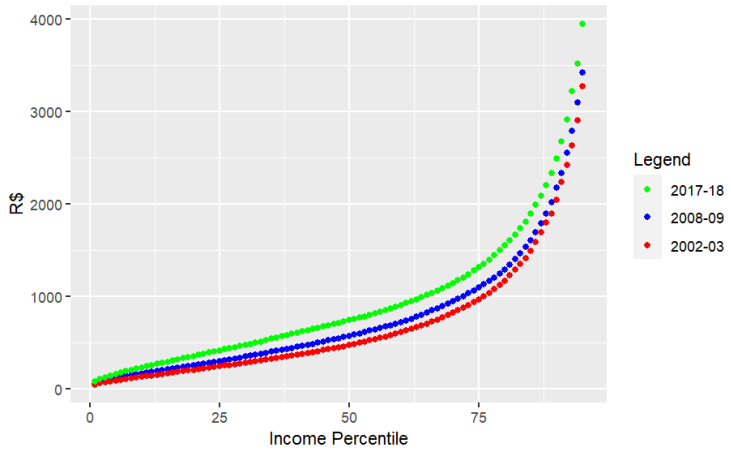

Figure A.5: Per-capita real household Income by percentile, cut in the 95^{th} percentile for better visualization

A.8
Mathematical appendix

A.8.1
Homothetic preferences

As defined by Mas-Colell et al. (1995), homothetic preferences are characterized by the indifference sets being related by "proportional expansion along rays; that is, if $\mathbf{x} \sim \mathbf{y}$, then $\alpha\mathbf{x} \sim \alpha\mathbf{y}$ for any $\alpha \geq 0$" (p. 45). As corollary, "a continuous

preference is homothetic if and only if it admits a utility function $u(\mathbf{x})$ that is homogeneous of degree one" (p. 50). Well, if $u(\mathbf{x})$ has this property, it follows that the derivative, $u'(\mathbf{x})$, is homogeneous of degree zero. This means that two points $\mathbf{x}$ and $\alpha\mathbf{x}$ have the same derivative with respect to some x_i (see Varian (1992) p.496 for proof). Then, for a price vector $\mathbf{p}$, the utility maximization condition implies, let MRS_{ij} be the marginal rate of substitution between goods i and j, that $MRS_{ij} = p_i/p_j$. Since the

$$MRS_{ij} = \frac{\partial u(\mathbf{x})/\partial x_i}{\partial u(\mathbf{x})/\partial x_j}$$

and we argued that $u(\mathbf{x})/\partial x_s = u(\alpha\mathbf{x})/\partial x_s$ for all $x_s \in \mathbf{x}$ and $\alpha \geq 0$, then if $\mathbf{x}^*$ is an optimal point given $\mathbf{p}$ for the income $\bar{m}$, so is $\alpha\mathbf{x}^*$ for some income $\hat{m}$. In fact, by the budget constraint:

$$\mathbf{p}.\mathbf{x}^* = \bar{m} \Leftrightarrow \mathbf{p}.\alpha\mathbf{x}^* = \alpha\bar{m}$$

as for Walras' law, $\alpha\mathbf{x}^*$ being optimal implies $\mathbf{p}.\alpha\mathbf{x}^* = \hat{m}$, it follows $\alpha\bar{m} = \hat{m}$, which means **that when income is multiplied by α, the optimality conditions implies that the consumer will simples increase the amount of each good by α**. In other words, Engels' curve is going to be linear.

Using the criterion above, it is easy to show that CES utility functions represent homothetic preferences. This class of functions has the form:

$$\begin{aligned} u(x_i, x_j) &= \left(w_i x_i^\rho + w_j x_j^\rho\right)^{1/\rho},\ w_i, w_j \in \mathbb{R} \\ \Rightarrow u(\alpha x_i, \alpha x_j) &= \left(w_i \alpha^\rho x_i^\rho + w_j \alpha^\rho x_j^\rho\right)^{1/\rho} \\ &= [(\alpha^\rho)(w_i \alpha^\rho x_i^\rho + w_j \alpha^\rho x_j^\rho)]^{1/\rho} \\ &= \alpha[w_i \alpha^\rho x_i^\rho + w_j \alpha^\rho x_j^\rho]^{1/\rho} = \alpha u(x_i, x_j) \end{aligned}$$

A.8.2
Non-homothetic prefences

We begin by showing that quasi-linear functions cannot be used to incorporate the theory and empirical facts of this thesis. Take the form of these functions as, let x_i be the Informal good, x_j the Formal good and $f(x)$ a strictly concave function:

$$u(x_i, x_j) = f(x_i) + x_j$$

$$\Rightarrow MRS_{ij} = \frac{\partial u(\mathbf{x})/\partial x_i}{\partial u(\mathbf{x})/\partial x_j} = f'(x_i) = \frac{p_i}{p_j} \Rightarrow x_i = f'^{-1}\left(\frac{p_j}{p_i}\right)$$

In this case, upon the condition that the agent has enough income to buy the optimal level, $x_i^* = f'^{-1}(p_j/p_i)$ is fixed as it doesn't depend on income, and therefore, again by Walras' Law (let m denote the income), $x_j^* = \frac{m}{p_j} - \frac{f'^{-1}(p_j/p_i)}{p_j}$, so x_j^* grows with income, but x_i^* doesn't. It is obvious at Figure A.6 that the optimal basket when income increases is not on the same linear array as before.

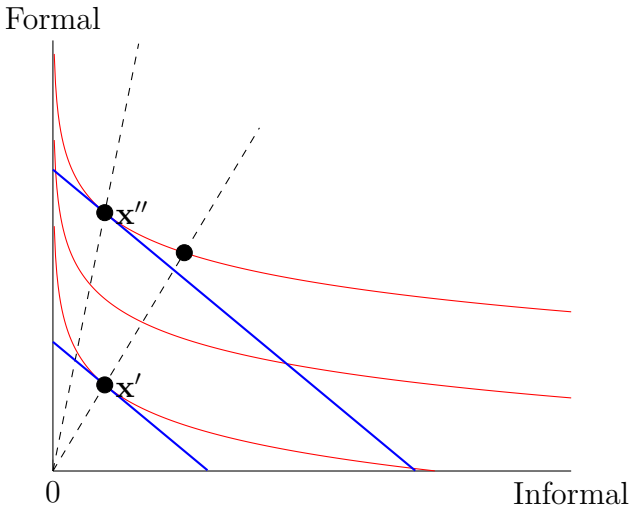

Figure A.6: Example of non-homothetic utility function.

Although this case is a common example in the literature, it does not nest the empirical observation discussed in Chapter 2. This is because this utility implies that: a) for income too low such that the optimal level of informal consumption can't be bought, households will spend all their income on informal goods and b) for incomes that are big enough, any marginal income above the necessary to consume the optimal level of informal goods will be spent in formal goods, meaning that the income elasticity of demand for informal goods should be zero (we can see that in Figure A.6 the amount of "Informal" goods consumption, i.e. the x-axis, is the same in both x' and x'', even though income increased, whereas "Formal" goods increased by a lot). In our empirical section (Chapter 2), we observe otherwise, as even those in the first percentile of income consume a relevant share of formal goods (see Figure 2.2) and the income elasticity for informal goods expenditures was close to 0.65 in all three surveys evaluated, meaning that when income increase a positive share of this marginal income will be spent in informal goods.

Next, we will show that — denoting x_i as the Informal good and x_j as the Formal good — and assuming that $\lambda_i + \lambda_j = 1$, $\gamma_j = 0$, and $0 < \gamma_i < x_i$, then the Stone-Geary preference can generate an income elasticity of demand of less than one for informal goods and more than one for formal goods, being thus a non-homothetic preference:

$$MRS_{ij} = \frac{p_i}{p_j} \Rightarrow \frac{\lambda_i x_j}{\lambda_j(x_i - \gamma_i)} = \frac{p_i}{p_j} \Rightarrow x_j = \frac{p_i}{p_j}\frac{\lambda_j}{\lambda_i}(x_i - \gamma_i)$$

plugging in the budget constraint thus yields:

$$x_i = \frac{\lambda_i}{p_i}m + \lambda_j\gamma_i$$
$$x_j = \frac{\lambda_j}{p_j}(m - p_i\gamma_i)$$

finally, the income elasticities are:

$$e_m^s \equiv \frac{\partial x_s}{\partial m}\frac{m}{x_s},\ s \in \{i, j\}$$
$$e_m^i = \frac{\frac{\lambda_i}{p_i}m}{\frac{\lambda_i}{p_i}m + \lambda_j\gamma_i} = \frac{1}{1 + \frac{\lambda p_i\gamma_i}{\lambda_i m}} < 1$$
$$e_m^j = \frac{\frac{\lambda_j}{p_j}m}{\frac{\lambda_j}{p_j}(m - p_i\gamma_i)} = \frac{1}{1 - \frac{p_i\gamma_i}{m}} > 1$$

where the last inequality is granted by the assumption that $\gamma_i < x_i$ thus $p_i\gamma_i < p_i x_i \leq m$ by the budget constraint, so $\frac{p_i\gamma_i}{m} < 1$. Note also that $e_m^i > 0\ \forall\ \lambda_i > 0$.