THE ECONOMICS OF TOBACCO CONTROL IN SOME AFRICAN COUNTRIES

GRIEVE CHELWA

Thesis Presented for the Degree of

DOCTOR OF PHILOSOPHY

in the School of Economics

UNIVERSITY OF CAPE TOWN

August 2015

<u>SUPERVISOR</u>:

PROF. CORNÉ VAN WALBEEK

SCHOOL OF ECONOMICS

UNIVERSITY OF CAPE TOWN

THE ECONOMICS OF TOBACCO CONTROL IN SOME AFRICAN COUNTRIES

ABSTRACT

This thesis examines some aspects of the economics of tobacco control in South Africa, Uganda and Zambia. The first part of the thesis examines whether tobacco expenditure displaces (or "crowds out") expenditure on other goods and services within Zambian households. In so doing, I make two contributions to the literature. Firstly, I use expenditure data from a low-income sub-Saharan African country where most households are poor. Secondly, I use the standard instrumental variable used in the literature, the adult sex ratio, to instrument for the tobacco smoking status of Zambian households. But unlike previous studies, I relax the strict exclusion restriction and allow for the adult sex ratio to be correlated with the error term. That is, I allow the instrumental variable to be imperfect. I consider the relaxation of the exclusion restriction to be reasonable given that the adult sex ratio is just as likely to influence tobacco expenditure as it is to influence expenditure on other goods and services. Even after relaxing the exclusion restriction, I, however, confirm many findings in the literature. For instance, I find that smoking households allocate less expenditure towards food, schooling, clothing, water, electricity, transportation, equipment maintenance and remittances. In addition, the crowding out patterns I uncover are in some ways related to the geographical location of households which in turn is related to socioeconomic status in Zambia. In sum, the results in this part of the thesis show that a broader accounting of tobacco's costs in Zambia should include other costs over and above mortality and morbidity considerations.

We know from several studies that tax and price measures are the single most effective policy tool for reducing tobacco consumption. However, most of this evidence is based on studies conducted in developed countries with very few published studies on African countries. The second part of my thesis, therefore, contributes to the recent literature that uses expenditure data to estimate price and expenditure elasticities of demand for tobacco products in Low- and Middle-Income countries. I use expenditure data from Uganda and exploit the fact that prices of cigarettes vary across geographical space. I also adjust my demand elasticity estimates for measurement error and quality heterogeneity. I find price and expenditure elasticities that are in line with international evidence. For instance, I find that cigarette demand is expected to decline by between 3% and 4%, at the very least, for every 10%

increase in cigarette prices. The authorities in Uganda can, therefore, reduce cigarette consumption by increasing excise taxes on cigarettes without reducing tax revenues.

The third and final part of my thesis evaluates the impact on per capita cigarette consumption of South Africa's consistent excise tax increases that began in 1994. The tax rises have overtime translated into large increases in the inflation-adjusted price of cigarettes. For instance, the average real price per pack increased by 110% between 1994 and 2004. The main challenge in conducting policy evaluations is that of creating a credible counterfactual. That is, we want to know what would have happened to per capita cigarette consumption in South Africa if the excise tax increases had not occurred. This is particularly important in the case of South Africa because per capita cigarette consumption had already started declining by the time the tax rises started. I, therefore, use a transparent and data-driven technique, the Synthetic Control method, to create a credible counterfactual of South Africa's cigarette consumption after 1994. The counterfactual is constructed as a linear combination of the per capita cigarette consumption of countries that are similar to South Africa but did not engage in large-scale tobacco control efforts over the period 1994 to 2004. I find that per capita cigarette consumption would not have continued declining in the absence of the consistent tax rises that began in 1994. Specifically, I find that by 2004, per capita cigarette consumption was 36% lower than it would have been had the tax increases not occurred. This result is robust to several falsification (or placebo) exercises. Based on these results, I conclude that countries in Africa can achieve substantial reductions in cigarette consumption and prevent uptake from new smokers by consistently increasing excise taxes in the manner of South Africa.

ACKNOWLEDGEMENTS

I would like to begin by thanking my supervisor Prof. Corné van Walbeek for the wonderful guidance he has given me since I began working on this thesis in April 2012. He has been very supportive of all my research ideas, always encouraging me to “go and give it a try”. He has fashioned me into the academic I always wanted to be. Above all else, he has been my father away from home.

Secondly, I would like to thank my colleagues in the Economics of Tobacco Control Project at the University of Cape Town. Prof. Hana Ross and Dr. Evan Blecher were key in pushing for the initial funding from which I and many other students have benefited. I would also like to thank Nicole Vellios, Jodie Polsen and Lara Kruger for the constructive feedback during the time I was preparing to defend my thesis proposal in January 2013. Sharon de Bruyns always made sure that my scholarship was paid in time for me to meet my bills. I am also reminded of the wider staff in the School of Economics including Prof. Edwin Muchapondwa, Paula Bassingthwaighte, Yasmina Charles and Brenda Adams.

I would like to thank the following people at the American Cancer Society who in one way or the other have contributed to this thesis: Dr. Jeffrey Drope, Jacqui Drope, Alex Liber and Michal Stoklosa. A special mention here for Dr. Drope and Alex Liber. Dr. Drope sent me one of my first reading lists on the economics of tobacco control in March of 2012. Alex Liber always and patiently responded to my email requests on the *Tobacco Atlas*.

The following people have been key in shaping the direction of this thesis: Prof. Emmanuel Guindon for his seminar presentation at the Emerging Researcher Programme in 2012 that got me interested in Angus Deaton’s methods on the use of expenditure surveys to estimate price elasticities of demand. He followed this up by sending me, in pdf format, the entire literature on the subject. Prof. Rijo John for tolerating my email queries on the methods used in Chapters 2 and 3. Dr. Miquel Pellicer for sending me Alberto Abadie’s paper on Synthetic Control methods which eventually led to my writing Chapter 4. Dr. Evan Blecher and Prof. Corné Van Walbeek’s work on cigarette affordability was incredibly crucial in my writing of Chapter 4. Writing this chapter would have been a thousand times more difficult if their work on affordability did not exist. Prof. Steven Koch for the incredibly useful discussion that I had with him following my presentation of Chapter 2 at the 2013 Economic Society of South Africa meetings at the University of the Free State. Prof. Frank Chaloupka for giving me feedback and encouragement during the Economics of Tobacco Control Conference held at the University of Cape Town in 2014. Prof. Ken Warner for taking the time to have lunch with me at the

World Conference on Tobacco or Health (WCTOH) in Singapore in 2012 and following it up with dinner at the 2015 WCTOH in Abu Dhabi. Prof. John Paul Dunne who, as convenor of the PhD Brownbag Seminar series at the University of Cape Town, helped clear a lot of cobwebs for me about the PhD journey. I also want to thank attendees to the Emerging Researcher Programmes held in 2013 and 2015 and the Economics of Tobacco Control Conference held in 2014 at the University of Cape Town for their valuable comments on different aspects of this thesis.

I would also like to thank my peers in the PhD lab in the School of Economics Building at the University of Cape Town. I have learnt a lot from them over the last 3.5 years. I would like to make special mention of Ramaele Moshoeshoe for alerting me to the work by Aviv Nevo and Adam Rosen that was important for Chapter 2. Godfrey Charles-Cadogan read an early draft of Chapter 2 and gave useful feedback. Other comrades who have been very instrumental in the lab include David Fadiran, Mashekwa Maboshe, Dale Mudenda, Sulemana Mahayiwa, Mhbuba Shifa, David Khaoya, Alfred Mukong, Brian Albert Monroe and the recent graduate Dr. Chijioke Nwosu.

I also like to thank my friends in South Africa and Zambia who, in their own way, have provided me with the encouragement needed to finish this task. These include Goodwell Mateyo, Dr. Kondwelani Mateyo, Chitongwa Siame, Masuzyo Chilwesa, Singumbe Muyeba, Bob Mwiinga Munyati, Vinayak Bhardwaj, Dr. Sebastien Bodenstein, Kilweza Unene, Joseph Simumba, Jeffrey Mashiane, Katherine Eyal, Valeri Sokolovski, Karmen Naidoo, Tabitha Kamoto, Christopher Simusokwe and Benjamin Stanwix. In some ways this thesis is dedicated to my late friend Arthur Shipekesa. I know that he would have been proud of my having finished the thesis.

I should not forget to thank the Bill and Melinda Gates Foundation and the Carnegie Corporation of New York for providing the generous funding that made my PhD studies possible. A special mention for Cynthia Lewis of the Gates Foundation who I have had the pleasure of discussing some findings from my research with.

I also want to thank my family for the support and prayers. These include my sisters Nachela Chelwa and Chibuye Chelwa. My uncles Patrick Kampengele and Victor Kampengele and my aunt Marjorie

Kampengele. A very special mention for my parents Eunice and Jevarson Chelwa. I thank them for the wonderful gifts of life and education.

And finally, Mwansa Mushinge. Words escape me in describing how this work is every bit hers as it is mine.

Grieve Chelwa

Cape Town

6th August, 2015

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

For Mwansa;

For my parents

CHAPTER 1: INTRODUCTION

According to the World Health Organization (WHO), 100 million people died from tobacco-related causes in the 20th Century (WHO, 2008). Another 1 billion people are expected to die in this century if current smoking trends continue (ibid.).

This is the background against which the WHO, along with its member states, developed the Framework Convention on Tobacco Control (FCTC). The FCTC, which entered into force on 27th February 2005, encourages Parties to put into place a wide range of tobacco control measures. Currently there are 180 Parties to the treaty including many Low- and Middle-Income countries where the burden of the tobacco epidemic is likely to be greatest in the 21st Century (WHO, 2012a).[1] Even though most of the 47 countries in WHO's Africa region (also known as AFRO region)[2] are Parties to the FCTC, the region is yet to make significant progress in fully implementing the treaty.

One of the ways of assessing this is to measure how the region has fared in implementing Article 6 of the FCTC which is an important Article that encourages countries to periodically increase prices and taxes on tobacco products. The Article derives its importance from decades of research showing the primacy of tax and price measures in reducing tobacco consumption (Chaloupka and Warner, 2000; IARC, 2011). The most recent round of assessments on the implementation of the FCTC showed that WHO's AFRO region achieved an implementation rate of 49% for Article 6 against a global average of 61% (WHO, 2014a).[3] In addition and related to this point, the AFRO region was the only region whose excise taxes on cigarettes declined in real terms over the period 2008 to 2012 (see Figure 1.1). The

---

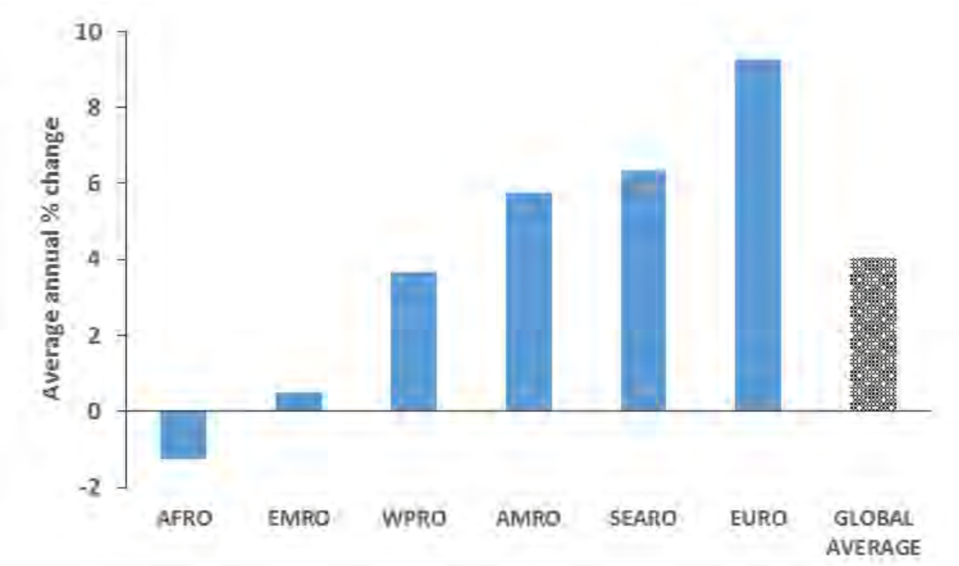
[1] The up-to-date list of Parties to the FCTC can be found here: https://treaties.un.org/pages/ViewDetails.aspx?src=TREATY&mtdsg_no=IX-4&chapter=9&lang=en.

[2] The WHO is made up of 6 regional groupings. These are AFRO for the African regional group, AMRO for countries in the Americas region and EMRO for the Eastern Mediterranean region. Other groups are EURO for the European region, SEARO for the South East Asia region and WPRO for the Western Pacific region. The list of countries that make up each region can be accessed here: http://www.who.int/about/regions/en/.

[3] Countries are scored on the basis of how many of an Article's indicators they fully comply with. In this case, countries in the AFRO region had by 2014 only complied with 49% of Article 6's indicators. The indicators on which countries are scored for Article 6 are the following: (1) Tax policies to reduce tobacco consumption implemented, (2) sales to international travellers of tobacco products prohibited or restricted, (3) tobacco imports by international travellers prohibited or restricted. See WHO (2014a, p75) for more details.

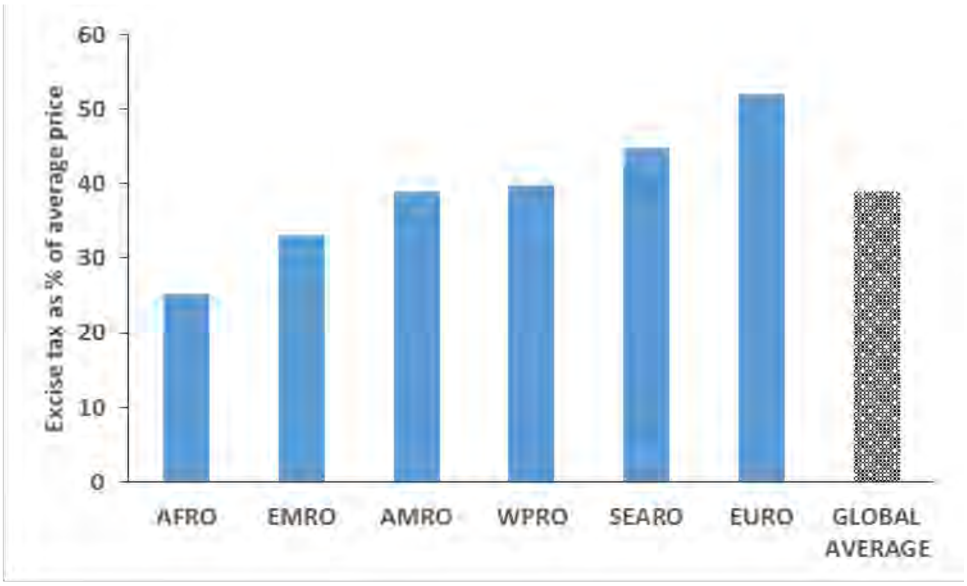
real excise tax in AFRO declined on average by about 1% per annum over the period 2008 to 2012 whereas globally, real excise taxes increased by about 4% per annum (Eriksen *et al.*, 2015). Further, excise taxes are relatively low in the AFRO region. A measure of the level of excise taxes is the excise tax burden, which is calculated as the percentage of the retail selling price that is due to the excise tax. The most recent internationally comparable data show that the excise tax burden on cigarettes in AFRO was 25% in 2012 against a global average of 39% (WHO, 2013a). This excise tax burden was the lowest in all 6 regions (see Figure 1.2) and was well below the 70% target recommended by WHO (WHO, 2010).

Figure 1.1: Average annual percentage changes in the real excise tax on cigarettes across WHO regions, 2008 to 2012



Notes: The figure shows average annual percentage changes in the real excise tax levied on the most popular brand of cigarettes in different WHO regions and the global average over the period 2008 to 2012. AFRO refers to the Africa region, AMRO to the Americas region and EMRO to the Eastern Mediterranean region. EURO refers to the European region, SEARO to the South East Asia region and WPRO to the Western Pacific region. The data are taken from Eriksen *et al.* (2015).

Figure 1.2: Excise tax burdens on cigarettes across WHO regions as at 2012

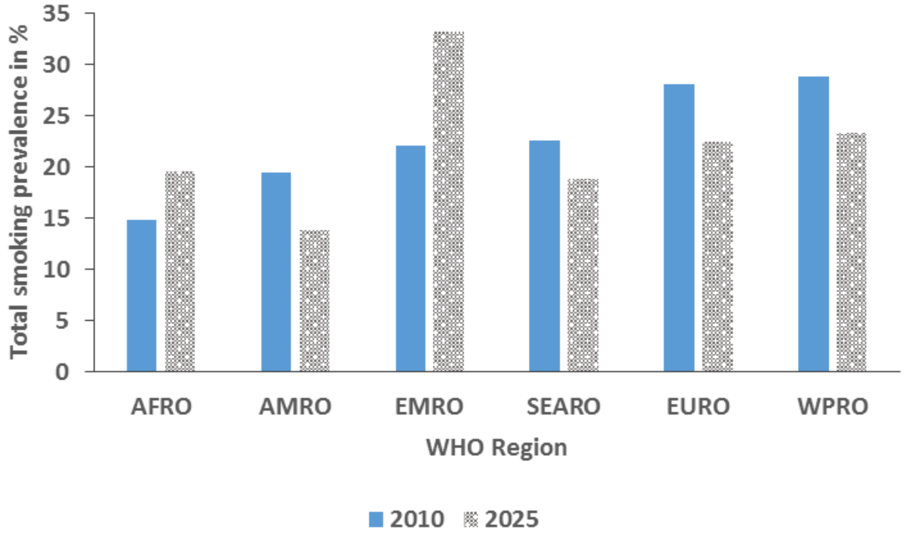


Notes: The figure shows the excise tax burden on cigarettes in different WHO regions and the global average in 2012. The excise burden is calculated as the percentage of the average selling price that is due to the excise tax (whether the tax is levied as a specific tax or an *ad valorem tax* or a combination of the two). AFRO refers to the Africa region, AMRO to the Americas region and EMRO to the Eastern Mediterranean region. EURO refers to the European region, SEARO to the South East Asia region and WPRO to the Western Pacific region. The data are from WHO (2013a).

Currently the AFRO region has the lowest smoking prevalence estimates in all of WHO's 6 regions (see Figure 1.3). Total smoking prevalence for AFRO in 2010, the latest year for which we have internationally comparable data, was 15% (WHO, 2015). However, AFRO is only one of two regions where total smoking prevalence is projected to increase by 2025 if the current level of implementation of tobacco control efforts is maintained (see Figure 1.3). Total smoking prevalence in AFRO is projected to increase to 20% by 2025 (Bilano *et al.*, 2015; WHO, 2015). This means that in absolute terms, the number of smokers in Africa will double from 71 million people in 2010 to 145 million people in 2025.[4] The only other region projected to experience an increase in smoking prevalence is the Eastern Mediterranean region (EMRO) where it is expected to increase from 22% in 2010 to 33% in 2025 (Bilano *et al.*, 2015; WHO, 2015).

[4] The 2010 estimate for the number of smokers is arrived at by applying adult smoking prevalence estimates from WHO (2015) on the population of adult Africans in 2010. The population estimates are from the United Nations (2012). The number of smokers for 2025 is similarly calculated using projections for smoking prevalence in 2025 (Bilano *et al.*, 2015; WHO, 2015) and projections for Africa's adult population in 2025 (United Nations, 2012).

Figure 1.3: 2010 and 2025 (projected) total adult smoking prevalence in different WHO regions



Notes: The figure shows 2010 and projected total smoking prevalence for 2025 in different WHO regions. Total smoking prevalence is defined as the average of adult male and adult female smoking prevalence estimates. Smoking prevalence is in turn defined as the percentage of adults (15 years and above) who currently smoke. AFRO refers to the Africa region, AMRO to the Americas region and EMRO to the Eastern Mediterranean region. EURO refers to the European region, SEARO to the South East Asia region and WPRO to the Western Pacific region. The method and assumptions behind the projections are contained in Bilano *et al.* (2015). The data are taken from WHO (2015).

The predicted increase in smoking prevalence and in the number of smokers over the next 10 to 15 years in Africa will coincide with an increase in the burden of diseases associated with tobacco consumption. Tobacco use is a leading risk factor for many Non-communicable diseases (NCDs) such as cardiovascular disease, stroke, diabetes, chronic respiratory disease and lung and other cancers (WHO, 2014b). The percentage of deaths due to these diseases in Africa is projected to increase from 12% in 2002 to 17% in 2030 with the total number of deaths doubling from 1.3 million to 2.5 million (Mathers and Loncar, 2006).

Article 20 of the FCTC requires that “Parties undertake to develop and promote national research and to coordinate research programmes at the regional and international level in the field of tobacco control” (WHO, 2005, p17). The Article further requires that Parties “initiate the conduct of research and scientific assessments, and in so doing promote and encourage research that addresses the determinants and consequences of tobacco consumption” (ibid., p18). To date, countries in the AFRO region have not fared well in implementing this Article. According to the FCTC Secretariat, the AFRO region achieved an implementation rate of 33% for Article 20 in their most recent assessment against a global average of 51% (WHO, 2014a).[5] This is also reflected in the fact that there is little published research focussing on tobacco control in Africa outside of South Africa. For instance, a recent comprehensive review of the large literature on price elasticities of demand for tobacco products by the International Agency for Research on Cancer (IARC) was only able to count 3 African countries, outside of South Africa, as having published elasticity estimates (IARC, 2011). Research that speaks to local conditions is vital in building the case for policy action especially in view of the prognosis for the tobacco epidemic on the continent as detailed above. This thesis, therefore, aims to fill some gaps in our knowledge of the economics of tobacco control, with particular attention paid to Africa.

1.1. Themes of the thesis

This thesis concerns itself with three broad but related themes in the economics of tobacco control. The first theme deals with the quantification of the costs of tobacco consumption and in this way sets up the economic justification for policy intervention. The second theme, which is one of the oldest research areas in the economics of tobacco control, attempts to quantify the responsiveness of tobacco demand to price and/or tax changes. This follows directly from policymakers' desire to influence demand given the costs of tobacco consumption. The last theme concerns itself with conducting policy evaluations of countries that have engaged in large-scale tobacco control interventions with the aim of drawing wider policy lessons. In what follows, I discuss each one of these themes in some detail.

1.1.1. The economic costs of tobacco consumption

In economics, particularly neo-classical economics, policy interventions are often justified on market failure grounds. A market failure or an externality occurs when the actions of one agent directly affect the well-being of another agent (Varian, 1992). When a market failure occurs, policy actions are required to correct or to "internalize" the externality. The external costs associated with tobacco consumption have traditionally been thought of in terms of mortality and morbidity. When smokers fall ill, a portion of their medical costs are met by non-smokers especially in countries with publically provided healthcare. By some estimates, healthcare costs attributable to smoking take up between 6% and 15% of total healthcare costs in high-income countries (Jha *et al.*, 2000). Secondly, premature death as a result of tobacco smoking also imposes costs on the wider economy due to productivity

[5] See WHO (2014a, p78) for a complete listing of Article 20's indicators.

losses. For instance, in the United States, the Department of Health and Human Services estimates that more than USD156 billion is lost per annum in productivity due to premature mortality (United States Department of Health and Human Services, 2014). It is, however, important to point out that some of the productivity losses are borne by smokers themselves. For instance, smokers are known to face lower wages in the labour market (Levine *et al.*, 1997; Viscusi and Hersch, 2001; Lye and Hirschberg, 2004). External costs also include the mortality and morbidity costs of third parties who are exposed to environmental tobacco smoke. Further, new work has pointed to the importance of cigarette excise taxes as a self-control device for smokers. That is, in addition to the role that taxes play in correcting for externalities, they are useful in helping smokers commit to quitting in the future. The theoretical and empirical evidence for this so-called "internalities" argument for cigarette taxes has been shown by Gruber and Koszegi (2001, 2004) and Gruber and Mullainathan (2005), among others.

Recently, researchers have started to think about the implications of tobacco consumption for intra-household resource allocation. In many Low- and Middle-Income countries (LMICs), household consumption decisions are often made by the head of the household. In contexts where incomes are low, the head's decisions can have far reaching consequences for household consumption patterns. That is, in deciding to spend money on cigarettes, the head is choosing to spend less on other goods and services. Spending less on, for example, education or health can consign poor households into a vicious cycle of poverty (John *et al.*, 2012). These displacement costs are referred to as the "opportunity costs" of tobacco consumption in the literature.

The opportunity costs of tobacco consumption can be quite large as demonstrated by Efroymson *et al.* (2001), who were the first researchers to quantify the costs in the context of Bangladesh. They found that the average male smoker spent on tobacco more than twice the per capita expenditure on clothing, housing, health and education combined. Further, they found that the typical poor smoker's daily cigarette expenditure could easily pay for an additional 500 calories for one or two children in Bangladesh. Following on from their study, other researchers have attempted to quantify these opportunity costs for Cambodia (John *et al.*, 2012), India (John, 2008b), Indonesia (Block and Webb, 2009), South Africa (Koch and Tshiswaka-Kashalala, 2008), Taiwan (Pu *et al.*, 2008) and the United States (Busch *et al.*, 2004). Even though these later studies have employed more sophisticated econometric techniques, their conclusions have largely been in line with what Efromyson *et al.* (2001)

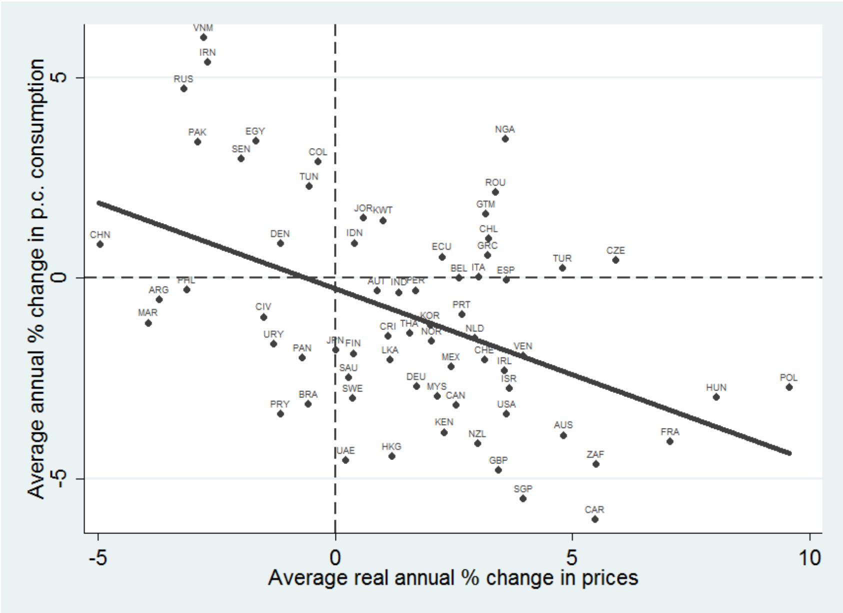
initially found in their landmark study on Bangladesh. That is, tobacco consumption often leads households to sacrifice expenditure on desirable goods and services.

1.1.2. Measuring the responsiveness of cigarette demand to changes in prices

Of all demand reduction tools that are available to policymakers, raising taxes so that prices rise has been found to be the most effective (Chaloupka and Warner, 2000; IARC, 2011). This is a consequence of the Law of Demand which states that the quantity demanded of a good falls whenever the price of the good rises, holding everything else constant. The nature of the relationship between tobacco demand and tobacco prices can also be illustrated in the form of a graph. In Figure 1.4, I plot the average annual percentage change in per capita cigarette consumption against the average annual percentage change in real cigarette prices for 67 countries. The annual percentage changes are calculated over the period 1990 to 2009 using a constant growth regression (i.e. $lnY_t = \alpha + \beta t + u_t$, where Y_t is cigarette consumption per capita or cigarette prices and β is the growth rate over the period 1990 to 2009). As is clear from the figure, the relationship between the two variables is negative. In other words, countries whose cigarette prices have increased over the period 1990 to 2009, have, on average, experienced a reduction in per capita cigarette consumption over the same period. On the other hand, countries whose real cigarette prices have declined, have on average experienced an increase in per capita cigarette consumption.[6]

[6] The relationship between changes in cigarette prices and changes in per capita consumption is also influenced by other variables, such as the rate of growth in real GDP per capita. For example, cigarette consumption might grow in fast growing economies even when real prices are increasing (or consumption could decline in countries experiencing an economic slowdown even when real prices are declining). This might explain the presence of some countries in the 1st and 3rd quadrants in Figure 1.4.

Figure 1.4: Relationship between percentage changes in real prices and percentage changes in per capita cigarette consumption



Notes: The figure shows a scatterplot of the percentage change in per capita cigarette consumption against the percentage change in real cigarette prices for 67 countries for which I have data. It covers the period 1990 to 2009. The figure also shows a fitted line of a regression of changes in consumption on changes in prices. The beta coefficient on the change in price variable is -0.425 [p-value = 0.00; confidence interval (-0.226, -0.623); n = 67]. The per capita cigarette consumption data are taken from ERC Group (2010) and the cigarette price data are taken from the Economist Intelligence Unit's World Cost of Living Survey. The cigarette price is the price of the cheapest brand (in packs) found in each city in a particular country expressed in constant 2000 US dollars (I discuss each data source in detail in Chapter 4). The countries and their symbols are Argentina (ARG), Australia (AUS), Austria (AUT), Belgium (BEL), Brazil (BRA), Cameroon (CAR), Canada (CAN), Chile (CHL), China (CHN), Colombia (COL), Costa Rica (CRI), Ivory Coast (CIV), Czech Republic (CZE), Denmark (DEN), Ecuador (ECU), Egypt (EGY), Finland (FIN), France (FRA), Germany (DEU), Greece (GRC), Guatemala (GTM), Hong Kong (HKG), Hungary (HUN), India (IND), Indonesia (IDN), Iran (IRN), Ireland (IRL), Israel (ISR), Italy (ITA), Japan (JPN), Jordan (JOR), Kenya (KEN), South Korea (KOR), Kuwait (KWT), Malaysia (MYS), Mexico (MEX), Morocco (MAR), Netherlands (NLD), New Zealand (NZL), Nigeria (NGA), Norway (NOR), Pakistan (PAK), Panama (PAN), Paraguay (PRY), Peru (PER), Philippines (PHL), Poland (POL), Portugal (PRT), Romania (ROU), Russia (RUS), Saudi Arabia (SAU), Senegal (SEN), Singapore (SGP), South Africa (ZAF), Spain (ESP), Sri Lanka (LKA), Sweden (SWD), Switzerland (CHE), Thailand (THA), Tunisia (TUN), Turkey (TUR), United Arab Emirates (UAE), United Kingdom (GBP), United States of America (USA), Uruguay (URY), Venezuela (VEN) and Viet Nam (VNM).

On the other hand, excise taxes are often an important source of revenue. The World Health Organization (WHO) estimates that governments across the globe collected at least USD145 billion in tobacco excise taxes for the 2012 fiscal year (WHO, 2013a). Governments understandably worry that as tobacco consumption falls in response to tax hikes, tax revenues might fall along with it. One of the biggest achievements in the economics of tobacco control has been to point out that a contradiction does not necessarily exist between public health goals, on the one hand, and tax revenue goals, on the other. It turns out that tobacco demand decreases less than proportionately to a price increase. And the parameter that has been crucial in driving home this point has been the price elasticity of demand for tobacco products.

Economists working in tobacco control have spent the last three decades or so estimating price elasticities of demand for tobacco products. These estimates give us a sense of the nature of the relationship between tobacco demand and tobacco prices. This work has focussed on the demand for cigarettes rather than on other forms of tobacco simply because data and prices for cigarettes are readily available. Secondly, cigarettes are a standardized product and thus allow researchers to easily conduct cross-country comparisons.

For example, an estimate of the price elasticity of demand can be obtained using the data in Figure 1.4. This is done by running a regression of the percentage change in per capita cigarette consumption on the percentage change in cigarette prices. The coefficient on the change in price variable is the estimate of the price elasticity of demand (i.e. the slope of the line in Figure 1.4 is the estimate of the price elasticity of demand). The regression coefficient on the change in price variable is -0.425 (p-value = 0; 95% confidence interval [-0.226, -0.623], n = 67). This suggests that for the data in Figure 1.4, a 1% increase in prices results, on average, in a 0.425% reduction in per capita cigarette consumption.[7]

The predominant approach in the literature has been to use time series data and appropriate time series techniques to estimate elasticities with most elasticity estimates coming from developed countries. For instance, a sizable number of the studies reviewed in the comprehensive literature surveys by Chaloupka and Warner (2000) and by the International Agency for Research on Cancer (2011) were in this tradition. The reasons for this were largely practical: during the time that researchers were beginning to estimate price elasticities of demand for tobacco products, time series econometrics was the dominant way of doing empirical economics. Furthermore, Low-and-Middle Income Countries (LMICs) often did not maintain time series records of a time span long enough to justify the use of time series estimation techniques. On the other hand, these datasets have been in abundance in developed countries.

[7] This regression is underpinned by the following simplifying assumptions: (1) changes in per capita GDP (or some other variable) do not influence per capita cigarette consumption and (2) there is no reverse causation from per capita consumption to prices. One of the preoccupations of the academic work on estimating price elasticities of demand has been to figure out how best to control for these confounding factors. Interestingly, however, even with these caveats, this elasticity estimate (and its confidence interval) is well within the range of estimates in the literature (see Chapter 3 for a detailed discussion of the literature).

The last decade or two has seen an increase in the number of household budget surveys or household expenditure surveys across LMICs. This, combined with theoretical, econometric and computer advances in using household surveys has led to an increase in the number of researchers interested in estimating price elasticities for LMICs. There are now a lot more estimates of the price elasticity of demand for cigarettes in, for example, China (Lance *et al.*, 2004; Bishop *et al.*, 2007; Chen and Xing, 2011) and India (John, 2005; John, 2008a; Guindon *et al.*, 2011) than 20 years ago. These estimates are generally in line with those from developed countries, namely that cigarette demand responds less than proportionately to price increases.

Even though we now have abundant datasets and the necessary econometric tools, only a handful of countries in Africa have estimated price elasticities of demand. The IARC (2011) *Handbook on the Effectiveness of Tax and Price Policies for Tobacco Control*, the most comprehensive up-to-date review of the literature, was only able to find elasticity estimates for the following African countries: Egypt, Morocco, South Africa and Zimbabwe. Two of these countries (Egypt and Morocco) are not part of WHO's AFRO region.

1.1.3. Evaluating tobacco control initiatives

As is clear from Figure 1.4, some countries have been particularly successful at reducing per capita cigarette consumption over the period 1990 to 2009. For instance, France, where real cigarette prices increased by about 7% per annum over the period 1990 to 2009, experienced a halving of consumption over the same period (Jha and Peto, 2014). This substantial decline in cigarette consumption coincided with a 50% reduction in the lung cancer death rate among French men (Hill, 2013). Another country which has been very successful in reducing tobacco use is the United States (US), which pioneered tobacco control policies in a number of dimensions following the 1964 Surgeon-General's report. According to Figure 1.4, the US experienced a 3% average increase in real prices with consumption declining by about 3% per annum over the 1990 to 2009 period. Over a longer time horizon (1964 to 2012), US smoking prevalence declined from 42% to 18%, with much of this attributable to an aggressive tax policy (United States Department of Health and Human Services, 2014). Much of the decline in prevalence in the United States has been attributed to increases in taxes at the state rather than at the federal level and the declines have, consequently, been uneven across states (Peterson *et al.*, 1992; Fichtenberg and Glantz, 2000; Abadie *et al.*, 2010).

Some Low- and Middle-Income countries (LMICs) have also had their fair share of success with tobacco control. South Africa, which was one of the first LMICs to adopt an aggressive tobacco control policy in the 1990s, saw aggregate cigarette consumption decline by a third between 1994 and 2004 (Van Walbeek, 2005). According to Figure 1.4, real cigarette prices increased at the rate of 5% per annum during most of this time. Thailand managed to reduce smoking prevalence from 24% in 1991 to 15% in 2007 (Sangthong *et al.*, 2012). During most of this time, real cigarette prices in Thailand increased at the average rate of 2% per annum (see Figure 1.4).

Given that a sizable number of countries have been implementing tobacco control measures over the last couple of decades, the interest of some researchers has shifted to evaluating the effectiveness of these interventions. The 2008 publication of the International Agency for Research on Cancer's (IARC) handbook on the *Methods for Evaluating Tobacco Control Policies* is testament to this. The overriding focus of this research agenda is the construction of policy counterfactuals. That is, the research tries to figure out what the world would have looked like in the absence of tobacco control measures. The success (or lack thereof) of any intervention is arrived at by comparing outcomes under tobacco control with outcomes under a counterfactual scenario.

Warner (1977) was one of the first to tackle this problem. He predicted what cigarette consumption patterns would have been in the US if anti-smoking measures had not been introduced in the 1960s by using regression coefficients from the pre-smoking campaign period. Comparing predicted to actual consumption data, he calculated that per capita cigarette consumption was about 30% lower in 1975 than it would have been had the anti-smoking campaign measures not been implemented. Fichtenberg and Glantz (2000) and Holford *et al.* (2014) have used versions of Warner (1977) in evaluating tobacco control initiatives in California and the entire United States respectively. Recent research attention has, however, shifted to methods that rule out the possibility that the treatment effects identified in, for instance, Warner (1977) or Fichtenberg and Glantz (2000) did not happen by coincidence. That is, the aim of this research is to rule out the possibility that the decline in cigarette consumption would have happened even in the absence of tobacco control measures. For example, the Synthetic Control method developed by Abadie and Gardeazabal (2003) and applied in the context of tobacco control in Abadie *et al.* (2010) tries to do exactly that.

1.2. Outline of the thesis

This thesis comprises of three main chapters in line with the themes outlined above. I discuss each chapter's main arguments below.

Chapter 2 uses data from the 2006 edition of Zambia's Living Conditions Monitoring Survey (LCMS), a household expenditure survey, to investigate whether tobacco consumption influences household expenditure patterns in Zambia. Previous research has approached this problem by estimating a system of conditional demand functions (or conditional Engel curves) where the demand for each expenditure item is estimated as a function of a household's expenditure on tobacco (or a household's tobacco smoking status). That is, conceptually households are thought of as initially allocating expenditure to tobacco before allocating expenditure to other items. However, since there is likely to be reverse causation from, say, expenditure on food to whether a household smokes or not (or indeed even simultaneity bias), researchers have had to employ instrumental variable techniques to obtain consistent estimates. The instrumental variable of choice in the literature for tobacco expenditure (or a household's tobacco smoking status) has been the adult male to adult female ratio (or adult sex ratio). That is, the adult sex ratio is a good predictor of whether a household has at least one smoker because adult males are more likely than adult females to smoke in LMICs. I argue in the chapter that even though the adult sex ratio is informative about the endogenous regressor (that is, it predicts smoking status), it is likely to violate the exclusion restriction in the sense that it also influences the left-hand side variables. For example, a household with more males relative to females might spend less on health care. I, therefore, use the adult sex ratio as per the literature but allow for it to violate the exclusion restriction in the manner suggested by Nevo and Rosen (2012). This allows me to obtain consistent estimates of the effect of tobacco expenditure on a household's expenditure patterns in the presence of an imperfect instrumental variable. A previous version of this Chapter was published in 2014 as an Economic Research Southern Africa (ERSA) Working Paper (Chelwa and Van Walbeek, 2014).

In Chapter 3, I use the 2005 and 2009 editions of Uganda's National Panel Survey (UNPS), a household expenditure survey, to estimate price and expenditure elasticities of demand for cigarettes in Uganda using a method developed by Deaton (1988). Given that the UNPS does not contain cigarette price data, I construct a measure of price as a ratio of household expenditure on cigarettes to total quantity of cigarettes purchased over a reference period. Deaton refers to this measure of price as a "unit value". I show that unit values vary across geographical space in Uganda and use this fact to obtain price and expenditure elasticities of demand that are largely free of reverse causation or simultaneity bias concerns. Deaton's method also allows one to correct the elasticity estimates for measurement error and cigarette quality differences. This is the first study to apply Deaton's method in estimating price and expenditure elasticities of demand for tobacco products in the context of an African country.

In tobacco control, South Africa is often held up as an exemplar country. Between 1994 and 2004, the average real price per pack of cigarettes in South Africa increased by 110% owing largely to consistent tax increases by the government. During the same period, per capita cigarette consumption declined by 46% (Van Walbeek, 2005). However, comparing 1994's per capita cigarette consumption with the 2004 number is not particularly useful for evaluating policy. Many other factors, aside from the price rises, could have influenced cigarette consumption during the period 1994 to 2004. Furthermore, per capita cigarette consumption had already started declining by the time the tax started increasing. In Chapter 4, I create a counterfactual per capita cigarette consumption trend line for South Africa for the period 1994 to 2004 using the method developed by Abadie and Gardeazabal (2003). That is, I create what per capita cigarette consumption would have looked like in South Africa had the tax increases not happened. In addition, I conduct several falsification (or placebo) exercises to verify that I am not incorrectly attributing the changes in cigarette consumption to changes in cigarette prices.

Chapter 5 concludes by giving a summary of the main findings of the thesis and discusses ideas for future research.

1.3. Some comments on the choice of countries

Ideally, one would want to conduct the analysis in Chapters 2 and 3 on the same country. That is, one chapter would establish the economic case for policy intervention while the other chapter would give policymakers an idea of what the pattern of demand responses would look like should they choose to intervene. I had initially started off with the idea of writing chapters 2 and 3 exclusively on Zambia. Unfortunately, none of the available household expenditure surveys in Zambia contain the quantity variable for cigarettes, which is crucial if one wants to use Deaton's (1988) method. On the other hand, Uganda is a good substitute. For one thing, the first two rounds of the Uganda National Panel Survey (UNPS) contain all the relevant variables. Secondly, Uganda, from a context point of view, is sufficiently similar to Zambia, and to many other sub-Saharan African countries, to allow for some generalization of the results. Lastly, Uganda is itself in the process of implementing several tobacco control measures including increasing the excise tax on cigarettes.

The choice of South Africa for Chapter 4 follows from the fact that it is one of the few countries in sub-Saharan Africa that has consistently increased taxes over a sufficiently long time span to allow for a proper policy evaluation.

CHAPTER 2: DOES TOBACCO EXPENDITURE INFLUENCE HOUSEHOLD SPENDING PATTERNS? EVIDENCE FROM ZAMBIA

2.1. Introduction

Tobacco use is the leading cause of premature death in the world. According to the World Health Organization (WHO), nearly 6 million people die per annum from tobacco related deaths (WHO, 2014c) including about 600,000 from passive smoking (Oberg, *et al.*, 2010). There is another aspect to the cost of tobacco use that is separate from mortality and morbidity costs, namely that tobacco expenditure tends to crowd out the consumption of other commodities within the household. This chapter adds to the literature on the crowding out effect of tobacco expenditure in two ways. Firstly, the chapter uses expenditure data from a low income sub-Saharan African country, Zambia, where most of the households are poor. Aside from the work by Koch and Tshiswaka-Kashalala (2008) on South Africa, I am unaware of studies that investigate this issue in the context of sub-Saharan Africa. Secondly, in identifying the causal impact of tobacco expenditure I use the standard instrumental variable from the literature, the adult sex ratio, but unlike previous work, I allow for this instrument to be correlated with the error term. I consider this approach to be plausible given that intra-household resource allocations, and as a consequence household expenditure patterns, often display biases related to gender (Senauer *et al.*, 1988; Deaton *et al.*, 1989; Thomas, 1990, 1994; Deaton, 1997; Duflo, 2003; Doss, 2013). Therefore, excluding the adult sex ratio from structural demand equations is likely to result in biased estimates of the impact of tobacco expenditure on household decision making.

My results, even after allowing for a correlation between the instrumental variable and the error term, confirm many of the findings in the literature. For instance, I find that households with at least one smoker allocate less expenditure towards food, schooling, clothing and water. Other expenditure categories that are crowded out include transportation, equipment maintenance, entertainment and remittances. I also find some evidence that the crowding out of food is more pronounced for poorer households. But unlike previous studies, I do not find that spending on tobacco leads households to allocate more expenditure towards alcohol. My empirical analysis, which is based on less restrictive assumptions on the instrumental variable, leads to the conclusion that the positive association between tobacco and alcohol found in previous work is more likely a correlational relationship than a causal relationship.

The results in this chapter show that the costs of tobacco consumption in Zambia are more than the direct costs associated with mortality and morbidity. I agree with Wang *et al.* (2006), Block and Webb (2009) and John *et al.* (2012) who conclude that by displacing expenditure categories like food, education and the maintenance of equipment, households in poor countries are likely to be trapped in a cycle of poverty.

My interest in studying the impact of tobacco consumption on household decision making in Zambia is motivated by the fact that per capita cigarette consumption has recently started rising after declining over much of the 1990s (Chelwa, 2012). For the period 1990 to 2001, per capita cigarette consumption declined by 75%. During this period, real GDP per capita declined at the average rate of 1% per year.[8] Further, the 1990s were a period of great economic hardship as the country implemented a series of austerity measures including a controversial privatization programme that resulted in thousands of job losses (Muuka, 1997; Craig, 2000). From 2002 to 2009, per capita cigarette consumption increased by 37% largely supported by the recovery in the economy's performance (Chelwa, 2012). Growth in real GDP per capita averaged 4% per year during the previous decade. There is an expectation that per capita cigarette consumption will continue to increase in the current decade (ERC, 2010). This suggests that tobacco is increasingly becoming an important part of the expenditure decisions of most households in Zambia. Given that households in general face a budget constraint, it, therefore, becomes important to investigate which goods and services tobacco displaces, if at all any, in the household's budget.

The rest of the chapter is structured as follows: Section 2.2 gives an overview of the relevant literature. Section 2.3 discusses the conceptual framework of how crowding out might occur. Section 2.4 discusses the empirical strategy and the data is described in section 2.5. The empirical results are presented in Sections 2.6 and 2.7. Section 2.8 concludes.

[8] http://data.worldbank.org/data-catalog/world-development-indicators (accessed December 2015)

2.2. Literature review

The costs associated with household tobacco consumption are often conceptualised in one of two ways: (1) costs on the macroeconomy attributed to smoking via death, increased healthcare expenditure and lost productivity (Chaloupka and Warner, 2000; Kang *et al.*, 2003; Max *et al.*, 2004; Liu *et al.*, 2006) and those related to the displacement of some commodities by tobacco in the household. My review of the literature focusses on the latter as it bears directly on the focus of this chapter.

Efroymson, *et al.* (2001) was one of the first studies to highlight the potential crowding out effect of tobacco consumption using several datasets from Bangladesh. Theirs was not so much an econometric study as a simple comparison of the expenditure profiles of smoking versus non-smoking households. One of the study's main findings was that male cigarette smokers spent more than twice as much on cigarettes as on clothing, housing, health and education combined. Further, the typical smoker could add more than 500 calories to the diet of one or two children with the money spent on cigarette purchases. Since their analysis was not econometric in nature, it did not account for observable confounders, variables that were likely, alongside tobacco, to influence expenditure allocations between the two types of households. Further, Efroymson *et al.* did not account for the possibility that the decision to allocate expenditure towards tobacco might be determined endogenously within the household. If observable and unobservable confounders are not accounted for, then it becomes unclear whether reducing a smoking household's expenditure on tobacco will elevate that household's consumption profile to that of a non-smoking household.

Busch *et al.* (2004) and Wang *et al.* (2006) added to the literature by estimating demand systems that adequately controlled for social, demographic and geographic variables likely to impact expenditure decisions. The former estimated an Almost Ideal Demand System (AIDS) using the Consumption and Expenditure Survey from the United States and found that tobacco crowded out food and clothing. The latter used data from rural China and found that tobacco crowded out expenditure on education, agriculture equipment maintenance and savings. Wang *et al.* also found that tobacco smoking households were likely to spend more money on alcohol, further exacerbating the negative expenditure impact of smoking.

The next generation of studies in this literature attempted to address the issue of endogeneity which had not been adequately dealt with up to that point. John (2008b), using data from India's National Sample Survey was the first to use instrumental variables to account for the possible endogeneity of tobacco use in the demand system. His choice of instrument was the adult sex ratio motivated by the fact that smoking in India was mainly done by adult males. In addition, John's analysis controlled for possible preference differences between smoking and non-smoking households using a method introduced by Vermeulen (2003). John found that tobacco expenditure crowded out food, education and entertainment while crowding in expenditure on health, clothing and fuels. Pu *et al.* (2008) using John's (2008b) method and expenditure data from Taiwan found that tobacco crowded out clothing, medical care and transportation amongst others. Pu *et al.*'s contribution was to treat alcohol and tobacco as complements in the demand system, which allowed them to separately study the impact of both on household expenditure decisions. Koch and Tshiswaka-Kashalala (2008) added to the literature by using a different instrument for tobacco expenditure. Their preferred instrument was a composite smoking prevalence measure based on prevalence estimates for South Africa computed in Van Walbeek (2002). Their results showed that tobacco crowded out expenditure on education, fuel, clothing, healthcare and transportation for the full sample of smoking households. On the other hand, spending on tobacco was associated with increased expenditure on housing, food, entertainment and alcohol in some data specifications.

More recently Block and Webb (2009), lacking appropriate instrumental variables, have used an indirect approach to identify the causal impact of tobacco expenditure on household's expenditure decisions in Indonesia. The authors estimate a series of reduced form equations for food, tobacco and child height against a common set of covariates. The basic idea is that if a common set of covariates reduces the allocation to food and reduces child height but at the same time increases the allocation to tobacco, then this is suggestive of crowding out. Their indirect empirical strategy "demonstrate[s] that the same exogenous covariates that are associated with improved dietary quantity and quality are also associated with reduced allocation of household resources to tobacco" (ibid., p. 18).

My approach in this chapter is closely aligned with the most recent generation of empirical studies on the crowding out impact of tobacco consumption. This chapter's main contribution is to use the standard instrumental variable used in the literature, the adult sex ratio, while making less stringent assumptions about the behaviour of this instrumental variable.

2.3. Conceptual framework

Following the theoretical framework laid out in John (2008b), I assume that each household seeks to maximize a single utility function in the manner made precise in Samuelson (1956) and Becker (1974, 1981). The need to make this assumption is driven by the limitations in my dataset that make it difficult to incorporate intra-household interactions in the analysis. The household's utility maximization problem results, in general, in a set of n household Marshallian demand functions of the form $x_i(p_i, \dots, p_n; Y; \boldsymbol{a})$, where x_i is the quantity purchased of the ith commodity, p_i is the price of the ith commodity, $\boldsymbol{a}$ is a vector of household characteristics and Y is total household income.

The model assumes that a household that spends on tobacco, in the sense that at least one household member is a smoker, first decides on the quantity of tobacco to be purchased before deciding on the quantities of the other commodities. In such a situation, the household's utility maximization problem results in a set of conditional demand functions of the form $x_i = g_i(p_i, \dots, p_n; M; \boldsymbol{a};\ d)$, where d is an indicator variable for whether the household spends on tobacco and M is the remainder of household income after spending on tobacco. That is, the household's demand for commodity i is conditional on the household's smoking status. Pollak (1969) formally introduced and discussed the properties of conditional demand functions and showed that they obeyed the theory of demand.

In this chapter, I seek to estimate and compare Marshallian demand functions for non-smoking households with conditional demand functions for households with at least one smoker for a common set of commodities. If, on average, the quantity demanded of a commodity for the typical non-smoking household is less than the quantity demanded of the same commodity for a typical smoking household, then the difference can be attributed, *ceteris paribus*, to tobacco.

2.4. Empirical strategy

I conduct my empirical analysis in two parts. In the first part, I compare the mean expenditure shares for various commodity groups between smoking and non-smoking households. In particular, the comparisons are conducted for the following commodities: food, alcohol, healthcare, schooling, water, housing, electricity, alternative energy sources, transportation, equipment maintenance (boats, cars, motor bikes and bicycles), telephone usage, entertainment, house care, personal care and "other".[9]

The second part of my empirical strategy formally tests the crowding out hypothesis of tobacco expenditure for the commodities listed in the previous paragraph. To do so, I estimate Engel curves using the Quadratic Almost Ideal Demand System (QUAIDS) developed by Banks, *et al.* (1997). The QUAIDS has the advantage of not only being consistent with utility theory but also nesting the popular Almost Ideal Demand System (AIDS) of Deaton and Muellbauer (1980a) and further allows commodities to be modelled as luxuries at some income levels and necessities at others. John (2008b), Pu *et al.* (2008), Koch and Tshiswaka-Kashalala (2008) and John *et al.* (2012) have used the QUAIDS to conduct similar analyses. I assume that the household is a single utility maximizer and therefore estimate a system of household-level Engel curves with each one taking the following form:

$$w_{ij} = \alpha_{1i} + \alpha_{2i}d_j + \alpha_{3i}lnM_j + \alpha_{4i}(\ln M)_j^2 + \alpha_{5i}FE + \gamma_i\boldsymbol{a}_j + u_{ij} \qquad (2.1)$$

where w_{ij} is the monthly expenditure share of expenditure category i in household j after deducting the expenditure on tobacco. d_j is a dummy variable that takes the value of 1 if household j reports positive monthly expenditure on tobacco and zero otherwise. $\ln M_j$ is the natural logarithm of total monthly expenditure (in Zambian Kwacha) in household j excluding expenditure on tobacco. $(\ln M)_j^2$ is the square of lnM in household j. Equation (2.1) is the empirical implementation of the Marshallian and conditional demand functions discussed in Section (2.3).

[9] The category with "other" contains goods and services that are difficult to classify or too small to stand alone.

Ordinarily, budget share equations such as those in equation (2.1) should be estimated with price as an explanatory variable. In the absence of price data, I augment the Engel curves with cluster-level fixed effects (FE) under the assumption that households within the same cluster (or the same village) face the "same" price (or face the same relative prices). This assumption has empirical support especially in Low- and Middle-Income Countries (LMICs) where transportation costs are significant determinants of prices and the isolated nature of markets prevents the exploitation of arbitrage opportunities (see Deaton, 1988, 1989, 1990, 1997; Deaton and Grimard 1992).[10] FE is exogenous in my specification because an individual household's demand is too small to influence the determination and structure of cluster-level prices (see previous list of references). The FE can also be thought of as controlling for tastes which are likely to be the same within clusters but different across clusters.

$\boldsymbol{a}_j$ is a vector of household-specific characteristics that includes the natural logarithms of household size, age of household head, average age of adults in a household, average age of children in a household, years of schooling of the household head and years of schooling of the most educated member of the household. Other household characteristics in $\boldsymbol{a}$ include the proportion of adults in a household, the number of employed persons in the household and a dummy variable for whether the household head receives a wage income or not. I define adults as those who are 18 years old or older. $\boldsymbol{a}$ also includes a number of indicator variables for the type of household as classified by the local authority in which the household is located. In some data specifications, namely those that span both rural and urban households, I include a dummy variable in $\boldsymbol{a}$ for whether the household is located in an urban or rural area. I also include in $\boldsymbol{a}$ a dummy variable for whether the household grows tobacco. The controls in $\boldsymbol{a}$ are the standard ones used in the literature on the crowding out effect of tobacco (John, 2008b; Pu *et al.* 2008; John *et al.* 2012). u_{ij} is the usual error term which is assumed to be normally distributed with mean zero. I conduct the empirical analysis on the full sample, urban households, rural households and by expenditure category. Crowding out is established if α_{2i} (the coefficient on d) in equation (2.1) is negative and statistically significant.

$d, \ln M$ and $(\ln M)^2$ in equation (2.1) are likely endogenous, in the sense that they are each correlated with the error term u_{ij}. Such a situation would preclude my giving a causal interpretation to the regression coefficients in my demand system. In this case, it is desirable to use instrumental variables

[10] This point is explored further in the context of Uganda in Chapter 3.

to ensure consistent estimates. John (2008b) and Pu *et al.* (2008) instrument for tobacco expenditure using the adult sex ratio in a household. I follow their approach and instrument for d using the adult sex ratio. This choice of instrument is motivated by the fact that adult males are more likely than adult females to consume tobacco in Zambia. According to the 2007 round of the Zambia Demographic and Health Survey (ZDHS, 2007), whose timing coincides with the expenditure survey I use in this chapter, smoking prevalence among adult males was estimated at 24% while among adult females it was estimated at 0.7% (for the 2002 round of the survey, adult male and adult female smoking prevalence were estimated at 15% and 0.5% respectively). The fact that adult males are more likely to smoke than adult females has also been shown by Pampel (2008). I expect the adult sex ratio to explain a sizable proportion of the variation in d (I formally test this assertion in section 2.6.2). I also assume that the adult sex ratio is not correlated with unobservables that are contained in the error term u_{ij}. It is almost impossible to test this assumption in a just-identified case, such as I have here, where the number of instruments is equal to the number of endogenous regressors. I believe, however, that any determinants of w_{ij} that are not contained in my specification of equation (2.1) are unlikely to be correlated in a significant way with my choice of instrument for d.[11] In any case, I relax this assumption, formally known as the exclusion restriction, in section 2.7.

I instrument for $\ln M$ and $(\ln M)^2$ using the logarithm of the value of total household assets and the square of this logarithm respectively. I expect that these two variables explain a significant proportion of the variation in $\ln M$ and $(\ln M)^2$. Recent work has shown that household assets often predict household income or household expenditure in Africa (Sahn and Stifel, 2003; Young, 2012).[12] After all,

[11] Following Koch and Tshiswaka-Kashalala (2008), I constructed a cluster-level tobacco prevalence estimate calculated from the previous (2002) round of Zambia's Living Conditions Monitoring Survey (LCMS) as a possible instrument for households' tobacco smoking status in 2006 (I use the 2006 edition of the LCMS in this chapter). This is an attractive instrument because it is unlikely to be correlated with the error term in the individual households since it is calculated at the cluster level using the previous round's cross-sectional survey (the probability that a household appears in two consecutive cross-sectional surveys is very small). Whereas the direction of the relationship between this instrument and d was positive, the relationship was not strong enough to overcome the problem of identification with weak instruments (Stock, *et al.*, 2002).

[12] An alternative approach would be to use total household expenditure (and its square) as instruments for $\ln M$ and $(\ln M)^2$ respectively. One challenge to this approach is that household expenditure is also likely endogenous in the sense that it appears in the denominator on the Left-Hand Side (LHS) of equation (2.1). Recall that in obtaining the shares on the LHS, I need to subtract expenditure on tobacco from total household expenditure. Total household expenditure is, therefore, implicit in the denominator.

one of the defining characteristics of assets is that they can generate a flow of income for the holder. In any case, in Section 2.6.2 I formally test the extent to which household assets explain the variation in residual household expenditure. In addition, the value of total household assets satisfies the exclusion restriction, the assumption of no correlation with the error term, since demand functions are rarely, if ever, specified with the value of total household assets as an explanatory variable.

Since d is dichotomous, the first-stage regression relationship between d and its instrument is likely to be non-linear, best estimated by, for example, a probit. In addition, estimating a first-stage probit ensures that the predicted values for d, $\hat{d}$, are bounded between zero and one. This is not assured with a linear estimation. Estimating a first-stage probit, however, introduces the complication of the so-called forbidden regression (Angrist and Pischke, 2009, p190-192) whereby predicted values from a non-linear first-stage are directly applied to a linear second-stage regression. Doing so would risk a non-zero correlation between the first-stage residuals and $\hat{d}$ (ibid.). A way around this is to use the predicted values from the first-stage probit regression, $\hat{d}$ in my case, as an instrument for d (Heckman, 1978; Wooldridge, 2002; Angrist and Pischke, 2009). This is the approach I adopt in this chapter.

To implement the instrumental variables technique outlined above, I estimate the system in equation (2.1) by Three Stage Least Squares (3SLS) combined with Seemingly Unrelated Regression (SURE).[13] The SURE allows me to account for any within-household correlation of error terms by exploiting the structure of the covariance matrix of the errors (see Zellner, 1962). In addition, in estimating the system by 3SLS/SURE, one is required to arbitrarily drop one of the demand equations in the system otherwise the covariance matrix of error terms is singular and therefore not invertible (Takada, *et al.*, 1995). I opt to drop the equation for "other" goods[14]. I estimate the system using the *ireg* optional command in Stata which provides maximum likelihood estimates and ensures that my estimated coefficients are not sensitive to the choice of equation that is dropped.

[13] In effect, the demand system I estimate is a four-stage least squares procedure since the first two stages involve estimating a probit function for d and using this function to generate predicted values, $\widehat{d}$, which are in turn used as instruments for d in the third-stage. The fourth stage corrects the standard errors associated with my regression coefficients for within-household correlation of error terms using the SURE method.

[14] John (2008b), Pu *et al.* (2008) and John *et al.* (2012) use a similar procedure.

John (2008b), in estimating the system in equation (2.1), controlled for possible differences in preferences between smoking and non-smoking households using a method introduced by Vermeulen (2003). The way the demand system is set-up in equation (2.1) does not allow me to control for preference differences in the manner of John (2008b). In any case, the instrumental variables approach that I adopt should allow me to obtain consistent estimates even in the presence of an omitted variable that captures any differences in preferences between the two types of households.

2.5. Description of the data

The data for this chapter come from the 2006 round of the Living Conditions Monitoring Survey (LCMS) conducted by the Central Statistical Office (CSO) in Zambia. The survey was nationally representative and used a two-stage stratified cluster sample design whereby 988 clusters were selected in the first stage. The second stage saw the selection of 18,662 households from the 988 clusters distributed as 9,530 households in urban areas and 9,132 households in rural areas. Data collection took place in its entirety over a single month in December 2006. Urban households were classified as low-cost, medium-cost or high-cost according to the local authority's classification of residential areas. In rural areas, households were classified as either small-scale, medium-scale, large-scale or non-agriculture households. The survey collected a rich set of data on the living conditions of households in the areas of education, health, economic activities and employment, child nutrition, death in the households, income sources, income levels, food production, household consumption expenditure, access to clean water and more.[15] The household expenditure section of the survey asked each household to recall and report on the total expenditure allocated to a particular commodity over a reference period. In most cases the reference period was the month prior to the survey but for some commodities, such as expenditure on health care or schooling, the recall period was a year. In such cases, I converted annual expenditure to monthly expenditure by dividing it by 12. As stated in section 2.4, I focus on the following commodities: food, alcohol, healthcare, schooling, water, housing, electricity, alternative energy sources, transportation, equipment maintenance, telephone usage, entertainment, house care, personal care and "other". Table 2.1 shows some summary statistics from the sample.

As reported in Table 2.1, a total of 18622 households are in the full sample, of which 9530 (51%) are urban and 9132 (49%) are rural. The expenditure criteria (top 50% vs bottom 50%) split the sample

[15] Additional information on the Survey can be accessed here: http://catalog.ihsn.org/index.php/catalog/2258

into two equally populated samples. Poverty in the sample has a geographic profile: the bottom 50% of households are mainly of composed of rural households whereas the top 50% of households are mainly urban households. Urban households report spending five times as much on average per month than rural households (see line 7 in Table 2.1).

9% of households in the sample report spending on tobacco in the month prior to the survey. This crude measure of prevalence also exhibits regional and expenditure category differences: rural households and those in the bottom 50% have a higher prevalence than those in urban areas and the top 50% respectively. The average smoking household reports spending USD3.34 per month on tobacco products. This also has a regional and expenditure category profile: urban households and households in the top 50% spend on average five times as much on tobacco products than rural and households in the bottom 50% respectively.

The most important expenditure item in the sample is food, which was allocated 48% of total monthly expenditure for the sample as a whole. Lower income households (bottom 50%) and those in rural areas allocate more to food than richer (top 50%) and urban households respectively.

Table 2.1: Descriptive statistics from the 2006 Living Conditions Monitoring Survey

Line	Statistic	Full Sample	Urban	Rural	Top 50%	Bot. 50%
1	Number of households	18 662	9 530	9 132	9 331	9 331
2	Percentage of households in Urban areas	51.00%	100.00%	N/A	81.00%	21.00%
3	Percentage of households in Rural areas	49.00%	N/A	100.00%	19.00%	79.00%
4	Average monthly tobacco expenditure[16]	USD 3.34	USD 5.00	USD 1.60	USD 5.00	USD 1.12
5	Percent of households reporting positive tobacco expenditure	9.45%	7.00%	12.00%	8.00%	11.00%
6	Tobacco share among tobacco spending households	4.37%	3.81%	4.95%	3.34%	6.38%
7	Monthly household expenditure	USD 132.00	USD 217.00	USD 45.00	USD 221.00	USD 22.00
8	Percentage of adult females	53.00%	52.00%	54.00%	51.00%	55.00%
9	Average household size	5.25	5.30	5.20	5.70	4.74
10	Percentage of adults in household	57.00%	59.00%	54.00%	57.00%	56.00%
11	Average age of household head	41.00	40.00	43.00	40.00	42.00
12	Average age of adults in household	34.00	33.00	36.00	33.00	36.00
13	Average age of children in household	8.55	8.90	8.20	9.00	8.00
14	Average years of schooling for household head	8.53	10.00	7.00	11.00	6.80
15	Average schooling years for most educated household member	9.53	11.00	8.00	11.00	7.40
16	Average number of employed people in household	1.50	1.40	1.60	1.50	1.50
17	Percentage of household heads with wage(regular) income	31.00%	50.00%	11.00%	53.00%	8.34%
18	Percentage of small-scale agriculture households	37.00%	N/A	75.00%	11.00%	62.00%
19	Percentage of medium-scale agricultural households	5.14%	N/A	11.00%	4.00%	6.00%
20	Percentage of large-scale agricultural households	0.20%	N/A	0.42%	0.32%	0.09%
21	Percentage of fish farming households	7.00%	N/A	14.00%	3.00%	11.00%
22	Percentage of non-Agriculture households	34.00%	67.00%	N/A	51.00%	18.00%
23	Percentage of low-cost households	10.00%	19.00%	N/A	17.00%	2.00%
24	Percentage of medium cost households	7.00%	14.00%	N/A	13.00%	1.00%
Line	**Average Budget Share of Non-Tobacco Expenditure**	**Full Sample**	**Urban**	**Rural**	**Top 50%**	**Bot. 50%**
25	Food	48.00%	47.00%	50.00%	45.00%	56.00%
25	Alcohol	2.00%	2.00%	2.00%	2.00%	2.00%
27	Health	1.05%	0.80%	1.31%	0.80%	1.25%
28	School	5.72%	6.14%	5.29%	7.12%	4.32%
29	Clothing	9.20%	6.17%	12.37%	6.65%	11.76%
30	Housing	3.00%	5.70%	0.20%	5.42%	0.60%
31	Water	0.98%	1.79%	0.14%	1.60%	0.40%
32	Electricity	1.32%	2.40%	0.20%	2.48%	0.10%
33	Alternative Energy	4.36%	4.17%	4.55%	3.27%	5.45%
34	Daily Transport	1.80%	2.75%	0.8%	3.30%	0.30%
35	Other Transport	1.95%	1.99%	1.90%	2.78%	1.10%
36	Equipment Maintenance	0.71%	0.4%	1.03%	0.60%	0.80%
37	Entertainment	0.56%	0.90%	0.20%	0.90%	0.08%
38	Telephone	2.17%	3.57%	0.70%	4.08%	0.20%
39	Remittance	0.70%	0.90%	0.50%	1.10%	0.30%
40	House Care	0.88%	1.43%	0.30%	1.56%	0.20%
41	Personal Care	10.92%	9.81%	12.08%	9.81%	12.03%
42	Households per cluster	19	25	15	11	19
43	Number of clusters	988	383	605	838	987

[16] Zambian Kwacha converted to United States Dollar using the end-of-year exchange rate in 2006 obtained from www.oanda.com which was ZMK 4,200 to USD 1.

2.6. Empirical results

I present the results of my empirical analysis in two parts: part one conducts difference of means tests for expenditure shares between smoking and non-smoking households. In the second part, I present the results of the econometric implementation of equation (2.1).

2.6.1. Differences in expenditure shares between smoking and non-smoking households

Table 2.2 contains my motivation for investigating whether the decision to spend on tobacco influences households' spending decisions. The table reports differences in expenditure shares between smoking and non-smoking households expressed as percentage points. A positive percentage point difference implies that smoking households on average allocate a greater share to that category than non-smoking households. While a negative percentage point difference implies that smoking households allocate a smaller share. The actual expenditure shares from which the differences are calculated are reported in Appendix A as Table A1. In the sequel, I present the results for the full, urban, rural, upper and lower income samples. I, however, limit the discussion of the results to the latter four samples to make it easier to follow.

Table 2.2 shows that there are statistically significant differences in expenditure allocations between smoking and non-smoking households. **Food** constitutes the biggest expenditure item for the two types of households. The table shows that on average, non-smoking households allocate a greater share of expenditure towards food than smoking households with the differences being statistically significant. Upper income households[17] are the exception: the difference between the two types of households is not statistically different from zero. Differences between the two types of households are largest for rural and poorer households; non-smoking households allocate between 5 and 7 percentage points more to food than smoking households. This pattern combined with the geographic nature of poverty noted in Table 2.1 is suggestive of binding constraints faced by poorer households in Zambia.

[17] In the rest of this chapter, I refer to the top 50% of households in terms of household expenditure as upper income households and the bottom 50% as lower income households.

Non-smoking households allocate, on average, 2 percentage points more to **schooling** than smoking households in all four sample types. The data on **Telephone** expenses and expenditure on **personal care** shows that non-smoking households allocate significantly more towards those expenditure categories than smoking households.

The other expenditure categories exhibit a mixed pattern. For instance, for urban and upper income samples, non-smoking households allocate more towards **housing** than smoking households. The differences are not significant for rural and lower income households. This is not a surprising finding as rural and lower income households are unlikely to pay explicit rentals as they mostly reside in their own houses. For **clothing**, non-smoking households in the rural and lower income samples allocate more expenditure towards this item. The difference, though, for lower income households is only significant at 10%. When it comes to **electricity**, urban and upper income non-smoking households allocate a bigger share than smoking households. Lower income and rural households with at least one smoker, on the other hand, allocate more towards **alternative energy** sources (kerosene and firewood) which are cheaper than electricity. It appears as though these households are trading-off an expensive energy source for a cheaper one given that tobacco is an additional expenditure item for them.

Non-smoking households allocate significantly more of their monthly budget towards **water** than smoking households. This is true in all data specifications except rural areas. This is not surprising given that rural households are more likely than not to obtain their water "free of charge" from streams and water wells. Water is, therefore, not an explicit expenditure item in rural areas. The data on **house care** shows statistically significant differences in favour of non-smoking households for urban and upper income households were spending on home improvements is likely to be important.

On the other hand, smoking households spend significantly more on **alcohol** than non-smoking households. In all four samples, smoking households allocate between 5 and 7 percentage points more towards alcohol than non-smoking households. Smoking households allocate more of their monthly expenditure towards **entertainment** in all four samples. The differences, however, are only statistically significant for rural households and lower income households. There are no statistically significant differences in **health** expenditures between the two types of households in all four samples.

In summary, the information in Table 2.2 shows that there are differences in the way that smoking and non-smoking households allocate their monthly expenditure for most of the items. In the rest of the chapter, I interrogate whether the patterns in Table 2.2 can be given a causal interpretation.

Table 2.2: Differences in expenditure shares between smoking and non-smoking households

Share on:	Full Sample	Urban	Rural	Top 50%	Bottom 50%
Food	-3.46***	-1.97***	-5.99***	-0.40	-7.62***
Alcohol	5.99***	5.19***	6.64***	5.10***	6.68***
Health	0.04	0.05	-0.08	-0.04	0.04
School	-2.15***	-2.30***	-1.88***	-2.19***	-1.69***
Clothing	0.07	-0.02	-1.29***	0.02	-0.72*
Housing	-1.43***	-1.70***	0.00	-2.35***	0.02
Water	-0.49***	-0.59***	-0.05	-0.69***	-0.14**
Electricity	-0.81***	-1.15***	-0.07	-1.26***	-0.10*
Alternative Energy	0.14	0.49**	-0.18	0.46***	-0.46*
Daily Transport	-0.54***	-0.32	-0.24	-0.43	-0.14
Other Transport	0.03	0.21	-0.08	0.36	0.05
Equipment Maintenance	0.16**	0.11	0.05	0.19*	0.10
Entertainment	0.03	0.12	0.12**	0.04	0.15***
Telephone	-0.82***	-0.66***	-0.27***	-0.91***	-0.13**
Remittances	-0.15**	-0.09	-0.13	-0.13	-0.04
House Care	-0.25***	-0.31***	0.05	-0.35***	0.05
Personal Care	-1.05***	-1.37***	-1.36***	-1.22***	-1.29***
Other	0.20**	0.50***	-0.07	0.48***	-0.11

Notes: The numbers reported in the table are differences in mean expenditure shares between smoking and non-smoking households expressed as percentage points. A positive percentage point difference implies that smoking households report a higher expenditure share than non-smoking households. The actual shares from which the differences in the table are calculated are reported in Appendix A as Table A1. *, **, *** implies that the percentage point difference is statistically significant at the 10%, 5% and 1% level respectively.

2.6.2. Econometric results

This section investigates whether the differences discussed in Section 2.6.1 can be given a causal interpretation. That is, are the expenditure share differences between smoking and non-smoking households *caused* by a household allocating some of its expenditure towards tobacco?

Ordinary Least Squares (OLS)

It is possible that the expenditure share patterns observed in Table 2.2 are the result of confounding variables. That is, there might be some characteristics of the household, other than tobacco smoking status, such as household structure or the household's socio-economic status that influences the results in Table 2.2. I can control for possible observable confounders by using OLS to regress expenditure shares on the household's smoking status and a number of control variables representing household structure and measures of a household's socio-economic status. In essence, this involves estimating equation (2.1) by OLS. I report the results of such an exercise in Table 2.3. (The table only reports the results of the coefficient on d in equation (2.1). The full set of results for the OLS estimation are contained in Tables A2 to A6 in Appendix A).

In Table 2.3, a negative coefficient on d implies that smoking households allocate less expenditure to the category in question when compared with non-smoking households after controlling for observables. The table largely reproduces the results in Table 2.2 at least in a qualitative sense. Food is given a smaller expenditure allocation by smoking households in all data specifications excluding upper income households (although the difference for urban households is only statistically significant at 10%). Even though the expenditure differences are now smaller in absolute terms, expenditure differences for food between smoking and non-smoking households are still largest for low income and rural households, a pattern noted in Table 2.2. Similarly, the coefficient on the schooling variable is negative in all specifications. Smoking households allocate, on average, between 1 and 2 percentage points less towards schooling than non-smoking households. The findings with regards to housing and water are similar to those in Table 2.2: smoking households in urban and upper income samples allocate a smaller share to the two expenditure categories than non-smoking households. The findings on electricity and personal care are also largely reproduced: smoking households allocate a smaller share to those two categories than non-smoking households.

Table 2.3: Ordinary least squares (OLS) estimates for the coefficient on *d*

Coefficient on *d* in:	Full Sample	Urban	Rural	Top 50%	Bottom 50%
Food	-2.100***	-1.226*	-2.614***	-0.399	-3.565***
	(0.479)	(0.638)	(0.708)	(0.608)	(0.742)
Alcohol	5.817***	4.841***	6.496***	5.026***	6.499***
	(0.163)	(0.219)	(0.242)	(0.211)	(0.251)
Health	-0.058	-0.021	-0.080	-0.128	-0.001
	(0.010)	(0.121)	(0.157)	(0.106)	(0.172)
School	-1.698***	-1.374***	-1.980***	-1.391***	-1.898***
	(0.259)	(0.374)	(0.364)	(0.389)	(0.343)
Clothing	-0.177	0.205	-0.496	-0.271	-0.124
	(0.254)	(0.274)	(0.415)	(0.269)	(0.435)
Housing	-0.683***	-1.583***	-0.062	-1.538***	0.107
	(0.208)	(0.431)	(0.065)	(0.388)	(0.122)
Water	-0.181**	-0.352**	-0.042	-0.365***	0.030
	(0.078)	(0.162)	(0.033)	(0.137)	(0.072)
Electricity	-0.329***	-0.673***	-0.039	-0.496***	-0.040
	(0.104)	(0.206)	(0.071)	(0.191)	(0.071)
Alt. Energy	0.125	0.203	0.064	-0.006	0.063
	(0.163)	(0.224)	(0.235)	(0.185)	(0.268)
Daily Tport	-0.133	-0.143	-0.175	-0.235	-0.172*
	(0.175)	(0.306)	(0.190)	(0.326)	(0.104)
Other Tport	0.065	0.492**	-0.206	0.293	-0.129
	(0.187)	(0.244)	(0.280)	(0.300)	(0.219)
Equipment	0.069	0.171*	-0.023	0.163	-0.055
	(0.082)	(0.102)	(0.127)	(0.123)	(0.109)
Entertain.	0.107	0.104	0.075	0.105	0.060
	(0.080)	(0.160)	(0.051)	(0.152)	(0.039)
Telephone	-0.144	0.015	-0.233*	0.004	-0.141*
	(0.136)	(0.253)	(0.127)	(0.253)	(0.086)
Remittance	-0.081	0.021	-0.153	-0.050	-0.109
	(0.076)	(0.114)	(0.102)	(0.127)	(0.080)
House Care	0.004	-0.135	0.098	-0.106	0.119**
	(0.065)	(0.122)	(0.060)	(0.120)	(0.048)
Personal	-0.807***	-1.032***	-0.617	-0.866***	-0.718*
	(0.264)	(0.339)	(0.402)	(0.312)	(0.431)
Sample size	**13,679**	**7,275**	**6,404**	**7,501**	**6,178**

Notes: The results shown above are only for the coefficient on d in equation (2.1). The full set of OLS results are contained in tables A2 to A6 in Appendix A. Standard errors are reported in parentheses. *, **, *** implies that the coefficient on d is statistically significant at the 10%, 5% and 1% significance respectively. Tport is an abbreviation for transport. Alt is an abbreviation for alternative, entertain is short for entertainment and personal is short for personal care.

The coefficients on d in the alcohol equation in Table 2.3 also show the same pattern as in Table 2.2, namely smoking households allocate a larger expenditure share to alcohol than non-smoking households even after controlling for other observable variables that might confound the analysis. The magnitudes of the expenditure differences for alcohol are similar to those in Table 2.2.

For most of the remaining expenditure categories, the patterns in Table 2.2 do not hold out after controlling for observable confounders. For instance, whereas Table 2.2 shows statistically significant differences in the allocation of expenditure shares for telephone expenses in all 4 samples, Table 2.3 shows that the differences are only statistically significant in 2 subsamples (rural and low income households) and only at the 10% level.

Three-stage least squares (3SLS)

As highlighted in Section 2.4, it is likely that d in equation (2.1) is endogenous. For example, a household might only decide to spend on tobacco after other household expenses have been made. Alternatively, some other variable, not explicitly specified in equation (2.1) and therefore contained in the error term, might simultaneously influence a household's decision to spend on tobacco and the decision to spend on other commodities. The OLS procedure in the previous section assumes that d is exogenous, i.e. d and the error term are not correlated. Further, $\ln M$ and $(\ln M)^2$ are also likely to be endogenous in a similar manner. In Table A14 in Appendix A, I report the chi-square statistics associated with a Durbin-Wu-Hausman test for exogeneity. The null hypothesis that the three variables are exogenous is rejected for most of the expenditure categories. Previous work (Vermeulen, 2003; John, 2008b; Pu *et al.*, 2008) has also shown that the decision to spend on tobacco and, as a consequence, the residual expenditure $\ln M$ [and $(\ln M)^2$] are endogenous.

This section, therefore, uses the method of instrumental variables to account for the endogeneity of d, $\ln M$ and $(\ln M)^2$ in equation (2.1). The method proceeds in a series of steps: In the first step, I estimate the first-stage (or reduced form) regressions involving the endogenous variables and the candidate instruments. Recall that I use the adult sex ratio as an instrument for d and the logarithm of the value of assets within a household and its square as instruments for $\ln M$ and $(\ln M)^2$ respectively. In the second stage, the predicted values from the first stage regression are substituted for the endogenous variables in an OLS estimation of equation (2.1).

The first-stage probit results for the regression of d on the adult sex ratio (the instrumental variable) alongside the other covariates for all four samples are reported in Table A7 in Appendix A. In all the five samples, the instrumental variable is a strong predictor of whether a household reports spending on tobacco or not after controlling for other variables. The F statistics associated with the coefficient on this variable are equal to 78, 50, 27, 26 and 44 in the full, urban, rural, upper and lower income samples respectively. These F statistics satisfy the standard rule of thumb that an instrument be considered "strong" if its associated F statistic is equal to or greater than 10 (Stock, *et al.*, 2002). Tables A8 and A9 in Appendix A repeat this exercise for $\ln M$ and $(\ln M)^2$ using the logarithm of household assets and its square, respectively, as instruments. I find that the two instruments explain a substantial proportion of the variation in $\ln M$ and $(\ln M)^2$. The F statistics associated with the coefficient on the instrumental variable for lnM (the logarithm of the value of household assets) are equal to 3364, 1971, 1127, 1444 and 560 in the full, urban, rural, upper and lower income samples respectively. For $(\ln M)^2$, the corresponding F statistics are 3136, 1866, 1444, 2162 and 697.

Table 2.4: Three stage least squares (3LS) estimates for the coefficient on *d*

Coefficient on *d* in:	Full Sample	Urban	Rural	Top 50%	Bottom 50%
Food	-5.076***	-0.728	-7.086***	-4.275*	-6.315***
	(1.634)	(2.543)	(2.250)	(2.504)	(2.234)
Alcohol	9.696***	8.656***	10.060***	8.672***	8.989***
	(0.668)	(1.019)	(0.928)	(1.068)	(0.956)
Health	0.012	0.267	-0.064	0.769*	-0.113
	(0.337)	(0.434)	(0.507)	(0.447)	(0.505)
School	-1.975**	-2.925*	-1.662	-0.735	-2.096**
	(0.863)	(1.523)	(1.084)	(1.608)	(0.935)
Clothing	-2.665***	-2.751**	-2.836**	0.661	-3.174**
	(0.890)	(1.194)	(1.328)	(1.143)	(1.411)
Housing	-0.366	-1.829	0.134	-2.760*	0.571**
	(0.675)	(1.666)	(0.177)	(1.611)	(0.269)
Water	-0.795***	-2.228***	-0.119	-2.187***	-0.095
	(0.273)	(0.679)	(0.111)	(0.600)	(0.209)
Electricity	-0.479	-2.164***	0.132	-1.684*	0.114
	(0.341)	(0.813)	(0.221)	(0.874)	(0.083)
Alt. Energy	1.788***	1.251	2.205***	-3.045***	2.204***
	(0.554)	(0.943)	(0.733)	(0.843)	(0.774)
Daily Tport	0.177	0.256	0.305	0.388	-0.325
	(0.586)	(1.264)	(0.563)	(1.363)	(0.227)
Other Tport	-0.661	-0.249	-0.761	0.191	1.504**
	(0.652)	(1.032)	(0.884)	(1.269)	(0.635)
Equipt.	0.405	1.038**	0.195	1.949***	-1.136***
	(0.294)	(0.449)	(0.423)	(0.520)	(0.373)
Entertain.	0.398	0.817	0.220	0.451	0.174
	(0.260)	(0.620)	(0.157)	(0.636)	(0.134)
Telephone	0.687	2.099*	0.061	3.861***	-0.218
	(0.487)	(1.111)	(0.410)	(1.199)	(0.338)
Remittance	0.627**	1.490***	0.279	2.050***	0.196
	(0.261)	(0.497)	(0.303)	(0.573)	(0.226)
House Care	0.213	-0.417	0.390**	-1.114**	0.464***
	(0.222)	(0.492)	(0.192)	(0.522)	(0.155)
Personal	-1.566*	-1.788	-1.116	-3.148**	-0.518
	(0.872)	(1.310)	(1.227)	(1.246)	(1.318)
Sample size	**8,555**	**4,545**	**4,010**	**4,092**	**3,212**

Notes: The results shown above are only for the coefficient on d in equation (2.1). The full set of 3LS results are contained in tables A10 to A13 Appendix A. Standard errors are reported in parentheses. *, **, *** signifies that the coefficient on d is statistically significant at the 10%, 5% and 1% levels respectively. Tport is an abbreviation for transport. Alt is an abbreviation for alternative, entertain is short for entertainment and personal is short for personal care.

Table 2.4 presents the results of the three-stage least squares (3SLS) implementation of equation (2.1). The table only reports estimates of the coefficient on d (the full set of the 3LS results are reported in tables A9 to A12 in Appendix A). A negative coefficient on d implies that smoking households allocate a smaller percentage of household expenditure to that particular category after accounting for observable confounders and the possible endogeneity of d, $\ln M$ and $(\ln M)^2$.

The 3SLS procedure confirms the OLS estimates for food, namely that smoking households spend a smaller proportion of their expenditure on this expenditure category. The results are significant for rural and lower income samples at 1% and marginally so for upper income households (at 10%) but not for urban households. The 3SLS estimates for food are generally larger in absolute terms implying that OLS underestimates the difference in expenditure allocations for food between the two types of households.

The 3SLS estimates in the **schooling** equation are in general larger than the OLS estimates. For urban and lower income samples, smoking households allocate significantly less expenditure to schooling than non-smoking households. For urban households, the difference is only statistically significant at the 10% level. In the OLS regression, the differences in expenditure allocation between the two types of households were statistically significant in all four samples. For **water**, the qualitative pattern of the coefficient estimates in Table 2.4 is similar to that in Table 2.3, namely that smoking households allocate less expenditure than non-smoking households. Further, the 3SLS estimates are larger in absolute terms.

The 3SLS coefficient estimates for clothing, housing and alternative energy tell a different story to the OLS estimates presented in Table 2.3. In Table 2.3, the expenditure allocations to **clothing** were not significantly different between smoking and non-smoking households. In Table 2.4, the 3SLS estimates show that non-smoking households allocate significantly more expenditure to clothing. Among upper income households, share allocations between smoking and non-smoking households for clothing are not statistically significant. For **housing**, significant differences in expenditure allocations are only observed for upper income and lower income households using the 3SLS procedure: for upper income households, non-smoking households allocate significantly more and the reverse is true for lower income households. With **alternative energy**, significant differences in expenditure allocations exist for three of the four samples. For rural and lower income samples, smoking households allocate more to alternative energy than non-smoking households whereas the reverse is true for richer households. This again suggests some sort of substitution where lower income smokers substitute cheaper energy sources for more expensive energy sources. A pattern noted in Table 2.2.

The qualitative patterns with **alcohol** in Tables 2.2 and 2.3 are reproduced in Table 2.4. Smoking households allocate more of their expenditure towards alcohol than non-smoking households. The 3SLS estimates are bigger than the OLS estimates. Other categories that result in smoking households allocating a larger share of their expenditure are **telephone** and **remittances**. For the two categories, the 3SLS estimates are, in almost all cases, the direct opposite of the OLS estimates. For urban and upper income households, smoking households allocate a significantly larger share of expenditure towards the two categories than non-smoking households.

The discussion of the results given above holds even when the OLS sample is restricted to the sample used under the 3SLS procedure (see Table A16 in Appendix A).

2.7. Relaxing the exclusion restriction

The previous sub-section utilised the method of instrumental variables (IV) to identify the causal impact of a household's tobacco smoking status on its expenditure decisions. One important requirement underpinning the IV procedure is that the instrumental variable should not be correlated with the error term in equation (2.1).[18] In other words, the instrument should only influence the outcome through its influence on the endogenous variable. Unfortunately, it is difficult to ascertain whether this requirement, referred to as the exclusion restriction in the IV literature, holds especially in the just-identified case such as I have in this chapter.[19] In the just-identified case, the researcher

[18] Another requirement is that the instrumental variable be relevant (or informative). That is, the instrumental variable should explain a substantial proportion of the variation in the endogenous regressor. I believe that the instrumental variables in this chapter satisfy this requirement based on the results in tables A6 to A8 in Appendix A and in the discussion in the previous section.

[19] In the just-identified case, the number of instruments is equal to the number of endogenous variables. In the alternative case where the number of instruments is greater than the number of endogenous variables, one can perform a test of over-identifying restrictions, such as the Sargan test, to check whether the exclusion restriction holds.

needs to motivate that the exclusion restriction holds. Whereas I am confident that the total value of household assets, the instrument for the residual expenditure, satisfies the exclusion restriction, I am less so about the adult sex ratio. As highlighted in Section 2.4, demand equations are rarely specified, if ever, with the logarithm of household assets as an explanatory variable. In other words, the logarithm of household assets is often excluded from demand equations.

In the same way the sex ratio influences expenditure on tobacco, it might also influence expenditure on other goods and services within the household. It is known that women have a preference for spending on welfare enhancing goods and services like food, schooling and healthcare (Senauer *et al.*, 1988; Deaton, *et al.*, 1989; Thomas, 1990, 1994; Hoddinott and Haddad, 1995; Lundberg *et al.*, 1997; Duflo, 2003; Gitter and Barham, 2008; Doss, 2013). A preponderance of women over men might, therefore, tilt household expenditure towards these goods and services. Deaton (1997) goes over much of the empirical and theoretical literature that explains and documents the existence of gender biases in the allocation of resources within the household. Excluding the adult sex ratio from structural demand equations might, therefore, result in biased coefficient estimates.

This section of the chapter uses the method introduced by Nevo and Rosen (2012) to test the robustness of my 3SLS estimates by allowing the adult sex ratio to be correlated with the error term. In other words, Nevo and Rosen's method allows for the instrumental variable to be imperfect and thus imposes a less restrictive assumption than the standard IV approach.

The method in Nevo and Rosen (2012) relies on two assumptions. The first assumption is that the direction of the correlation between the imperfect instrumental variable and the error term should be the same as that between the error term and the endogenous variable.[20] I am confident that this assumption holds because my instrumental variable is positively correlated with the endogenous variable (see the results in Table A6 in Appendix A). Therefore, the direction of correlation with the error term must be the same for the two variables. In other words, if the instrumental variable and the endogenous regressor are positively correlated, as they are in this chapter, and if, say, the endogenous regressor is positively correlated with the error term, then the instrumental variable must

[20] Formally, letting Z denote the imperfect instrumental variable, this assumption can be represented as: $Corr\ (Z, u)\ Corr\ (d, u)\ \geq 0$ (i.e. the product of the two correlations should be positive).

also be positively correlated with the error term. The second assumption requires that the magnitude of the correlation between the error term and the endogenous variable should, in absolute terms, be greater than the magnitude of the correlation between the imperfect instrument and the error term.[21] In other words, this assumption requires the imperfect instrumental variable to be "less endogenous" than the endogenous regressor. This assumption likely holds for the following reasons: One of the variables that I suspect to be in the error term is the rate of time preference as measured by the discount rate. Field and laboratory experiments show that smokers have higher discount rates than non-smokers (Chabris *et al.*, 2008; Harrison *et al.*, 2010). In addition, there is evidence that the discount rate influences expenditure decisions on some of the goods and services on the left-hand side (Fersterer and Winter-Ebmer, 2003; Chabris *et al.*, 2008). On the other hand, there appears to be no significant differences in discount rates between men and women (Harrison *et al.*, 2002; Harrison *et al.*, 2005; Andersen, *et al.*, 2010). To the extent that one can define a household-level discount rate, consistent with my assumption of a unitary model of household decision making, then the foregoing suggests that the correlation between d and u is likely to be at least greater than the correlation between the adult sex ratio and u. One possible challenge to this assumption is that not all goods and services on the left-hand side are influenced by the discount rate. For instance, how might the discount rate, *ceteris paribus,* influence a household's expenditure on water? Whereas there might be some goods and services where the relationship is not obvious, I believe that the discount rate is important in most of the goods and services that I study (at least those that are generally considered important such as food and schooling) to give credence to this assumption.

Nevo and Rosen (2012) propose a synthetic instrumental variable, V, formally defined as $V(\lambda) = \sigma_d Z - \lambda\sigma_z d$, where Z is the imperfect instrument and d is the endogenous regressor; σ_d and σ_z are respectively the standard deviations of d and Z; and λ is some parameter defined as the ratio of the correlation between Z and u to the correlation between d and u.[22] For λ^*, $V(\lambda^*)$ satisfies the exclusion restriction allowing us to consistently identify the causal impact of d on households' spending decisions. I do not, however, observe λ^* but given the assumptions above, I know that it must lie in

[21] Formally, and using the notation in footnote 20, this assumption requires: $|\,Corr\,(d,u)\,| \geq |\,Corr\,(Z,u)|$.

[22] That is, $\lambda = \frac{Corr(Z,u)}{Corr\,(d,u)}$.

the [0,1] interval.[23] By varying λ in the [0,1] interval, I can construct bounds within which the true causal impact lies given the presence of endogeneity and an imperfect instrumental variable.[24]

Nevo and Rosen further show that one can only compute one-sided bounds if the empirical relationship (correlation) between the imperfect instrumental variable and endogenous regressor is positive such as I have here (this relationship/correlation is positive given the results in Table A6 in Appendix A) . Otherwise one can obtain two-sided bounds. The choice of bound (whether it is the causal impact associated with Z or with V) and whether the bound is an upper or lower bound depends on the assumptions made about the direction of correlation between the error term and the endogenous regressor[25] (I motivate below why I think this correlation is positive). The method of constructing bounds in this way has been used to shed light on empirical debates around the causal impact of teen pregnancy on school completion rates (Reinhold and Woutersen, 2011) and on whether there is a causal relationship between financial hardship and obesity (Averett and Smith, 2014). Kortelainen and Saarimaa (2015) have also employed the method in studying whether increasing homeownership rates benefits urban neighbourhoods. Recently, Aragon and Rud (2015) have used the method to study the relationship between industrial pollution and agricultural productivity in Ghana.

As stated in the previous paragraph, whether the bound is a lower or upper bound depends on the assumption made about the correlation between d and the error term, u. I assume that this correlation is positive in all equations in my demand system. This assumption follows quite naturally from the discussion above on the experimental evidence on the relationship between smoking and the rate of time preference (the discount rate). This assumption implies upper-bound estimates for the causal impact of d on households' spending decisions and by consequence a lower bound of -∞ (see footnote 23). Table 2.5, therefore, reports upper bound estimates for the causal impact of d in

[23] Since $\lambda = \frac{Corr(Z,u)}{Corr\ (d,u)}$, then for $Corr(Z,u) = 0,\ \lambda = 0$. For $Corr(Z,u) = Corr(d,u), \lambda = 1$.

24 To see this, note that if $\lambda = 0$, V is equal to the imperfect instrument Z multiplied by a scalar. Whereas if $\lambda = 1$, $V = \sigma_d\, Z - \sigma_z d$. The fact that V varies in this way results in the construction of a set within which the true causal impact must lie.

25 Formally, proposition 2 in Nevo and Rosen (2012) states: if $Corr\ (d,u) < 0$, then $\beta\ \geq max\{\beta_Z, \beta_V\}$ and if $Corr\ (d,u) > 0$, then $\beta\ \leq min\{\beta_Z, \beta_V\}$. β is the true effect, β_Z is the coefficient associated with the imperfect instrument and β_V is the coefficient associated with the synthetic instrument evaluated at $\lambda = 1$.

all the four data specifications that I consider. For the purposes of making inference, bounds that do not overlap zero imply that the underlying causal effect is statistically significant. [26] Such bounds are in bold in Table 2.5.

After relaxing the exclusion restriction, I end up with the finding that in no expenditure category do tobacco smoking households allocate a greater share of total household expenditure. This is also true for alcohol which in Tables 2.2 to 2.4 showed a strong positive association with the tobacco smoking status of a household. This association is more likely a correlation than a causal one.

By and large, most of the qualitative findings in Table 2.4 hold up with some differences. For instance, smoking households allocate a smaller share of expenditure to food in only rural and lower income households. This is suggestive of the fact that budget constraints are more likely to be binding for low income households given the geographic nature of poverty documented in Table 2.1. With regards to schooling, in only upper income and lower income samples do smoking households allocate a smaller and statistically significant share. As for clothing, the results are in line with those in Table 2.4, namely smoking households significantly allocate less expenditure in all samples except for upper income households. The qualitative results for water are also reproduced: smoking households allocate a smaller share to water for urban and upper income samples. The result with regards to water is confirmation of the pattern noted in Section 2.6.1, namely that water is likely obtained "free of charge" for households residing in rural and poorer parts of the country. In this way, water is not purchased formally in a market place and is, therefore, not susceptible to being crowded out or crowded in for these types of households.

For electricity and alternative energy, the differences are only statistically significant in one of the data subsamples: urban households where electricity is crowded out for smoking households and upper income households where alternative energy is crowded out. In either case, the differences were statistically significant in more than one subsample with the 3SLS procedure.

[26] Technically, the upper bounds in Table 2.5 are the bootstrapped upper bounds of a 95% confidence interval around either β_Z or β_V. This is a much more stringent approach to making inferences than just looking at β_Z or β_V [Reinhold and Woutersen (2011) and Kortelainen and Saarimaa (2015) use a similar procedure for making inferences].

Table 2.5: Estimates of bounds for the causal effect of *d*

Category	Full Sample	Urban	Rural	Top 50%	Bottom 50%
Food	**(-∞, -1.958)**	(-∞, 0.305)	**(-∞, -2.677)**	(-∞, 0.633)	**(-∞, -0.194)**
Alcohol	(-∞, 5.389)	(-∞, 4.435)	(-∞, 6.280)	(-∞, 4.50)	(-∞, 6.472)
Health	(-∞, 0.263)	(-∞, 0.043)	(-∞, 0.928)	**(-∞, -0.037)**	(-∞, 0.877)
School	**(-∞, -0.284)**	(-∞, 0.060)	(-∞, 0.462)	**(-∞, -0.010)**	**(-∞, -0.263)**
Clothing	**(-∞, -0.920)**	**(-∞, -0.412)**	**(-∞, -0.233)**	(-∞, 0.256)	**(-∞, -0.408)**
Housing	(-∞, 0.958)	(-∞, 1.437)	(-∞, 0.263)	(-∞, 0.398)	(-∞, 0.444)
Water	**(-∞, -0.261)**	**(-∞, -0.898)**	(-∞, 0.098)	**(-∞, -1.010)**	(-∞, 0.314)
Electricity	(-∞ , 0.189)	**(-∞, -0.571**	(-∞, 0.221)	(-∞, 0.028)	(-∞, 0.186)
Alternative	(-∞, 0.610)	(-∞, 1.044)	(-∞, 0.424)	**(-∞, -1.393)**	(-∞, 0.621)
Daily Transport	**(-∞,-0.023)**	**(-∞, -0.262)**	(-∞, 0.338)	**(-∞, -0.277)**	(-∞, 0.120)
Other Transport	(-∞, 0.616)	(-∞, 1.177)	(-∞, 0.972)	(-∞, 2.679)	(-∞, 0.430)
Equipt Maintenance	**(-∞, -0.125)**	(-∞, 0.198)	**(-∞, -0.255)**	(-∞, 0.145)	**(-∞, -0.405)**
Entertain.	(-∞, 0.026)	**(-∞, -0.089)**	(-∞, 0.185)	**(-∞, -0.022)**	(-∞, 0.124)
Telephone	(-∞, 0.070)	(-∞, 0.518)	(-∞, 0.160)	(-∞, 0.584)	(-∞, 0.443)
Remittance	**(-∞, -0.092)**	(-∞, 0.068)	**(-∞, -0.050)**	**(-∞, -0.036)**	(-∞, 0.078)
House Care	(-∞, 0.303)	(-∞, 0.547)	(-∞, 0.264)	**(-∞, -0.090)**	(-∞, 0.252)
Personal Care	(-∞, 0.142)	(-∞, 0.780)	(-∞, 1.290)	**(-∞, -0.707)**	(-∞, 2.066)
Sample size	**8,555**	**4,545**	**4,010**	**4,092**	**3,212**

Notes: The table shows estimates of bounds for the causal effect of d using the method in Nevo and Rosen (2012). Bounds in bold imply that the causal effect of d is statistically significant since such bounds do not overlap zero.

The findings that are entirely different to those in Table 2.4 are for daily transportation, equipment maintenance, entertainment and remittances. For daily transportation, smoking households allocate less expenditure than non-smoking households in only urban and upper income subsamples were this expenditure item is likely important. For the rest, the findings in Table 2.5 point in the opposite direction to those reported in Table 2.4, namely smoking households allocate significantly less expenditure. For equipment maintenance, this pattern is observed for rural andlower income samples; for entertainment it is in the urban and upper income households and for remittances it is for rural and upper income households.

2.8. Summary and conclusion

This chapter adds to the literature on the crowding out effect of tobacco expenditure in two ways. In the first instance I use data from a low income sub-Saharan African country, Zambia, where most of the households are poor. Secondly, in identifying the causal impact of tobacco expenditure, I use the method of instrumental variables, which is the standard method in the literature, but instead use less stringent assumptions on the behaviour of the instrument. My econometric analysis shows that tobacco expenditure negatively impacts household expenditure on food, schooling, clothing and water. Other expenditure categories that are negatively impacted include transportation, equipment repair, entertainment and remittances. My analysis shows that the patterns and the magnitudes of crowding out are in some instances related to the household's geographical location and/or socio-economic status. For instance, food is more likely to be displaced by tobacco in lower income households than upper income households. I, therefore, confirm many of the findings in the literature. On the other hand, I am unable to find instances where tobacco crowds in expenditure on alcohol. The econometric analysis in this chapter, supported by what I consider to be plausible assumptions on the instrumental variable, leads to the conclusion that the positive association between tobacco and alcohol is more likely to be a correlational one than a causal one.

This chapter's main limitation is that it uses cross-sectional data where, with an exhaustive list of controls, one can compare the expenditure profiles of two households that are identical in every respect except for tobacco smoking status. Unfortunately, there are unmeasurable sources of heterogeneity between the two types of households that cannot be accounted for making it difficult to draw definitive causal statements. Panel datasets are ideal in the sense that they allow for one to compare the expenditure profile of the same household at different points in time and in this way control for fixed unobserved heterogeneity. Unfortunately, Zambia does not as yet collect panel data. In any case, using instrumental variables, along with less stringent assumptions, can substitute for some advantages of panel data.

The work in this chapter shows that a broader accounting of tobacco's consumption costs in Zambia should, for example, include the costs associated with under nutrition and under investment in education and in equipment maintenance by households. As pointed out by Wang *et al.* (2006), Block and Webb (2009) and John *et al.* (2012), under nutrition, which can affect the cognitive development of children, under investment in education and underinvestment in the maintenance of income-generating equipment are likely to trap households in a cycle of poverty.

APPENDIX A

Table A 1: Mean expenditure shares in percentages between smoking and non-smoking households

Share on:	Full Sample	Urban	Rural	Top 50%	Bottom 50%
Food	(50.67, 47.21)	(47.29, 45.32)	(54.41, 48.42)	(44.56, 44.13)	(57.05, 49.43)
Alcohol	(1.35, 7.34)	(1.43, 6.61)	(1.26, 7.80)	(1.52, 6.62)	(1.17, 7.86)
Health	(1.03, 1.08)	(0.78, 0.82)	(1.32, 1.24)	(0.83,0.79)	(1.24, 1.29)
School	(5.93, 3.78)	(6.30, 4.00)	(5.51, 3.63)	(7.29, 5.10)	(4.51, 2.83)
Clothing	(9.19, 9.26)	(6.17, 6.15)	(12.52, 11.22)	(6.65, 6.67)	(11.84, 11.12)
Housing	(3.15, 1.72)	(5.83, 4.12)	(0.20, 0.20)	(5.60, 3.25)	(0.60, 0.62)
Water	(1.03, 0.54)	(1.83, 1.25)	(0.14, 0.10)	(1.66, 0.97)	(0.38, 0.24)
Electricity	(1.38, 0.57)	(2.49, 1.34)	(0.16, 0.08)	(2.59, 1.32)	(0.13, 0.03)
Alternative	(4.34, 4.48)	(4.14, 4.62)	(4.57, 4.39)	(3.24, 3.70)	(5.50, 5.04)
Daily Transport	(1.84, 1.30)	(2.77, 2.46)	(0.81, 0.57)	(3.33, 2.90)	(2.85, 0.15)
Other Transport	(1.95, 1.97)	(1.97, 2.18)	(1.91, 1.84)	(2.76, 3.12)	(1.10, 1.15)
Equipt Maintenance	(0.71, 0.86)	(0.42, 0.53)	(1.02, 1.07)	(0.61, 0.80)	(0.81, 0.91)
Entertainment	(0.50, 0.53)	(0.84, 0.97)	(0.13, 0.26)	(0.93, 0.97)	(0.06, 0.21)
Telephone	(2.23, 1.41)	(3.62, 2.96)	(0.70, 0.43)	(4.15, 3.24)	(0.23, 0.10)
Remittance	(0.71, 0.56)	(0.86, 0.77)	(0.55, 0.42)	(1.11, 0.97)	(0.30, 0.26)
House Care	(0.89, 0.64)	(1.45, 1.14)	(0.27, 0.33)	(1.58, 1.24)	(0.17, 0.21)
Personal Care	(11.02, 9.97)	(9.91, 8.54)	(12.24, 10.87)	(9.91, 8.69)	(12.17, 10.88)
Other	(2.07, 2.27)	(1.89, 2.40)	(2.26, 2.19)	(1.70, 2.18)	(2.45, 2.34)

Notes: The pairs in each cell refer to the mean expenditure shares in percentages reported by smoking and non-smoking households. In each pair, the mean expenditure share for non-smoking households is reported first followed by that of the smoking households. The figures in this table are used to construct the percentage point differences reported in Table 2.2.

Table A 2: Ordinary Least Squares results, Full Sample

Variable	Food	Alc	Health	Sch	Clothing	Housing	Water	Electricity	Alt. Energ	Daily Tprt	Other Tprt	Equpt	Ent	Tele	Remit	Hse Care	Pers. Cre
d	-2.100***	5.817***	-0.0577	-1.698***	-0.177	-0.683***	-0.181**	-0.329***	0.125	-0.133	0.0649	0.0685	0.107	-0.144	-0.0811	0.00449	-0.807***
	(0.479)	(0.163)	(0.0999)	(0.259)	(0.254)	(0.208)	(0.0781)	(0.104)	(0.163)	(0.175)	(0.187)	(0.0823)	(0.0799)	(0.136)	(0.0756)	(0.0653)	(0.264)
LnM	45.20***	0.614	-1.082***	-4.690***	-11.31***	-3.786***	0.319	-0.546	4.565***	-17.07***	1.639***	-2.767***	-4.557***	-4.033***	-1.356***	-2.384***	-1.027
	(1.564)	(0.533)	(0.326)	(0.847)	(0.828)	(0.679)	(0.255)	(0.341)	(0.532)	(0.572)	(0.610)	(0.269)	(0.261)	(0.445)	(0.247)	(0.213)	(0.863)
LnM2	-2.000***	-0.00970	0.0431***	0.212***	0.391***	0.196***	-0.0113	0.0334**	-0.239***	0.781***	-0.00983	0.123***	0.204***	0.213***	0.0726***	0.115***	-0.00760
	(0.0631)	(0.0215)	(0.0132)	(0.0342)	(0.0334)	(0.0274)	(0.0103)	(0.0138)	(0.0215)	(0.0231)	(0.0246)	(0.0109)	(0.0105)	(0.0180)	(0.00996)	(0.00861)	(0.0348)
Adult prop	-1.082	0.900**	0.284	0.670	0.000281	-0.901*	0.606***	0.970***	-0.442	0.558	-1.090**	0.388*	-0.200	0.437	-0.781***	0.120	-0.780
	(1.222)	(0.417)	(0.255)	(0.662)	(0.648)	(0.531)	(0.199)	(0.267)	(0.416)	(0.447)	(0.477)	(0.210)	(0.204)	(0.348)	(0.193)	(0.167)	(0.674)
loghhsize	-1.642***	-0.159	0.186*	3.715***	1.146***	-1.997***	0.0731	0.259**	0.100	-0.0840	-0.874***	0.525***	-0.262***	-0.123	-0.365***	-0.302***	-0.260
	(0.506)	(0.173)	(0.106)	(0.274)	(0.268)	(0.220)	(0.0826)	(0.110)	(0.172)	(0.185)	(0.198)	(0.0870)	(0.0844)	(0.144)	(0.0799)	(0.0690)	(0.279)
Head Sch	-0.251	-0.219	-0.203*	0.855***	-0.00185	0.132	0.159**	0.322***	-0.292*	-0.201	-0.246	0.143*	-0.0465	0.278**	-0.0629	0.0184	-0.357
	(0.497)	(0.169)	(0.104)	(0.269)	(0.263)	(0.216)	(0.0811)	(0.109)	(0.169)	(0.182)	(0.194)	(0.0855)	(0.0830)	(0.142)	(0.0785)	(0.0678)	(0.274)
Head w age	-0.792**	0.170	-0.246***	0.225	-0.406**	1.834***	-0.145**	0.120	-0.480***	-0.887***	-0.603***	-0.513***	0.0866	0.659***	0.182***	0.0790	1.068***
	(0.355)	(0.121)	(0.0741)	(0.192)	(0.188)	(0.154)	(0.0579)	(0.0775)	(0.121)	(0.130)	(0.139)	(0.0610)	(0.0592)	(0.101)	(0.0560)	(0.0484)	(0.196)
Hhold Emp	-0.00192	0.137**	0.0247	-0.875***	0.246***	-0.0911	-0.0683**	-0.184***	0.0542	0.135**	0.0404	0.0330	0.0540*	0.0951*	0.116***	0.0542**	0.170*
	(0.172)	(0.0587)	(0.0359)	(0.0932)	(0.0912)	(0.0748)	(0.0281)	(0.0376)	(0.0586)	(0.0630)	(0.0672)	(0.0296)	(0.0287)	(0.0490)	(0.0272)	(0.0235)	(0.0949)
Adult Age	2.805**	0.719*	0.932***	-8.238***	0.457	0.905*	-0.120	-0.438	0.657	0.952**	-0.0371	0.548**	0.283	0.459	0.0999	0.359**	-0.587
	(1.253)	(0.427)	(0.262)	(0.679)	(0.664)	(0.544)	(0.204)	(0.273)	(0.427)	(0.458)	(0.489)	(0.215)	(0.209)	(0.357)	(0.198)	(0.171)	(0.691)
Child age	-0.493**	-0.184**	0.00162	2.413***	-0.474***	-0.380***	0.0336	0.193***	-0.288***	-0.0974	-0.123	-0.0359	-0.0125	0.0553	-0.0989***	0.0359	-0.385***
	(0.227)	(0.0774)	(0.0474)	(0.123)	(0.120)	(0.0987)	(0.0371)	(0.0496)	(0.0773)	(0.0831)	(0.0886)	(0.0390)	(0.0379)	(0.0647)	(0.0358)	(0.0310)	(0.125)
Head Age	-2.294**	-1.145***	-0.454*	10.08***	-2.683***	-2.175***	0.641***	0.888***	-0.182	-0.799*	0.651	-0.599***	-0.323*	-0.823**	-0.145	-0.253*	-0.0499
	(1.123)	(0.383)	(0.235)	(0.609)	(0.595)	(0.488)	(0.183)	(0.245)	(0.382)	(0.411)	(0.438)	(0.193)	(0.187)	(0.320)	(0.177)	(0.153)	(0.620)
Most Edu	-2.080***	-0.207	-0.0995	0.463	0.639*	-0.195	0.0938	0.453***	-0.271	-0.431*	-0.127	-0.131	0.00717	0.505***	0.312***	0.181**	0.731**
	(0.640)	(0.218)	(0.134)	(0.347)	(0.339)	(0.278)	(0.104)	(0.140)	(0.218)	(0.234)	(0.250)	(0.110)	(0.107)	(0.182)	(0.101)	(0.0873)	(0.353)
2.stratum	-3.033***	-0.801***	0.105	1.503***	0.985***	0.119	-0.233**	-0.348**	-0.139	1.236***	0.0573	0.519***	-0.192*	0.214	0.143	0.129	-0.404
	(0.642)	(0.219)	(0.134)	(0.348)	(0.340)	(0.279)	(0.105)	(0.140)	(0.218)	(0.235)	(0.250)	(0.110)	(0.107)	(0.183)	(0.101)	(0.0875)	(0.354)
3.stratum	-3.471	-2.146**	-0.394	0.384	-1.739	0.254	-0.434	-0.497	0.906	5.951***	-3.688***	2.389***	-0.600	3.409***	1.554***	0.913**	-2.859*
	(2.861)	(0.975)	(0.597)	(1.550)	(1.515)	(1.243)	(0.467)	(0.624)	(0.974)	(1.046)	(1.116)	(0.492)	(0.477)	(0.815)	(0.452)	(0.390)	(1.578)
4.stratum	6.328***	-0.778***	0.340**	-0.901**	-2.239***	-0.889***	0.107	0.0499	0.245	0.0818	0.411	-0.562***	-0.120	-0.0261	-0.173*	-0.00383	-1.655***
	(0.655)	(0.223)	(0.137)	(0.355)	(0.347)	(0.285)	(0.107)	(0.143)	(0.223)	(0.240)	(0.256)	(0.113)	(0.109)	(0.187)	(0.103)	(0.0893)	(0.361)
5.stratum	1.121*	0.425**	0.180	-0.702**	-4.297***	0.519**	-1.365***	-1.231***	2.425***	1.324***	1.086***	-0.176*	0.0489	0.0265	-0.420***	-1.149***	-0.924***
	(0.602)	(0.205)	(0.126)	(0.326)	(0.211)	(0.262)	(0.0829)	(0.131)	(0.136)	(0.220)	(0.235)	(0.104)	(0.0847)	(0.172)	(0.0630)	(0.0822)	(0.332)
6.stratum	0.658	0.0332	0.157	0.212	-4.223***	0.691**	Omitted	-0.192	1.558***	0.169	0.823***	-0.282**		-0.0441	-0.722***	-1.073***	-0.266
	(0.677)	(0.231)	(0.141)	(0.367)	(0.312)	(0.294)		(0.148)	(0.201)	(0.248)	(0.264)	(0.116)		(0.193)	(0.0931)	(0.0923)	(0.373)
7.stratum	Omitted	Omitted	Omitted	Omitted	-4.400***	Omitted	-0.272**	Omitted	1.898***	Omitted	Omitted	Omitted	0.815***	Omitted	-0.666***	Omitted	Omitted
					(0.366)		(0.110)		(0.236)				(0.113)		(0.109)		
Cluster	0.000732	-0.000280	0.000107	0.00184***	0.000261	0.000119	6.70e-05	-0.000537***	-0.00126***	0.000139	-0.000373*	-0.000322***	0.000135	0.000482***	0.000127	-0.000189***	-0.00141***
	(0.000501)	(0.000171)	(0.000105)	(0.000271)	(0.000265)	(0.000217)	(8.17e-05)	(0.000109)	(0.000170)	(0.000183)	(0.000195)	(8.61e-05)	(8.35e-05)	(0.000143)	(7.90e-05)	(6.83e-05)	(0.000276)
2.region	2.795***	-1.155***	-0.344**	-0.986***	Omitted	2.372***	2.505***	2.151***	Omitted	-1.958***	-2.784***	-0.797***	-0.279***	0.338*	Omitted	1.301***	-0.545
	(0.692)	(0.236)	(0.144)	(0.375)		(0.301)	(0.0962)	(0.151)		(0.253)	(0.270)	(0.119)	(0.0984)	(0.197)		(0.0944)	(0.382)
Tobacco	-1.263	-0.286	0.0909	0.450	1.806*	0.0753	-0.108	-0.517	-0.617	0.269	-0.482	0.563*	0.711**	-0.378	-0.0581	-0.215	-0.106
	(1.879)	(0.640)	(0.392)	(1.018)	(0.995)	(0.816)	(0.306)	(0.410)	(0.640)	(0.687)	(0.733)	(0.323)	(0.313)	(0.535)	(0.297)	(0.256)	(1.036)
Cons	-193.5***	-1.703	6.415***	10.51*	96.85***	26.84***	-4.794***	-1.876	-15.28***	93.29***	-14.94***	16.07***	25.92***	17.86***	6.921***	12.01***	29.00***
	(10.11)	(3.445)	(2.110)	(5.475)	(5.353)	(4.390)	(1.649)	(2.206)	(3.441)	(3.697)	(3.944)	(1.738)	(1.686)	(2.879)	(1.595)	(1.378)	(5.575)
Obs.	13,679	13,679	13,679	13,679	13,679	13,679	13,679	13,679	13,679	13,679	13,679	13,679	13,679	13,679	13,679	13,679	13,679
R-squared	0.204	0.098	0.012	0.199	0.176	0.177	0.130	0.147	0.102	0.197	0.037	0.053	0.092	0.185	0.046	0.153	0.037

Table A 3: Ordinary Least Squares results, Urban Sample

Variable	Food	Alc	Health	Sch	Clothing	Housing	Water	Electricity	Alt. Energ	Daily Tprt	Other Tprt	Equpt	Ent	Tele	Remit	Hse Care	Pers. Cre
d	-1.226*	4.841***	-0.0205	-1.374***	0.205	-1.583***	-0.352**	-0.673***	0.203	-0.144	0.492**	0.171*	0.104	0.0149	0.0212	-0.135	-1.032***
	(0.638)	(0.219)	(0.121)	(0.374)	(0.274)	(0.431)	(0.162)	(0.206)	(0.224)	(0.306)	(0.244)	(0.102)	(0.160)	(0.253)	(0.114)	(0.122)	(0.339)
LnM	40.47***	1.875*	-0.133	-10.58***	-9.695***	6.434***	2.467***	7.008***	-0.185	-27.96***	1.589	-6.083***	-9.032***	4.042***	-1.321**	-1.555***	3.383**
	(2.965)	(1.018)	(0.564)	(1.736)	(1.271)	(2.000)	(0.753)	(0.955)	(1.040)	(1.422)	(1.135)	(0.475)	(0.743)	(1.176)	(0.529)	(0.567)	(1.573)
LnM2	-1.833***	-0.0547	0.00873	0.426***	0.344***	-0.160**	-0.0944***	-0.254***	-0.0621	1.184***	-0.0230	0.245***	0.376***	-0.0915**	0.0687***	0.0849***	-0.166***
	(0.112)	(0.0384)	(0.0213)	(0.0655)	(0.0479)	(0.0754)	(0.0284)	(0.0360)	(0.0392)	(0.0536)	(0.0428)	(0.0179)	(0.0280)	(0.0443)	(0.0200)	(0.0214)	(0.0593)
Adult prop	2.080	1.043**	0.192	-0.743	-1.527**	-1.321	0.895**	1.339***	-0.696	0.786	-0.883*	0.103	-0.382	0.187	-0.844***	-0.0543	-0.193
	(1.395)	(0.479)	(0.265)	(0.817)	(0.598)	(0.941)	(0.354)	(0.449)	(0.489)	(0.669)	(0.534)	(0.223)	(0.349)	(0.553)	(0.249)	(0.267)	(0.740)
loghhsize	2.220***	-0.144	0.0760	3.159***	0.0103	-3.253***	0.00224	0.406**	0.182	-0.504*	-0.753***	0.165*	-0.492***	-0.297	-0.542***	-0.463***	0.240
	(0.597)	(0.205)	(0.114)	(0.350)	(0.256)	(0.403)	(0.152)	(0.192)	(0.209)	(0.286)	(0.229)	(0.0957)	(0.150)	(0.237)	(0.107)	(0.114)	(0.317)
Head Sch	-0.704	-0.276	-0.409***	1.056***	0.278	0.471	0.287	0.485**	-0.965***	-0.326	0.142	0.215*	-0.117	0.567**	-0.0226	-0.196	-0.166
	(0.698)	(0.240)	(0.133)	(0.409)	(0.299)	(0.471)	(0.177)	(0.225)	(0.245)	(0.335)	(0.267)	(0.112)	(0.175)	(0.277)	(0.125)	(0.133)	(0.370)
Head w age	-1.143***	0.241**	-0.216***	0.230	-0.0730	1.883***	-0.217**	-0.0352	-0.336***	-0.681***	-0.548***	-0.318***	0.140	0.419***	0.120*	-0.00102	0.862***
	(0.357)	(0.122)	(0.0679)	(0.209)	(0.153)	(0.241)	(0.0906)	(0.115)	(0.125)	(0.171)	(0.137)	(0.0572)	(0.0894)	(0.141)	(0.0637)	(0.0682)	(0.189)
Hhold Emp	-0.278	0.243***	0.00219	-0.815***	0.425***	-0.291**	-0.0588	-0.291***	-0.00956	0.150	0.0656	0.0632*	0.0654	0.322***	0.185***	0.0855**	0.0281
	(0.215)	(0.0739)	(0.0409)	(0.126)	(0.0922)	(0.145)	(0.0547)	(0.0693)	(0.0755)	(0.103)	(0.0824)	(0.0345)	(0.0539)	(0.0853)	(0.0384)	(0.0411)	(0.114)
Adult Age	4.289***	0.816	0.888***	-7.749***	-0.432	0.380	0.1000	-0.393	1.071**	0.591	0.178	0.145	0.260	0.558	0.0598	0.582**	-1.529*
	(1.506)	(0.517)	(0.286)	(0.882)	(0.646)	(1.016)	(0.383)	(0.485)	(0.528)	(0.722)	(0.576)	(0.241)	(0.377)	(0.597)	(0.269)	(0.288)	(0.799)
Child age	-0.634**	-0.111	-0.0617	2.389***	-0.130	-0.758***	0.0440	0.256***	-0.305***	0.00255	-0.181*	0.0123	0.0204	0.0569	-0.0818*	0.00919	-0.428***
	(0.266)	(0.0914)	(0.0507)	(0.156)	(0.114)	(0.180)	(0.0677)	(0.0858)	(0.0934)	(0.128)	(0.102)	(0.0427)	(0.0667)	(0.106)	(0.0475)	(0.0509)	(0.141)
Head Age	-4.398***	-1.405***	-0.198	9.139***	-0.717	-3.645***	1.287***	1.562***	0.237	-0.726	0.747	-0.0529	-0.425	-1.021*	-0.0448	-0.406	0.439
	(1.330)	(0.456)	(0.253)	(0.779)	(0.570)	(0.897)	(0.338)	(0.428)	(0.466)	(0.638)	(0.509)	(0.213)	(0.333)	(0.527)	(0.237)	(0.254)	(0.705)
Most Edu	-2.011**	-0.309	0.416**	1.032*	-0.280	-1.036	0.485**	1.420***	-1.249***	-0.331	-0.666*	-0.169	0.219	1.066***	0.452***	0.681***	0.404
	(0.939)	(0.322)	(0.179)	(0.550)	(0.403)	(0.634)	(0.239)	(0.302)	(0.329)	(0.450)	(0.359)	(0.150)	(0.235)	(0.372)	(0.168)	(0.180)	(0.498)
6.stratum	-0.317	-0.448***	-0.0923	0.985***	-0.156	0.0776	1.292***	0.934***	-0.634***	-1.034***	-0.0306	-0.141**	-0.130	-0.210	-0.295***	0.0570	0.566**
	(0.444)	(0.152)	(0.0845)	(0.260)	(0.190)	(0.300)	(0.113)	(0.143)	(0.156)	(0.213)	(0.170)	(0.0712)	(0.111)	(0.176)	(0.0793)	(0.0849)	(0.236)
7.stratum	-0.896*	-0.466**	-0.252**	0.691**	-0.435*	-0.730**	1.066***	1.253***	-0.227	-1.430***	-0.695***	0.0898	0.546***	0.0209	-0.206**	1.125***	0.841***
	(0.528)	(0.181)	(0.100)	(0.309)	(0.226)	(0.356)	(0.134)	(0.170)	(0.185)	(0.253)	(0.202)	(0.0846)	(0.132)	(0.209)	(0.0942)	(0.101)	(0.280)
Cluster	-0.000141	-0.000364*	0.000161	0.00201***	0.000784***	0.000180	1.40e-05	-0.000941***	-0.000783***	-0.000343	-0.000210	-1.43e-05	0.000210	0.00101***	0.000253**	-0.000347***	-0.00167***
	(0.000594)	(0.000204)	(0.000113)	(0.000348)	(0.000255)	(0.000401)	(0.000151)	(0.000191)	(0.000208)	(0.000285)	(0.000227)	(9.51e-05)	(0.000149)	(0.000235)	(0.000106)	(0.000113)	(0.000315)
Tobacco	-5.799	-1.604	-0.0988	-0.578	-0.706	2.726	-0.362	-2.003	-0.447	7.981***	0.186	1.466*	2.726**	-0.925	-0.454	-1.302	-0.991
	(5.177)	(1.777)	(0.985)	(3.031)	(2.219)	(3.493)	(1.315)	(1.667)	(1.816)	(2.482)	(1.981)	(0.829)	(1.297)	(2.053)	(0.924)	(0.989)	(2.746)
Cons	-160.0***	-10.52	-1.405	50.49***	78.30***	-30.83**	-21.75***	-55.63***	19.51***	167.3***	-14.82*	37.20***	55.21***	-36.65***	6.247*	6.012	-1.818
	(19.93)	(6.839)	(3.790)	(11.67)	(8.542)	(13.44)	(5.062)	(6.417)	(6.988)	(9.554)	(7.625)	(3.193)	(4.992)	(7.901)	(3.557)	(3.809)	(10.57)
Obs.	7,275	7,275	7,275	7,275	7,275	7,275	7,275	7,275	7,275	7,275	7,275	7,275	7,275	7,275	7,275	7,275	7,275
R-squared	0.305	0.077	0.008	0.203	0.029	0.105	0.051	0.090	0.209	0.209	0.023	0.062	0.099	0.121	0.050	0.119	0.021

Table A 4: Ordinary Least Squares, Rural Sample

Variable	Food	Alc	Health	Sch	Clothing	Housing	Water	Electricity	Alt. Energ	Daily Tprt	Other Tprt	Equpt	Ent	Tele	Remit	Hse Care	Pers. Cre
d	-2.614***	6.496***	-0.0801	-1.980***	-0.496	-0.0617	-0.0420	-0.0393	0.0641	-0.175	-0.206	-0.0232	0.0747	-0.233*	-0.153	0.0980	-0.617
	(0.708)	(0.242)	(0.157)	(0.364)	(0.415)	(0.0653)	(0.0328)	(0.0714)	(0.235)	(0.190)	(0.280)	(0.127)	(0.0508)	(0.127)	(0.102)	(0.0600)	(0.402)
LnM	47.84***	1.145	-0.455	-5.487***	-7.228***	-0.707**	-0.227	-2.660***	1.188	-14.84***	-6.296***	-2.753***	-0.645***	-6.657***	-2.479***	-1.727***	1.703
	(3.010)	(1.026)	(0.665)	(1.545)	(1.765)	(0.277)	(0.139)	(0.303)	(0.999)	(0.806)	(1.189)	(0.539)	(0.216)	(0.538)	(0.433)	(0.255)	(1.707)
LnM2	-2.103***	-0.0350	0.0143	0.253***	0.203***	0.0366***	0.0124**	0.125***	-0.0862**	0.694***	0.350***	0.126***	0.0321***	0.322***	0.123***	0.0836***	-0.135*
	(0.130)	(0.0444)	(0.0288)	(0.0669)	(0.0764)	(0.0120)	(0.00603)	(0.0131)	(0.0432)	(0.0349)	(0.0514)	(0.0233)	(0.00934)	(0.0233)	(0.0187)	(0.0110)	(0.0739)
Adult prop	-6.025***	0.634	0.376	2.634**	2.347*	0.226	0.169*	0.317	0.232	0.131	-1.223	0.748**	-0.0171	0.746**	-0.715**	0.347*	-1.798
	(2.113)	(0.720)	(0.467)	(1.085)	(1.239)	(0.195)	(0.0978)	(0.213)	(0.701)	(0.566)	(0.835)	(0.379)	(0.152)	(0.378)	(0.304)	(0.179)	(1.199)
loghhsize	-6.287***	-0.201	0.279	4.642***	2.400***	-0.100	0.0187	-0.0572	-0.0168	0.359	-1.017***	0.905***	0.0270	0.0815	-0.183	-0.100	-0.931*
	(0.844)	(0.288)	(0.187)	(0.433)	(0.495)	(0.0778)	(0.0391)	(0.0851)	(0.280)	(0.226)	(0.333)	(0.151)	(0.0606)	(0.151)	(0.121)	(0.0715)	(0.479)
Head Sch	-0.0733	-0.180	-0.106	0.866**	-0.142	-0.0418	0.0159	0.0895	0.183	-0.0787	-0.440	0.133	0.0230	0.0828	-0.0843	0.132**	-0.572
	(0.713)	(0.243)	(0.158)	(0.366)	(0.418)	(0.0658)	(0.0330)	(0.0719)	(0.237)	(0.191)	(0.282)	(0.128)	(0.0512)	(0.127)	(0.103)	(0.0605)	(0.405)
Head w age	0.948	-0.0647	-0.342*	-0.0772	-0.963**	0.775***	0.0376	0.199**	-0.286	-1.068***	-1.078***	-0.944***	0.194***	0.736***	0.281**	0.237***	1.711***
	(0.812)	(0.277)	(0.179)	(0.417)	(0.476)	(0.0749)	(0.0376)	(0.0819)	(0.270)	(0.218)	(0.321)	(0.146)	(0.0583)	(0.145)	(0.117)	(0.0688)	(0.461)
Hhold Emp	0.373	0.0469	0.0467	-1.026***	0.0679	-0.0187	-0.0378***	-0.0198	0.114	0.0538	0.0370	-0.0134	0.00522	-0.0670	0.0609	0.0185	0.360**
	(0.270)	(0.0922)	(0.0597)	(0.139)	(0.159)	(0.0249)	(0.0125)	(0.0273)	(0.0897)	(0.0724)	(0.107)	(0.0484)	(0.0194)	(0.0483)	(0.0389)	(0.0229)	(0.153)
Adult Age	2.085	0.520	1.029**	-9.334***	1.740	0.0271	-0.00345	0.0757	0.637	0.941*	-0.0974	0.908**	0.0644	0.318	0.120	0.0696	0.582
	(2.043)	(0.696)	(0.451)	(1.049)	(1.198)	(0.188)	(0.0945)	(0.206)	(0.678)	(0.547)	(0.807)	(0.366)	(0.147)	(0.365)	(0.294)	(0.173)	(1.159)
Child age	-0.418	-0.262**	0.0764	2.476***	-0.792***	0.0920***	0.00859	0.0905**	-0.267**	-0.199**	-0.0838	-0.0720	-0.0263	0.0350	-0.126**	0.0654**	-0.365*
	(0.377)	(0.129)	(0.0834)	(0.194)	(0.221)	(0.0348)	(0.0175)	(0.0381)	(0.125)	(0.101)	(0.149)	(0.0676)	(0.0271)	(0.0674)	(0.0543)	(0.0320)	(0.214)
Head Age	-0.923	-0.789	-0.730*	11.26***	-4.551***	-0.303*	-0.0351	-0.0370	-0.740	-0.568	0.476	-1.041***	-0.0410	-0.606*	-0.196	-0.0583	-0.802
	(1.854)	(0.632)	(0.410)	(0.952)	(1.087)	(0.171)	(0.0858)	(0.187)	(0.615)	(0.497)	(0.732)	(0.332)	(0.133)	(0.331)	(0.267)	(0.157)	(1.052)
Most Edu	-1.833**	-0.137	-0.340*	0.0643	1.108**	0.101	-0.0569	0.0346	0.0497	-0.437*	0.0473	-0.0838	-0.0491	0.192	0.221*	-0.0989	1.027**
	(0.900)	(0.307)	(0.199)	(0.462)	(0.528)	(0.0829)	(0.0416)	(0.0907)	(0.299)	(0.241)	(0.355)	(0.161)	(0.0645)	(0.161)	(0.129)	(0.0762)	(0.510)
2.stratum	-2.105***	-0.712***	0.167	1.021***	1.193***	-0.177**	-0.0936***	-0.0986	-0.447*	1.021***	-0.448	0.382***	-0.113**	0.240*	0.0572	0.161**	0.0479
	(0.754)	(0.257)	(0.167)	(0.387)	(0.442)	(0.0695)	(0.0349)	(0.0760)	(0.250)	(0.202)	(0.298)	(0.135)	(0.0541)	(0.135)	(0.108)	(0.0639)	(0.427)
3.stratum	-1.328	-1.968*	-0.191	-0.780	-0.877	0.547*	-0.237	-0.230	-0.153	5.730***	-5.850***	2.010***	0.0555	3.054***	1.209**	1.088***	-1.385
	(3.319)	(1.131)	(0.733)	(1.704)	(1.946)	(0.306)	(0.154)	(0.335)	(1.101)	(0.889)	(1.311)	(0.595)	(0.238)	(0.593)	(0.477)	(0.281)	(1.883)
4.stratum	5.349***	-0.720***	0.383**	-0.699*	-2.116***	0.148**	0.0199	-0.0359	-0.0207	0.187	0.249	-0.502***	-0.0164	-0.0376	-0.196*	0.0424	-1.724***
	(0.758)	(0.258)	(0.167)	(0.389)	(0.444)	(0.0698)	(0.0351)	(0.0764)	(0.252)	(0.203)	(0.299)	(0.136)	(0.0544)	(0.135)	(0.109)	(0.0642)	(0.430)
Cluster	0.00152*	-0.000153	9.15e-05	0.00157***	-0.000110	0.000110	0.000202***	5.90e-05	-0.00196***	0.000536**	-0.000643**	-0.000664***	4.31e-05	2.13e-05	-1.22e-05	7.57e-06	-0.00108**
	(0.000823)	(0.000281)	(0.000182)	(0.000423)	(0.000483)	(7.59e-05)	(3.81e-05)	(8.30e-05)	(0.000273)	(0.000220)	(0.000325)	(0.000148)	(5.91e-05)	(0.000147)	(0.000118)	(6.98e-05)	(0.000467)
Tobacco	-0.158	-0.148	0.0981	0.409	1.670	-0.119	-0.0643	-0.167	-0.750	-0.600	-0.557	0.295	0.396**	-0.409	-0.0628	-0.0660	0.181
	(2.249)	(0.767)	(0.497)	(1.155)	(1.319)	(0.207)	(0.104)	(0.227)	(0.747)	(0.602)	(0.888)	(0.403)	(0.161)	(0.402)	(0.323)	(0.191)	(1.276)
Cons	-204.8***	-4.899	3.656	12.35	74.71***	4.092**	1.125	13.48***	3.934	77.84***	29.71***	15.32***	3.198**	34.24***	13.36***	8.666***	14.05
	(18.03)	(6.145)	(3.983)	(9.255)	(10.57)	(1.661)	(0.834)	(1.818)	(5.982)	(4.827)	(7.119)	(3.230)	(1.293)	(3.221)	(2.592)	(1.528)	(10.23)
Obs.	6,404	6,404	6,404	6,404	6,404	6,404	6,404	6,404	6,404	6,404	6,404	6,404	6,404	6,404	6,404	6,404	6,404
R-squared	0.099	0.111	0.008	0.199	0.069	0.052	0.012	0.046	0.031	0.140	0.057	0.037	0.015	0.140	0.040	0.054	0.025

Table A 5: Ordinary Least Squares, Top 50%

Variable	Food	Alc	Health	Sch	Clothing	Housing	Water	Electricity	Alt. Energ	Daily Tprt	Other Tprt	Equpt	Ent	Tele	Remit	Hse Care	Pers. Cre
d	-0.399	5.026***	-0.128	-1.391***	-0.271	-1.538***	-0.365***	-0.496***	-0.00643	-0.235	0.293	0.163	0.105	0.00432	-0.0501	-0.106	-0.866***
	(0.608)	(0.211)	(0.106)	(0.389)	(0.269)	(0.388)	(0.137)	(0.191)	(0.185)	(0.326)	(0.300)	(0.123)	(0.152)	(0.253)	(0.127)	(0.120)	(0.312)
LnM	18.50***	7.110***	-0.459	-6.857*	-6.134**	15.34***	4.250***	20.89***	-16.59***	-44.25***	3.527	-8.044***	-14.89***	21.53***	2.608**	-0.0103	12.48***
	(6.251)	(2.172)	(1.092)	(3.999)	(2.770)	(3.988)	(1.407)	(1.960)	(1.905)	(3.350)	(3.086)	(1.261)	(1.563)	(2.600)	(1.311)	(1.233)	(3.203)
LnM2	-1.055***	-0.238***	0.0195	0.283*	0.208**	-0.490***	-0.158***	-0.751***	0.531***	1.787***	-0.0884	0.321***	0.583***	-0.720***	-0.0690	0.0276	-0.490***
	(0.226)	(0.0785)	(0.0395)	(0.144)	(0.100)	(0.144)	(0.0508)	(0.0708)	(0.0689)	(0.121)	(0.112)	(0.0456)	(0.0565)	(0.0940)	(0.0474)	(0.0446)	(0.116)
Adult prop	-0.721	0.372	0.492**	0.146	-1.332**	-0.607	0.972***	1.303***	-0.420	1.077	-1.052	0.294	-0.441	0.420	-1.141***	0.0889	0.283
	(1.402)	(0.487)	(0.245)	(0.897)	(0.621)	(0.895)	(0.316)	(0.440)	(0.427)	(0.751)	(0.692)	(0.283)	(0.351)	(0.583)	(0.294)	(0.277)	(0.718)
loghhsize	1.332**	-0.371*	0.00119	3.729***	0.118	-3.008***	0.0870	0.272	0.262	-0.0581	-1.022***	0.349***	-0.457***	-0.282	-0.630***	-0.530***	0.206
	(0.589)	(0.205)	(0.103)	(0.377)	(0.261)	(0.376)	(0.133)	(0.185)	(0.180)	(0.316)	(0.291)	(0.119)	(0.147)	(0.245)	(0.124)	(0.116)	(0.302)
Head Sch	-0.392	-0.604**	-0.516***	1.958***	-0.446	0.368	0.218	0.529**	-0.772***	-0.280	0.0499	0.192	-0.0484	0.483*	-0.0832	-0.0136	-0.372
	(0.683)	(0.237)	(0.119)	(0.437)	(0.303)	(0.436)	(0.154)	(0.214)	(0.208)	(0.366)	(0.337)	(0.138)	(0.171)	(0.284)	(0.143)	(0.135)	(0.350)
Head w age	-0.514	0.176	-0.169***	0.223	0.0818	1.767***	-0.197**	-0.179	-0.375***	-0.635***	-0.839***	-0.458***	0.218**	0.313**	0.107	0.00597	0.839***
	(0.361)	(0.126)	(0.0632)	(0.231)	(0.160)	(0.231)	(0.0813)	(0.113)	(0.110)	(0.194)	(0.178)	(0.0729)	(0.0904)	(0.150)	(0.0758)	(0.0713)	(0.185)
Hhold Emp	-0.190	0.148**	0.0278	-0.878***	0.221**	-0.0969	-0.0915**	-0.237***	0.101	0.137	0.0770	0.0634	0.0607	0.225***	0.182***	0.0825**	0.0937
	(0.204)	(0.0710)	(0.0357)	(0.131)	(0.0906)	(0.130)	(0.0460)	(0.0641)	(0.0623)	(0.110)	(0.101)	(0.0413)	(0.0511)	(0.0850)	(0.0429)	(0.0403)	(0.105)
Adult Age	2.597*	0.461	1.146***	-9.039***	0.623	0.524	0.270	0.207	0.583	1.390*	0.0605	0.503*	0.109	0.319	-0.0468	0.497*	-0.472
	(1.502)	(0.522)	(0.263)	(0.961)	(0.666)	(0.958)	(0.338)	(0.471)	(0.458)	(0.805)	(0.742)	(0.303)	(0.376)	(0.625)	(0.315)	(0.296)	(0.770)
Child age	-0.830***	-0.325***	-0.0416	2.596***	-0.350***	-0.616***	0.0627	0.254***	-0.117	-0.0512	-0.113	0.00317	0.0104	0.0421	-0.141**	0.0131	-0.253*
	(0.278)	(0.0966)	(0.0486)	(0.178)	(0.123)	(0.177)	(0.0626)	(0.0872)	(0.0848)	(0.149)	(0.137)	(0.0561)	(0.0696)	(0.116)	(0.0583)	(0.0549)	(0.143)
Head Age	-3.247**	-0.956**	-0.158	10.71***	-0.886	-3.438***	1.005***	1.157***	-0.185	-1.419**	0.696	-0.249	-0.278	-1.328**	-0.0817	-0.299	-0.801
	(1.314)	(0.456)	(0.230)	(0.840)	(0.582)	(0.838)	(0.296)	(0.412)	(0.400)	(0.704)	(0.649)	(0.265)	(0.328)	(0.546)	(0.275)	(0.259)	(0.673)
Most Edu	-1.662*	-0.264	0.490***	0.250	0.245	-1.179**	0.413**	1.116***	-0.771***	-0.886*	-0.680	-0.276	0.219	1.163***	0.598***	0.581***	0.510
	(0.934)	(0.325)	(0.163)	(0.598)	(0.414)	(0.596)	(0.210)	(0.293)	(0.285)	(0.501)	(0.461)	(0.189)	(0.234)	(0.389)	(0.196)	(0.184)	(0.479)
2.stratum	-4.975***	-0.729**	0.149	1.627***	1.039***	0.728	-0.399**	-0.535*	-0.459*	3.275***	-0.541	1.035***	-0.200	0.538	0.0523	0.309*	-0.770*
	(0.881)	(0.306)	(0.154)	(0.564)	(0.390)	(0.562)	(0.198)	(0.276)	(0.269)	(0.472)	(0.435)	(0.178)	(0.220)	(0.367)	(0.185)	(0.174)	(0.451)
3.stratum	-6.651**	-2.088**	-0.300	-0.0955	-1.303	1.535	-0.490	-0.467	0.401	7.923***	-4.971***	3.380***	-0.540	4.099***	1.797***	1.344**	-3.486**
	(2.785)	(0.968)	(0.487)	(1.782)	(1.234)	(1.777)	(0.627)	(0.873)	(0.849)	(1.493)	(1.375)	(0.562)	(0.696)	(1.159)	(0.584)	(0.549)	(1.427)
4.stratum	4.105***	-0.396	0.0523	-2.788***	-0.818	-1.526**	0.138	-0.111	0.526	0.145	1.234**	-0.0887	-0.105	0.100	-0.240	0.0966	-0.721
	(1.155)	(0.401)	(0.202)	(0.739)	(0.512)	(0.737)	(0.260)	(0.362)	(0.352)	(0.619)	(0.570)	(0.233)	(0.289)	(0.480)	(0.242)	(0.228)	(0.592)
5.stratum	3.348***	0.487**	0.242**	-2.477***	-3.722***	3.671***	1.106***	-1.319***	1.808***	1.865***	-2.754***	-0.669***	0.0829	-0.122	-0.647***	-0.0590	-0.373
	(0.544)	(0.192)	(0.0967)	(0.348)	(0.241)	(0.347)	(0.122)	(0.174)	(0.166)	(0.297)	(0.269)	(0.110)	(0.118)	(0.230)	(0.114)	(0.0931)	(0.284)
6.stratum	3.822***	-0.00836	0.137	-1.885***	-3.596***	3.795***	2.401***	-0.389**	1.054***	0.596*	-2.851***	-0.843***	Omitted	-0.346	-0.985***	Omitted	0.0346
	(0.660)	(0.213)	(0.107)	(0.422)	(0.293)	(0.421)	(0.149)	(0.192)	(0.201)	(0.328)	(0.326)	(0.133)		(0.254)	(0.138)		(0.313)
7.stratum	3.106***	Omitted	Omitted	-1.790***	-3.881***	3.183***	2.287***	Omitted	1.255***	Omitted	-3.556***	-0.642***	0.693***	Omitted	-0.919***	1.124***	Omitted
	(0.732)			(0.468)	(0.324)	(0.467)	(0.165)		(0.223)		(0.361)	(0.148)	(0.153)		(0.154)	(0.121)	
Cluster	-0.000783	-0.000152	8.49e-05	0.00188***	5.79e-05	0.000501	-0.000132	-0.000841***	6.47e-05	0.000200	-0.000530*	0.000121	0.000216	0.000865***	0.000307**	-0.000197*	-0.00175***
	(0.000593)	(0.000206)	(0.000104)	(0.000380)	(0.000263)	(0.000379)	(0.000134)	(0.000186)	(0.000181)	(0.000318)	(0.000293)	(0.000120)	(0.000148)	(0.000247)	(0.000124)	(0.000117)	(0.000304)
Tobacco	-0.101	-1.347	0.0435	0.176	1.254	1.011	-0.279	-0.946	-0.877	0.703	-0.280	0.459	1.083*	-0.403	-0.211	-0.388	0.0650
	(2.615)	(0.909)	(0.457)	(1.673)	(1.159)	(1.668)	(0.588)	(0.820)	(0.797)	(1.401)	(1.291)	(0.528)	(0.654)	(1.088)	(0.548)	(0.516)	(1.340)
Cons	-5.952	-45.76***	-0.0315	25.79	56.72***	-96.77***	-34.65***	-152.8***	131.1***	277.0***	-24.62	49.86***	96.13***	-156.0***	-20.27**	-4.952	-63.85***
	(43.11)	(14.98)	(7.534)	(27.58)	(19.10)	(27.50)	(9.701)	(13.52)	(13.14)	(23.10)	(21.28)	(8.700)	(10.78)	(17.93)	(9.040)	(8.504)	(22.09)
Obs.	7,501	7,501	7,501	7,501	7,501	7,501	7,501	7,501	7,501	7,501	7,501	7,501	7,501	7,501	7,501	7,501	7,501
R-squared	0.259	0.089	0.014	0.200	0.077	0.125	0.093	0.099	0.164	0.191	0.035	0.079	0.092	0.087	0.043	0.100	0.023

Table A 6: Ordinary Least Squares, Bottom 50%

Variable	Food	Alc	Health	Sch	Clothing	Housing	Water	Electricity	Alt. Energ	Daily Tprt	Other Tprt	Equpt	Ent	Tele	Remit	Hse Care	Pers. Cre
d	-3.564***	6.499***	-0.00107	-1.898***	-0.124	0.107	0.0304	-0.0402	0.0625	-0.172*	-0.129	-0.0549	0.0599	-0.141*	-0.109	0.119**	-0.719*
	(0.742)	(0.251)	(0.172)	(0.343)	(0.435)	(0.122)	(0.0718)	(0.0714)	(0.268)	(0.104)	(0.219)	(0.109)	(0.0388)	(0.0855)	(0.0795)	(0.0475)	(0.431)
LnM	32.58***	-0.111	-2.847**	-7.516***	-8.352***	-3.350***	-0.101	-0.196	0.628	-2.195***	-4.528***	2.160***	-0.208	-2.283***	-0.738	-0.0921	1.788
	(5.237)	(1.771)	(1.212)	(2.420)	(3.068)	(0.864)	(0.507)	(0.504)	(1.887)	(0.734)	(1.547)	(0.767)	(0.274)	(0.603)	(0.561)	(0.335)	(3.041)
LnM2	-1.383***	0.0217	0.128**	0.343***	0.251*	0.166***	0.00749	0.00787	-0.0605	0.114***	0.255***	-0.101***	0.0121	0.114***	0.0409	0.00731	-0.139
	(0.240)	(0.0812)	(0.0556)	(0.111)	(0.141)	(0.0396)	(0.0232)	(0.0231)	(0.0866)	(0.0337)	(0.0709)	(0.0352)	(0.0125)	(0.0277)	(0.0257)	(0.0154)	(0.139)
Adult prop	-2.989	1.797**	0.0202	1.039	2.510**	-0.538	0.0126	0.476**	-0.453	-0.203	-0.957	0.519*	0.0790	0.429*	-0.262	0.169	-2.150*
	(2.133)	(0.721)	(0.493)	(0.985)	(1.249)	(0.352)	(0.206)	(0.205)	(0.769)	(0.299)	(0.630)	(0.312)	(0.111)	(0.246)	(0.228)	(0.136)	(1.238)
loghhsize	-5.553***	0.137	0.341*	3.678***	2.454***	-0.442***	-0.0577	0.137	-0.180	-0.0415	-0.650**	0.719***	0.0425	0.0998	0.00317	0.00234	-0.833*
	(0.869)	(0.294)	(0.201)	(0.401)	(0.509)	(0.143)	(0.0841)	(0.0836)	(0.313)	(0.122)	(0.257)	(0.127)	(0.0454)	(0.100)	(0.0930)	(0.0556)	(0.504)
Head Sch	-0.167	0.0340	-0.0381	0.131	0.217	-0.00205	0.0328	-0.00532	0.166	-0.00624	-0.378*	0.131	0.00505	0.0790	-0.0253	0.0240	-0.411
	(0.730)	(0.247)	(0.169)	(0.337)	(0.427)	(0.120)	(0.0706)	(0.0702)	(0.263)	(0.102)	(0.216)	(0.107)	(0.0381)	(0.0840)	(0.0781)	(0.0467)	(0.424)
Head w age	-0.523	0.114	-0.429**	-0.108	-1.812***	0.915***	-0.0815	0.0995	0.401	0.0349	-0.327	-0.237*	0.0113	0.177*	0.137	0.0984*	1.329***
	(0.877)	(0.297)	(0.203)	(0.405)	(0.514)	(0.145)	(0.0849)	(0.0844)	(0.316)	(0.123)	(0.259)	(0.128)	(0.0458)	(0.101)	(0.0939)	(0.0561)	(0.509)
Hhold Emp	0.173	0.127	0.0194	-0.852***	0.285*	-0.0951**	-0.0216	-0.0595**	-0.0146	0.0232	0.0418	-0.0172	0.00518	-0.00216	0.0393	0.0148	0.305*
	(0.287)	(0.0971)	(0.0664)	(0.133)	(0.168)	(0.0473)	(0.0278)	(0.0276)	(0.103)	(0.0402)	(0.0848)	(0.0420)	(0.0150)	(0.0331)	(0.0307)	(0.0184)	(0.167)
Adult Age	3.123	0.873	0.826*	-7.094***	1.161	-0.0984	-0.145	-0.459**	0.883	-0.0938	0.0700	0.677**	0.154	0.476**	0.136	0.105	-0.796
	(2.094)	(0.708)	(0.484)	(0.967)	(1.227)	(0.345)	(0.203)	(0.201)	(0.755)	(0.293)	(0.618)	(0.307)	(0.109)	(0.241)	(0.224)	(0.134)	(1.216)
Child age	-0.0610	-0.0351	0.0508	2.211***	-0.535**	-0.219***	-0.00228	0.0870**	-0.403***	-0.0744	-0.190*	-0.0477	-0.00874	-0.0176	-0.0687*	0.0469**	-0.565***
	(0.365)	(0.124)	(0.0846)	(0.169)	(0.214)	(0.0603)	(0.0354)	(0.0352)	(0.132)	(0.0512)	(0.108)	(0.0535)	(0.0191)	(0.0421)	(0.0391)	(0.0234)	(0.212)
Head Age	-1.819	-1.354**	-0.719	9.143***	-4.677***	-0.345	0.200	0.340*	-0.233	0.433	0.448	-0.886***	-0.149	-0.421*	-0.178	-0.102	0.684
	(1.909)	(0.646)	(0.442)	(0.882)	(1.118)	(0.315)	(0.185)	(0.184)	(0.688)	(0.267)	(0.564)	(0.280)	(0.0997)	(0.220)	(0.204)	(0.122)	(1.108)
Most Edu	-1.776*	-0.289	-0.427**	0.744*	0.570	0.0377	-0.0355	0.128	0.0458	-0.00513	-0.0206	0.0148	-0.0523	0.0684	0.105	-0.0547	0.881*
	(0.914)	(0.309)	(0.212)	(0.422)	(0.536)	(0.151)	(0.0885)	(0.0880)	(0.330)	(0.128)	(0.270)	(0.134)	(0.0478)	(0.105)	(0.0979)	(0.0585)	(0.531)
2.stratum	-1.899**	-0.758**	0.136	0.911**	1.308**	0.0233	-0.0680	-0.0541	-0.342	0.00552	-0.0807	0.292**	-0.0465	0.0162	0.136	0.0965	0.243
	(0.964)	(0.326)	(0.223)	(0.445)	(0.565)	(0.159)	(0.0933)	(0.0927)	(0.347)	(0.135)	(0.285)	(0.141)	(0.0504)	(0.111)	(0.103)	(0.0617)	(0.560)
3.stratum	3.787	-1.351	0.0881	-0.639	0.329	0.119	-0.166	-0.0910	0.348	-0.495	-1.791	-0.742	-0.163	0.942	0.0209	-0.139	0.681
	(7.409)	(2.506)	(1.714)	(3.423)	(4.341)	(1.222)	(0.717)	(0.713)	(2.670)	(1.038)	(2.189)	(1.085)	(0.387)	(0.853)	(0.793)	(0.474)	(4.303)
4.stratum	6.561***	-0.833***	0.455**	-0.422	-2.546***	-0.167	0.0444	0.0265	-0.0239	0.264**	0.00468	-0.683***	-0.0381	-0.0839	-0.149	0.0144	-1.944***
	(0.856)	(0.289)	(0.198)	(0.395)	(0.501)	(0.141)	(0.0828)	(0.0823)	(0.308)	(0.120)	(0.253)	(0.125)	(0.0447)	(0.0985)	(0.0916)	(0.0547)	(0.497)
5.stratum	5.386***	-0.243	-0.261	0.00160	-0.589	0.737*	-1.723***	-0.689***	0.811	0.120	-0.213	0.0666	-0.00719	-0.0608	-0.169	0.0973	-8.009***
	(1.815)	(0.614)	(0.420)	(1.204)	(1.527)	(0.430)	(0.176)	(0.175)	(0.654)	(0.254)	(0.536)	(0.266)	(0.136)	(0.209)	(0.279)	(0.116)	(1.513)
6.stratum	Omitted	Omitted	Omitted	4.094***	-1.703	0.215	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	-0.0635	Omitted	-0.339	Omitted	-6.728***
				(1.419)	(1.800)	(0.507)							(0.161)		(0.329)		(1.784)
7.stratum	-4.032	0.257	-0.420	Omitted	Omitted	Omitted	-2.031***	0.551*	1.345	-0.204	-0.173	0.0507	Omitted	-0.0926	Omitted	0.216	Omitted
	(3.072)	(1.039)	(0.711)				(0.297)	(0.296)	(1.107)	(0.430)	(0.907)	(0.450)		(0.354)		(0.197)	
Cluster	0.00244***	-0.000454	0.000176	0.00173***	0.000500	-0.000224	0.000337***	-0.000123	-0.00285***	5.54e-05	-0.000242	-0.000834***	5.11e-05	4.83e-05	-8.60e-05	-0.000137**	-0.00108**
	(0.000834)	(0.000282)	(0.000193)	(0.000385)	(0.000489)	(0.000138)	(8.07e-05)	(8.03e-05)	(0.000301)	(0.000117)	(0.000246)	(0.000122)	(4.36e-05)	(9.61e-05)	(8.93e-05)	(5.34e-05)	(0.000484)
Tobacco	-2.090	0.431	0.132	0.481	2.248	-0.225	0.0588	-0.0396	-0.787	-0.272	-0.663	0.572	0.402***	-0.243	0.0522	-0.0783	-0.0155
	(2.719)	(0.920)	(0.629)	(1.256)	(1.593)	(0.448)	(0.263)	(0.262)	(0.980)	(0.381)	(0.803)	(0.398)	(0.142)	(0.313)	(0.291)	(0.174)	(1.579)
Cons	-128.4***	1.298	16.89**	25.97*	82.68***	19.82***	0.0166	0.899	5.747	9.452**	20.55**	-10.50**	0.849	10.55***	3.651	0.166	13.48
	(28.93)	(9.785)	(6.695)	(13.37)	(16.95)	(4.771)	(2.800)	(2.784)	(10.43)	(4.054)	(8.546)	(4.237)	(1.512)	(3.332)	(3.098)	(1.851)	(16.80)
Obs.	6,178	6,178	6,178	6,178	6,178	6,178	6,178	6,178	6,178	6,178	6,178	6,178	6,178	6,178	6,178	6,178	6,178
R-squared	0.054	0.110	0.009	0.168	0.103	0.119	0.090	0.032	0.053	0.012	0.024	0.039	0.005	0.017	0.009	0.011	0.024

Table A 7: First-stage probit regression results for *d*

Coefficient on:	Full Sample	Urban	Rural	Top 50%	Bottom 50%
Adult Sex ratio	0.895***	1.030***	0.747***	0.786***	1.050***
	(0.101)	(0.145)	(0.143)	(0.153)	(0.159)
Proportion of Adults	0.408**	0.084	0.788***	0.354	0.700**
	(0.172)	(0.245)	(0.246)	(0.248)	(0.289)
Household Size	0.279***	0.148	0.422***	0.131	0.484***
	(0.072)	(0.105)	(0.100)	(0.107)	(0.118)
Head School	-0.058	-0.040	0.062	-0.290***	0.075
	(0.061)	(0.106)	(0.076)	(0.103)	(0.088)
Head Wage Employment	0.083	0.066	0.118	0.060	0.162
	(0.052)	(0.062)	(0.098)	(0.065)	(0.128)
Household Employment	0.056**	0.074**	0.037	0.057	0.042
	(0.025)	(0.036)	(0.034)	(0.036)	(0.040)
Adult Age	0.931***	0.854***	0.996***	0.842***	1.08***
	(0.166)	(0.254)	(0.222)	(0.257)	(0.254)
Child Age	-0.092***	-0.076*	-0.102**	-0.096**	-0.080*
	(0.031)	(0.046)	(0.043)	(0.049)	(0.047)
Household Head Age	-0.407***	-0.308	-0.504**	-0.351	-0.525**
	(0.149)	(0.220)	(0.203)	(0.222)	(0.231)
Most Educated	-0.151*	-0.213	-0.138	-0.133	-0.221**
	(0.079)	(0.144)	(0.095)	(0.149)	(0.108)
Household Type 2	-0.209**	N/A	-0.231**	-0.499***	-0.140
	(0.090)		(0.092)	(0.190)	(0.127)
Household Type 3	-0.288	N/A	-0.351	-0.872*	Omitted
	(0.375)		(0.379)	(0.515)	
Household Type 4	-0.107	N/A	-0.106	0.001	-0.101
	(0.090)		(0.091)	(0.218)	(0.111)
Household Type 5	-0.507	N/A	N/A	0.212	0.413
	(0.757)			(1.290)	(0.770)
Household Type 6	-0.039	-0.558	N/A	-1.620	0.518
	(0.681)	(0.456)		(1.020)	(0.795)
Household Type 7	0.075	-.040	N/A	-0.274	1.370
	(0.560)	(0.468)		(0.765)	(1.070)
Region	Omitted	N/A	N/A	Omitted	Omitted
Tobacco Grow	0.253	0.428	0.220	0.319	0.301
	(0.242)	(0.645)	(0.262)	(0.467)	(0.308)
Observations	**8619**	**4567**	**4052**	**4099**	**3251**

Notes: The table shows results of estimating a probit model for the tobacco smoking status of a household. The dependent variable in the probit is $\boldsymbol{d}$ (the tobacco smoking status of a household). The independent variables include the instrument for $\boldsymbol{d}$, the adult sex ratio, alongside other control variables as discussed in Section 2.4. I do not report the coefficient results for the cluster fixed effects because of space considerations. Standard errors are in parenthesis. *, **, *** signify statistical significance at the 10%, 5% and 1% levels respectively. In all data specifications, the adult sex ratio is a strong predictor of whether the household spends on tobacco or not. The F statistic associated with the instrumental variable are equal to 78, 50, 27, 26 and 44 in the full, urban, rural, top 50% and bottom 50% samples respectively.

Table A 8: First-stage ordinary least squares (OLS) regression results for ln*M*

Coefficient on:	Full Sample	Urban	Rural	Top 50%	Bottom 50%
Log of Assets	0.232***	0.222***	0.235***	0.190***	0.142***
	(.004)	(0.005)	(.007)	(.005)	(0.006)
Proportion of Adults	0.072	0.054	0.118	0.061	0.060
	(0.046)	(0.054)	(0.080)	(0.049)	(0.072)
Household Size	0.222***	0.196***	0.262***	0.163***	0.239***
	(0.020)	(0.023)	(0.033)	(0.021)	(0.030)
Head School	0.134***	0.201***	0.087***	0.157***	0.056**
	(0.019)	(0.027)	(0.026)	(0.024)	(0.024)
Head Wage Employment	0.220***	0.149***	0.424***	0.103***	0.102***
	(0.014)	(0.014)	(0.032)	(0.013)	(0.031)
Household Employment	0.020***	0.055***	-0.020*	0.041***	-0.013
	(0.007)	(0.009)	(0.012)	(0.008)	(0.011)
Adult Age	-0.179***	-0.021	-0.382***	-0.003	-0.213***
	(0.047)	(0.058)	(0.075)	(0.052)	(0.069)
Child Age	0.040***	0.027***	0.054***	0.015	0.012
	(0.009)	(0.010)	(0.014)	(0.010)	(0.012)
Household Head Age	0.009	-0.079	0.143**	-0.084*	0.024
	(0.042)	(0.051)	(0.069)	(0.045)	(0.063)
Most Educated	0.249***	0.271***	0.224***	0.214***	0.161***
	(0.024)	(0.037)	(0.033)	(0.034)	(0.030)
Household Type 2	0.249***	N/A	0.242**	0.038	0.154***
	(0.026)		(0.030)	(0.038)	(0.033)
Household Type 3	0.833***	N/A	0.828***	0.593***	0.164
	(0.109)		(0.122)	(0.110)	(0.240)
Household Type 4	0.118***	N/A	0.099***	0.105**	0.021
	(0.026)		(0.029)	(0.045)	(0.029)
Household Type 5	Omitted	N/A	N/A	Omitted	Omitted
Household Type 6	Omitted	Omitted	N/A	Omitted	Omitted
Household Type 7	1.160***	Omitted	N/A	1.160***	-0.003
	(0.215)			(0.232)	(0.440)
Region	Omitted	N/A	N/A	Omitted	Omitted
Fixed Effect	F =	F=	F =	F =	F =
	11.420***	11.080***	5.700***	4.200***	5.000***
Tobacco Grow	0.119	0.302	0.109	0.158	0.084
	(0.074)	(0.200)	(0.088)	(0.109)	(0.091)
Sample size	**13661**	**7276**	**6385**	**7510**	**6151**

Notes: The table shows results of an OLS regression of $\ln M$ on its instrumental variable the logarithm of assets and other control variables as discussed in Section 2.4. Standard errors are in parenthesis. *, **, *** signify statistical significance at the 10%, 5% and 1% levels respectively. I also report the F test for the joint significance of the cluster fixed effects. In all data specifications, the logarithm of the value of household assets explains a substantial proportion of the variation in $\ln M$. The F statistics associated with the instrumental variable are equal to 3364, 1971, 1127, 1444 and 560 in the full, urban, rural, top 50% and bottom 50% samples respectively.

Table A 9: First-stage ordinary least squares (OLS) regression results for *(*ln*M)*^{2}

Coefficient on:	Full Sample	Urban	Rural	Top 50%	Bottom 50%
Log of Assets Squared	0.224***	0.216***	0.228***	0.186***	0.132***
	(0.004)	(0.005)	(0.006)	(0.004)	(0.005)
Proportion of Adults	1.760	1.360	2.600	1.440	1.380
	(1.130)	(1.430)	(1.830)	(1.330)	(1.560)
Household Size	5.240***	4.980***	5.730***	4.260***	5.190***
	(0.477)	(0.618)	(0.746)	(0.570)	(0.649)
Head School	3.280***	5.100***	2.030***	4.170***	1.280**
	(0.460)	(0.722)	(0.605)	(0.662)	(0.523)
Head Wage Employment	5.670***	3.930***	10.500***	2.820***	2.370***
	(0.336)	(0.371)	(0.735)	(0.350)	(0.692)
Household Employment	0.608***	1.490***	-0.392	1.150***	-0.264
	(0.175)	(0.232)	(0.266)	(0.207)	(0.242)
Adult Age	-4.020***	-0.281	-8.900***	0.024	-4.680***
	(1.140)	(1.520)	(1.720)	(1.410)	(1.490)
Child Age	0.978***	0.676**	1.290***	0.368	0.288
	(0.211)	(0.276)	(0.324)	(0.265)	(0.267)
Household Head Age	-0.069	-2.410*	3.370*	-2.440**	0.531
	(1.020)	(1.350)	(1.570)	(1.240)	(1.360)
Most Educated	5.720***	6.720***	4.960***	5.490***	3.470***
	(0.024)	(0.974)	(0.762)	(0.924)	(0.654)
Household Type 2	5.390***	N/A	5.360***	0.838	3.280***
	(0.633)		(0.676)	(1.050)	(0.718)
Household Type 3	19.640***	N/A	19.700***	15.100***	2.940
	(2.660)		(2.780)	(3.010)	(5.190)
Household Type 4	2.910***	N/A	2.390***	2.810**	0.480
	(0.628)		(0.661)	(1.220)	(0.633)
Household Type 5	Omitted	N/A	N/A	Omitted	Omitted
Household Type 6	Omitted	Omitted	N/A	Omitted	Omitted
Household Type 7	32.270***	Omitted	N/A	32.300***	0.226
	(5.250)			(6.350)	(9.530)
Region	Omitted	N/A	N/A	Omitted	Omitted
Fixed Effect	F = 10.970***	F= 10.570***	F = 5.600***	F = 4.200***	F = 5.000***
Tobacco Grow	2.870	8.810*	2.480	4.540	1.730
	(1.830)	(5.300)	(0.088)	(2.970)	(1.980)
Constant	92.700***	97.800***	89.400***	110.000***	105.000***
	(2.470)	(3.590)	(3.390)	(3.520)	(2.880)
Sample size	**13661**	**7276**	**6385**	**7510**	**6151**

Notes: The table shows results of an OLS regression of $(\ln M)^2$ on its instrumental variable, the square of the logarithm of assets and other control variables as discussed in Section 2.4. Standard errors are in parenthesis. *, **, *** signify statistical significance at the 10%, 5% and 1% levels respectively. I also report the F test for the joint significance of the cluster fixed effects. In all data specifications, the square of the logarithm of the value of household assets explains a substantial proportion of the variation in $(\ln M)^2$. The F statistics associated with the instrumental variable are equal to 3,136 1866, 1444, 2162 and 697 in the Full, urban, rural, top 50% and bottom 50% samples respectively.

Table A 10: Three-stage Least Squares, Full Sample

Variable	Food	Alc	Health	Sch	Clothing	Housing	Water	Electricity	Alt. Energ	Daily Tprt	Other Tprt	Equpt	Ent	Tele	Remit	Hse Care	Pers. Cre
d	-5.061***	9.696***	0.0118	-1.975**	-2.665***	-0.366	-0.795***	-0.479	1.788***	0.177	-0.661	0.405	0.398	0.687	0.627**	0.213	-1.566*
	(1.634)	(0.668)	(0.337)	(0.863)	(0.890)	(0.675)	(0.273)	(0.341)	(0.554)	(0.586)	(0.652)	(0.294)	(0.260)	(0.487)	(0.261)	(0.222)	(0.872)
LnM	26.55***	-5.832**	-1.624	-3.557	10.50***	5.881**	-0.754	-4.396***	9.116***	-28.57***	8.146***	-2.838**	-6.110***	0.361	-0.219	-5.036***	-10.06***
	(6.477)	(2.647)	(1.334)	(3.420)	(3.528)	(2.676)	(1.081)	(1.350)	(2.195)	(2.324)	(2.583)	(1.165)	(1.032)	(1.929)	(1.034)	(0.882)	(3.455)
LnM2	-1.360***	0.214**	0.0619	0.195	-0.429***	-0.232**	0.0424	0.233***	-0.435***	1.241***	-0.279***	0.157***	0.266***	0.0806	0.0380	0.230***	0.330**
	(0.250)	(0.102)	(0.0515)	(0.132)	(0.136)	(0.103)	(0.0417)	(0.0521)	(0.0847)	(0.0897)	(0.0997)	(0.0450)	(0.0398)	(0.0745)	(0.0399)	(0.0340)	(0.133)
Adult prop	-3.056*	0.960	0.259	1.307	1.422*	0.726	0.425	-0.241	0.00894	0.911	-0.431	-0.147	-0.168	-0.140	-1.018***	-0.349	-1.620*
	(1.587)	(0.648)	(0.327)	(0.838)	(0.864)	(0.656)	(0.265)	(0.331)	(0.538)	(0.569)	(0.633)	(0.285)	(0.253)	(0.473)	(0.253)	(0.216)	(0.846)
loghhsize	-0.436	-0.0387	0.242*	3.334***	0.746**	-1.192***	-0.0609	-0.349**	-0.0305	-0.0468	-0.566**	0.166	-0.241**	-0.557***	-0.504***	-0.414***	-0.308
	(0.673)	(0.275)	(0.139)	(0.355)	(0.366)	(0.278)	(0.112)	(0.140)	(0.228)	(0.241)	(0.268)	(0.121)	(0.107)	(0.200)	(0.107)	(0.0916)	(0.359)
Head Sch	-0.0615	-0.144	-0.205	1.007***	-0.233	0.521**	0.0512	0.145	-0.281	-0.139	-0.183	0.0811	-0.0557	0.0505	0.0796	-0.170**	-0.415
	(0.616)	(0.252)	(0.127)	(0.325)	(0.336)	(0.255)	(0.103)	(0.128)	(0.209)	(0.221)	(0.246)	(0.111)	(0.0982)	(0.184)	(0.0984)	(0.0839)	(0.329)
Head w age	0.172	0.563***	-0.161	0.160	-0.684***	2.075***	-0.214***	-0.419***	-0.0296	-1.125***	-0.506***	-0.911***	-0.0266	0.168	-0.0235	-0.0524	1.369***
	(0.482)	(0.197)	(0.0992)	(0.254)	(0.262)	(0.199)	(0.0804)	(0.100)	(0.163)	(0.173)	(0.192)	(0.0867)	(0.0767)	(0.143)	(0.0769)	(0.0656)	(0.257)
Hhold Emp	0.195	0.0849	0.0227	-0.815***	0.347***	-0.130	-0.0879**	-0.189***	0.0261	0.0205	0.0264	0.0201	0.0511	0.0152	0.127***	0.0264	0.231**
	(0.215)	(0.0880)	(0.0444)	(0.114)	(0.117)	(0.0890)	(0.0359)	(0.0449)	(0.0730)	(0.0773)	(0.0859)	(0.0387)	(0.0343)	(0.0642)	(0.0344)	(0.0293)	(0.115)
Adult Age	3.111*	-0.0858	0.728**	-8.060***	0.582	1.098*	-0.158	-0.632*	0.592	1.080*	0.228	0.501*	0.399	0.442	-0.205	0.306	-0.620
	(1.598)	(0.653)	(0.329)	(0.844)	(0.870)	(0.660)	(0.267)	(0.333)	(0.541)	(0.573)	(0.637)	(0.287)	(0.255)	(0.476)	(0.255)	(0.218)	(0.852)
Child age	-0.514*	-0.136	0.0591	2.314***	-0.682***	-0.420***	0.00827	0.0746	-0.141	-0.0248	-0.0418	-0.0993*	-0.0211	-0.0132	-0.0881*	-0.0117	-0.290*
	(0.290)	(0.119)	(0.0597)	(0.153)	(0.158)	(0.120)	(0.0484)	(0.0604)	(0.0983)	(0.104)	(0.116)	(0.0522)	(0.0462)	(0.0864)	(0.0463)	(0.0395)	(0.155)
Head Age	-2.565*	-1.031*	-0.338	10.30***	-2.379***	-2.286***	0.650***	0.973***	0.00349	-1.138**	0.588	-0.458*	-0.404*	-0.683	0.0955	-0.149	-0.392
	(1.417)	(0.579)	(0.292)	(0.748)	(0.772)	(0.585)	(0.236)	(0.295)	(0.480)	(0.508)	(0.565)	(0.255)	(0.226)	(0.422)	(0.226)	(0.193)	(0.756)
Most Edu	0.577	0.797**	-0.0221	-0.383	-0.855*	0.333	-0.0856	-0.348*	-0.0197	-0.376	-0.131	-0.711***	0.112	-0.210	-0.0589	0.188	1.239***
	(0.859)	(0.351)	(0.177)	(0.453)	(0.468)	(0.355)	(0.143)	(0.179)	(0.291)	(0.308)	(0.342)	(0.154)	(0.137)	(0.256)	(0.137)	(0.117)	(0.458)
2.stratum	-1.009	-0.364	0.118	0.919**	0.278	0.223	-0.378**	-0.845***	0.102	1.315***	-0.0134	0.190	-0.126	-0.396	-0.00863	0.167	-0.143
	(0.882)	(0.360)	(0.182)	(0.466)	(0.480)	(0.364)	(0.147)	(0.184)	(0.299)	(0.316)	(0.352)	(0.159)	(0.140)	(0.263)	(0.141)	(0.120)	(0.470)
3.stratum	0.767	-1.696	-0.300	-1.769	-3.910*	0.996	-0.967	-2.668***	2.099	6.046***	-3.165**	1.927***	-1.346**	2.346**	2.757***	-0.0135	-1.751
	(3.819)	(1.561)	(0.787)	(2.017)	(2.080)	(1.578)	(0.637)	(0.796)	(1.294)	(1.370)	(1.523)	(0.687)	(0.608)	(1.138)	(0.610)	(0.520)	(2.037)
4.stratum	7.446***	-1.224***	0.275	-0.732*	-2.829***	-0.591*	0.0315	-0.148	0.411	0.0621	0.441	-0.608***	-0.0599	-0.183	-0.171	0.00814	-1.892***
	(0.837)	(0.342)	(0.172)	(0.442)	(0.456)	(0.346)	(0.140)	(0.175)	(0.284)	(0.300)	(0.334)	(0.151)	(0.133)	(0.249)	(0.134)	(0.114)	(0.447)
5.stratum	6.829***	0.723**	-0.203	-2.089***	-0.994**	3.436***	0.788***	0.469**	-0.225	2.423***	0.0315	0.533***	-0.188*	-0.521**	0.286**	0.0197	-1.167***
	(0.683)	(0.364)	(0.141)	(0.360)	(0.485)	(0.282)	(0.114)	(0.186)	(0.302)	(0.320)	(0.355)	(0.128)	(0.109)	(0.203)	(0.142)	(0.0929)	(0.364)
6.stratum	7.302***	0.491	-0.303	-1.621***	-0.599	4.419***	1.894***	1.026***	-0.665**	0.460	-0.0577	Omitted	-0.486***	-0.952***	-0.197	-0.0341	-0.0473
	(0.939)	(0.384)	(0.194)	(0.496)	(0.512)	(0.388)	(0.157)	(0.196)	(0.318)	(0.337)	(0.375)		(0.150)	(0.280)	(0.150)	(0.128)	(0.501)
7.stratum	7.473***	Omitted	-0.373*	-2.226***	Omitted	3.440***	1.226***	Omitted	Omitted	Omitted	Omitted	0.186	0.584***	-0.513	Omitted	0.944***	-0.648
	(1.051)		(0.217)	(0.555)		(0.434)	(0.175)					(0.169)	(0.167)	(0.313)		(0.143)	(0.561)
Cluster	4.06e-05	-0.000462*	-2.43e-05	0.00151***	0.00106***	-0.000324	3.16e-05	-0.000465***	-0.000915***	8.13e-05	-0.000322	-0.000212*	0.000205*	0.000948***	8.82e-05	-0.000210**	-0.00148***
	(0.000660)	(0.000270)	(0.000136)	(0.000348)	(0.000359)	(0.000273)	(0.000110)	(0.000138)	(0.000224)	(0.000237)	(0.000263)	(0.000119)	(0.000105)	(0.000197)	(0.000105)	(8.99e-05)	(0.000352)
Tobacco	-1.426	-0.302	0.146	0.198	2.374*	0.416	-0.152	-0.657	-1.174	0.349	-0.316	0.659	0.562	-0.603	-0.107	-0.359	0.470
	(2.310)	(0.944)	(0.476)	(1.220)	(1.258)	(0.954)	(0.385)	(0.482)	(0.783)	(0.829)	(0.921)	(0.416)	(0.368)	(0.688)	(0.369)	(0.315)	(1.232)
Cons	-70.59*	42.64**	10.27	-0.0729	-43.03*	-31.07*	1.586	19.84**	-43.19***	164.8***	-55.74***	14.00*	35.22***	-12.98	-0.713	27.78***	89.40***
	(41.34)	(16.90)	(8.517)	(21.83)	(22.52)	(17.08)	(6.897)	(8.617)	(14.01)	(14.83)	(16.48)	(7.436)	(6.585)	(12.31)	(6.599)	(5.628)	(22.05)
Obs.	8,555	8,555	8,555	8,555	8,555	8,555	8,555	8,555	8,555	8,555	8,555	8,555	8,555	8,555	8,555	8,555	8,555
R-squared	0.172	0.039	0.015	0.207	0.117	0.153	0.089	0.092	0.078	0.173	0.029	0.002	0.081	0.148	0.030	0.139	0.049

Table A 11: Three-stage Least Squares, Urban Sample

Variable	Food	Alc	Health	Sch	Clothing	Housing	Water	Electricity	Alt. Energ	Daily Tprt	Other Tprt	Equpt	Ent	Tele	Remit	Hse Care	Pers. Cre
d	-0.728	8.656***	0.267	-2.925*	-2.751**	-1.829	-2.228***	-2.164***	1.251	0.256	-0.249	1.038**	0.817	2.099*	1.490***	-0.417	-1.788
	(2.543)	(1.019)	(0.434)	(1.523)	(1.194)	(1.666)	(0.679)	(0.813)	(0.943)	(1.264)	(1.032)	(0.449)	(0.620)	(1.111)	(0.497)	(0.492)	(1.310)
LnM	32.17**	-7.443	-0.222	-1.429	10.06	16.23*	3.475	13.02***	-12.97***	-53.64***	6.456	-16.27***	-13.08***	27.71***	1.528	-6.976***	4.414
	(13.38)	(5.362)	(2.283)	(8.011)	(6.279)	(8.763)	(3.569)	(4.274)	(4.960)	(6.651)	(5.429)	(2.363)	(3.259)	(5.841)	(2.616)	(2.587)	(6.891)
LnM2	-1.558***	0.270	0.00811	0.123	-0.387*	-0.615*	-0.115	-0.406**	0.368**	2.156***	-0.228	0.652***	0.530***	-0.926***	-0.0273	0.300***	-0.212
	(0.497)	(0.199)	(0.0849)	(0.298)	(0.233)	(0.326)	(0.133)	(0.159)	(0.184)	(0.247)	(0.202)	(0.0878)	(0.121)	(0.217)	(0.0972)	(0.0962)	(0.256)
Adult prop	-2.786	1.330*	0.698**	0.291	0.134	1.452	0.755	-0.0159	-0.400	0.824	-0.326	-0.481	-0.433	0.246	-0.910**	-0.718**	-0.327
	(1.811)	(0.726)	(0.309)	(1.085)	(0.850)	(1.187)	(0.483)	(0.579)	(0.672)	(0.901)	(0.735)	(0.320)	(0.441)	(0.791)	(0.354)	(0.350)	(0.933)
loghhsize	0.986	0.100	0.181	3.019***	0.194	-1.775***	-0.193	-0.499**	0.555*	-0.416	-0.171	-0.0819	-0.432**	-0.862**	-0.701***	-0.517***	0.316
	(0.772)	(0.310)	(0.132)	(0.463)	(0.363)	(0.506)	(0.206)	(0.247)	(0.286)	(0.384)	(0.313)	(0.136)	(0.188)	(0.337)	(0.151)	(0.149)	(0.398)
Head Sch	-1.508*	-0.0825	-0.471***	0.958*	0.332	1.443**	0.0185	0.148	-0.801**	-0.340	0.629*	0.175	-0.121	0.195	-0.0876	-0.581***	0.334
	(0.861)	(0.345)	(0.147)	(0.515)	(0.404)	(0.564)	(0.230)	(0.275)	(0.319)	(0.428)	(0.349)	(0.152)	(0.210)	(0.376)	(0.168)	(0.166)	(0.443)
Head w age	-0.768	0.492***	-0.130	0.280	-0.219	2.004***	-0.245*	-0.521***	0.345*	-0.647***	-0.592***	-0.452***	0.0257	-0.197	-0.0925	-0.0259	1.009***
	(0.476)	(0.191)	(0.0812)	(0.285)	(0.223)	(0.312)	(0.127)	(0.152)	(0.176)	(0.236)	(0.193)	(0.0840)	(0.116)	(0.208)	(0.0930)	(0.0920)	(0.245)
Hhold Emp	0.184	0.174	0.0414	-0.779***	0.489***	-0.265	-0.0769	-0.277***	-0.147	-0.0529	0.0510	-0.0193	0.0458	0.212*	0.167***	0.0323	0.134
	(0.269)	(0.108)	(0.0458)	(0.161)	(0.126)	(0.176)	(0.0717)	(0.0858)	(0.0996)	(0.134)	(0.109)	(0.0475)	(0.0654)	(0.117)	(0.0525)	(0.0520)	(0.138)
Adult Age	2.425	0.291	1.220***	-7.071***	-0.0215	1.228	-0.00510	-0.805	0.793	0.0707	1.009	-0.335	0.276	0.748	-0.0703	0.438	-0.811
	(1.944)	(0.779)	(0.332)	(1.164)	(0.912)	(1.273)	(0.519)	(0.621)	(0.721)	(0.966)	(0.789)	(0.343)	(0.473)	(0.849)	(0.380)	(0.376)	(1.001)
Child age	-0.669**	-0.0655	0.0734	2.247***	-0.251	-0.852***	0.00567	0.0854	-0.179	0.0955	-0.0574	8.93e-05	0.00127	-0.0397	-0.0607	-0.0442	-0.335*
	(0.339)	(0.136)	(0.0578)	(0.203)	(0.159)	(0.222)	(0.0903)	(0.108)	(0.126)	(0.168)	(0.137)	(0.0598)	(0.0825)	(0.148)	(0.0662)	(0.0655)	(0.174)
Head Age	-1.943	-1.515**	-0.494*	9.218***	-1.012	-4.080***	1.285***	1.685***	0.614	-0.959	0.108	0.185	-0.471	-0.978	0.0495	-0.295	-0.466
	(1.663)	(0.666)	(0.284)	(0.996)	(0.780)	(1.089)	(0.444)	(0.531)	(0.616)	(0.827)	(0.675)	(0.294)	(0.405)	(0.726)	(0.325)	(0.322)	(0.857)
Most Edu	0.642	0.935*	0.632***	-0.614	-1.957***	0.501	-0.0719	-1.022**	0.572	0.399	-0.506	-0.577**	0.376	-0.945	0.0981	0.923***	0.208
	(1.320)	(0.529)	(0.225)	(0.791)	(0.620)	(0.865)	(0.352)	(0.422)	(0.490)	(0.656)	(0.536)	(0.233)	(0.322)	(0.577)	(0.258)	(0.255)	(0.680)
6.stratum	0.605	-0.194	-0.145	0.418	-0.177	0.867**	1.017***	0.314	-0.00861	-1.496***	0.0535	-0.456***	-0.338**	-0.736***	-0.459***	-0.0121	0.998***
	(0.609)	(0.244)	(0.104)	(0.364)	(0.286)	(0.399)	(0.162)	(0.194)	(0.226)	(0.303)	(0.247)	(0.108)	(0.148)	(0.266)	(0.119)	(0.118)	(0.314)
7.stratum	0.969	-0.667**	-0.194	-0.208	0.188	-0.0293	0.474**	-0.335	0.349	-2.444***	0.157	-0.584***	0.510***	0.217	-0.238	0.946***	0.794**
	(0.780)	(0.313)	(0.133)	(0.467)	(0.366)	(0.511)	(0.208)	(0.249)	(0.289)	(0.388)	(0.317)	(0.138)	(0.190)	(0.341)	(0.153)	(0.151)	(0.402)
Cluster	-6.31e-05	-0.000805**	-6.62e-05	0.00223***	0.00132***	-0.000579	-9.52e-05	-0.000744***	-0.00141***	-0.000657	-0.000180	-0.000111	0.000306	0.00202***	0.000340**	-0.000424***	-0.000835**
	(0.000814)	(0.000326)	(0.000139)	(0.000488)	(0.000382)	(0.000534)	(0.000217)	(0.000260)	(0.000302)	(0.000405)	(0.000331)	(0.000144)	(0.000198)	(0.000356)	(0.000159)	(0.000158)	(0.000420)
Tobacco	-5.535	-2.021	-0.0399	0.139	0.204	5.689	-0.282	-2.401	-0.0952	8.740***	-0.0992	0.924	-1.625	-0.789	-0.570	-1.648	-0.596
	(5.483)	(2.198)	(0.936)	(3.283)	(2.573)	(3.592)	(1.463)	(1.752)	(2.033)	(2.726)	(2.225)	(0.968)	(1.335)	(2.394)	(1.072)	(1.060)	(2.824)
Cons	-102.5	53.77	-1.252	-15.47	-51.80	-91.47	-28.46	-98.24***	106.6***	337.3***	-46.53	103.2***	81.30***	-197.1***	-13.23	40.99**	-7.869
	(89.15)	(35.73)	(15.21)	(53.38)	(41.84)	(58.40)	(23.79)	(28.48)	(33.05)	(44.32)	(36.18)	(15.75)	(21.71)	(38.92)	(17.43)	(17.24)	(45.92)
Obs.	4,545	4,545	4,545	4,545	4,545	4,545	4,545	4,545	4,545	4,545	4,545	4,545	4,545	4,545	4,545	4,545	4,545
R-squared	0.308	0.016	0.012	0.199	-0.024	0.074	-0.000	-0.016	0.139	0.149	0.011	-0.052	0.079	0.017	-0.004	0.107	0.018

Table A 12: Three-stage Least Squares, Rural Sample

Variable	Food	Alc	Health	Sch	Clothing	Housing	Water	Electricity	Alt. Energ	Daily Tprt	Other Tprt	Equpt	Ent	Tele	Remit	Hse Care	Pers. Cre
d	-7.086***	10.06***	-0.0644	-1.662	-2.836**	0.134	-0.119	0.132	2.205***	0.305	-0.761	0.195	0.220	0.0606	0.279	0.390**	-1.116
	(2.250)	(0.928)	(0.507)	(1.084)	(1.328)	(0.177)	(0.111)	(0.221)	(0.733)	(0.563)	(0.884)	(0.423)	(0.157)	(0.410)	(0.303)	(0.192)	(1.227)
LnM	29.76**	-5.094	-1.309	-7.040	29.17***	-0.363	-0.680	-7.627***	10.64**	-27.50***	3.892	1.252	-0.300	-9.989***	-2.699	-5.397***	-17.28**
	(13.60)	(5.612)	(3.062)	(6.555)	(8.029)	(1.072)	(0.670)	(1.337)	(4.434)	(3.405)	(5.346)	(2.557)	(0.948)	(2.477)	(1.833)	(1.158)	(7.420)
LnM2	-1.509***	0.186	0.0479	0.334	-1.253***	0.0169	0.0316	0.342***	-0.455**	1.243***	-0.0827	0.00403	0.0184	0.483***	0.142*	0.248***	0.627**
	(0.562)	(0.232)	(0.127)	(0.271)	(0.332)	(0.0443)	(0.0277)	(0.0553)	(0.183)	(0.141)	(0.221)	(0.106)	(0.0392)	(0.102)	(0.0758)	(0.0479)	(0.307)
Adult prop	-3.440	0.352	-0.307	3.064**	3.230*	0.0386	0.137	0.108	-0.145	0.0463	-0.456	-0.251	-0.100	0.404	-1.044***	0.0162	-2.989*
	(2.801)	(1.156)	(0.631)	(1.350)	(1.654)	(0.221)	(0.138)	(0.275)	(0.913)	(0.701)	(1.101)	(0.527)	(0.195)	(0.510)	(0.378)	(0.238)	(1.528)
loghhsize	-2.397**	-0.211	0.252	4.012***	1.381**	-0.111	0.0335	-0.0745	-0.904**	0.163	-0.928**	0.314	-0.0553	-0.0476	-0.304*	-0.272***	-1.012
	(1.160)	(0.479)	(0.261)	(0.559)	(0.685)	(0.0914)	(0.0571)	(0.114)	(0.378)	(0.290)	(0.456)	(0.218)	(0.0809)	(0.211)	(0.156)	(0.0987)	(0.633)
Head Sch	0.714	-0.178	-0.0753	1.131***	-0.499	-0.0411	0.0187	0.0716	0.0457	0.000993	-0.630*	0.0611	0.00795	-0.0596	0.172	0.0699	-0.900*
	(0.889)	(0.367)	(0.200)	(0.429)	(0.525)	(0.0701)	(0.0438)	(0.0874)	(0.290)	(0.223)	(0.350)	(0.167)	(0.0620)	(0.162)	(0.120)	(0.0757)	(0.485)
Head w age	3.689***	0.752	-0.307	-0.453	-0.961	0.694***	-0.0221	-0.429***	-0.179	-2.103***	-0.542	-1.816***	0.161*	0.155	0.0611	-0.145	1.802***
	(1.239)	(0.511)	(0.279)	(0.597)	(0.732)	(0.0977)	(0.0610)	(0.122)	(0.404)	(0.310)	(0.487)	(0.233)	(0.0864)	(0.226)	(0.167)	(0.105)	(0.676)
Hhold Emp	0.308	0.0113	0.0145	-0.919***	0.201	-0.0324	-0.0424**	-0.0141	0.142	-0.0194	-0.00448	-0.00382	0.0204	-0.104*	0.0853*	0.0111	0.415**
	(0.345)	(0.142)	(0.0777)	(0.166)	(0.204)	(0.0272)	(0.0170)	(0.0339)	(0.112)	(0.0864)	(0.136)	(0.0649)	(0.0241)	(0.0628)	(0.0465)	(0.0294)	(0.188)
Adult Age	4.534*	-0.505	0.268	-9.295***	1.276	-0.189	0.0615	0.140	0.781	1.447**	-0.389	1.049**	0.168	0.366	-0.349	0.127	-0.0461
	(2.594)	(1.070)	(0.584)	(1.250)	(1.532)	(0.204)	(0.128)	(0.255)	(0.846)	(0.649)	(1.020)	(0.488)	(0.181)	(0.473)	(0.350)	(0.221)	(1.415)
Child age	-0.454	-0.229	0.0464	2.425***	-1.036***	0.0906**	0.00217	0.0621	-0.124	-0.162	-0.0376	-0.180**	-0.0172	-1.71e-05	-0.125*	0.0258	-0.288
	(0.482)	(0.199)	(0.109)	(0.232)	(0.285)	(0.0380)	(0.0238)	(0.0474)	(0.157)	(0.121)	(0.190)	(0.0907)	(0.0336)	(0.0879)	(0.0650)	(0.0411)	(0.263)
Head Age	-3.800	-0.467	-0.0967	11.46***	-3.467**	-0.184	-0.0692	-0.112	-0.778	-0.958	0.968	-0.856*	-0.0900	-0.540	0.188	-0.00158	-0.650
	(2.360)	(0.974)	(0.531)	(1.138)	(1.393)	(0.186)	(0.116)	(0.232)	(0.769)	(0.591)	(0.928)	(0.444)	(0.165)	(0.430)	(0.318)	(0.201)	(1.288)
Most Edu	0.383	0.734	-0.349	-0.351	-0.210	0.199**	-0.0624	-0.0818	-0.232	-0.645**	0.0767	-0.660***	0.0235	0.0154	-0.157	-0.192*	1.772***
	(1.189)	(0.491)	(0.268)	(0.573)	(0.702)	(0.0937)	(0.0586)	(0.117)	(0.388)	(0.298)	(0.467)	(0.224)	(0.0829)	(0.217)	(0.160)	(0.101)	(0.649)
2.stratum	-0.329	-0.360	0.183	0.678	1.218**	-0.0920	-0.140***	-0.363***	-0.405	0.441*	-0.266	-0.118	-0.106	-0.0838	-0.101	0.0790	0.0723
	(1.022)	(0.422)	(0.230)	(0.493)	(0.604)	(0.0806)	(0.0504)	(0.101)	(0.333)	(0.256)	(0.402)	(0.192)	(0.0713)	(0.186)	(0.138)	(0.0870)	(0.558)
3.stratum	3.143	-1.660	-0.161	-2.600	0.947	-0.163	-0.380	-1.697***	0.561	3.621***	-4.558**	1.460	-0.396	2.008**	2.252***	-0.265	-2.215
	(4.897)	(2.021)	(1.103)	(2.360)	(2.891)	(0.386)	(0.241)	(0.481)	(1.596)	(1.226)	(1.925)	(0.921)	(0.341)	(0.892)	(0.660)	(0.417)	(2.672)
4.stratum	6.490***	-1.250***	0.306	-0.497	-2.670***	0.232***	0.0320	-0.0190	0.0312	0.119	0.282	-0.543***	0.00700	-0.0707	-0.185	0.0480	-1.990***
	(0.978)	(0.403)	(0.220)	(0.471)	(0.577)	(0.0771)	(0.0482)	(0.0961)	(0.319)	(0.245)	(0.384)	(0.184)	(0.0682)	(0.178)	(0.132)	(0.0832)	(0.534)
Cluster	0.000155	-0.000148	0.000110	0.000732	0.00120*	8.32e-05	0.000257***	0.000160	-0.000872**	0.000311	-0.000583	-0.000476**	6.53e-05	0.000269	-0.000166	-2.49e-05	-0.00203***
	(0.00110)	(0.000454)	(0.000248)	(0.000530)	(0.000649)	(8.67e-05)	(5.42e-05)	(0.000108)	(0.000359)	(0.000275)	(0.000432)	(0.000207)	(7.67e-05)	(0.000200)	(0.000148)	(9.36e-05)	(0.000600)
Tobacco	-0.476	0.00740	0.196	0.0420	2.289	-0.0673	-0.0486	-0.0584	-1.489	-1.058	-0.428	0.319	0.730***	-0.445	-0.0619	-0.133	0.782
	(2.852)	(1.177)	(0.642)	(1.375)	(1.684)	(0.225)	(0.140)	(0.280)	(0.930)	(0.714)	(1.121)	(0.536)	(0.199)	(0.520)	(0.384)	(0.243)	(1.556)
Cons	-86.05	37.97	9.890	19.81	-148.1***	2.999	3.715	42.16***	-55.06**	150.7***	-30.94	-13.04	0.836	51.66***	14.14	29.34***	133.4***
	(81.69)	(33.71)	(18.39)	(39.37)	(48.23)	(6.439)	(4.024)	(8.031)	(26.63)	(20.45)	(32.11)	(15.36)	(5.695)	(14.88)	(11.01)	(6.954)	(44.57)
Obs.	4,010	4,010	4,010	4,010	4,010	4,010	4,010	4,010	4,010	4,010	4,010	4,010	4,010	4,010	4,010	4,010	4,010
R-squared	0.040	0.055	0.010	0.211	-0.013	0.046	0.013	-0.004	-0.022	0.095	0.053	-0.053	0.012	0.111	0.052	0.023	0.035

Table A 13: Three-stage Least Squares, Top 50%

Variable	Food	Alc	Health	Sch	Clothing	Housing	Water	Electricity	Alt. Energ	Daily Tprt	Other Tprt	Equpt	Ent	Tele	Remit	Hse Care	Pers. Cre
d	-4.275*	8.672***	0.769*	-0.735	0.661	-2.760*	-2.187***	-1.684*	-3.045***	0.388	0.191	1.949***	0.451	3.861***	2.050***	-1.114**	-3.148**
	(2.504)	(1.068)	(0.447)	(1.609)	(1.143)	(1.611)	(0.600)	(0.874)	(0.843)	(1.363)	(1.269)	(0.520)	(0.636)	(1.199)	(0.573)	(0.522)	(1.246)
LnM	-55.32	-32.12*	-4.884	34.84	45.86***	24.40	9.749	71.36***	-83.32***	-100.2***	27.83	-22.37***	-15.31	117.1***	22.85***	-12.56	7.759
	(38.68)	(16.50)	(6.907)	(24.84)	(17.65)	(24.88)	(9.270)	(13.49)	(13.01)	(21.04)	(19.60)	(8.026)	(9.826)	(18.52)	(8.853)	(8.067)	(19.24)
LnM2	1.572	1.152*	0.176	-1.199	-1.663***	-0.931	-0.335	-2.488***	2.887***	3.849***	-1.003	0.882***	0.607*	-4.127***	-0.788**	0.501*	-0.331
	(1.392)	(0.594)	(0.249)	(0.894)	(0.636)	(0.896)	(0.334)	(0.486)	(0.468)	(0.758)	(0.706)	(0.289)	(0.354)	(0.667)	(0.319)	(0.290)	(0.693)
Adult prop	-4.022**	-0.0406	0.664*	0.619	0.00958	2.163*	0.857*	0.422	-0.295	1.320	-0.615	-0.174	-0.494	0.479	-1.273***	-0.655	0.477
	(1.950)	(0.832)	(0.348)	(1.253)	(0.890)	(1.255)	(0.468)	(0.680)	(0.656)	(1.061)	(0.988)	(0.405)	(0.496)	(0.934)	(0.446)	(0.407)	(0.970)
loghhsize	0.992	-0.314	0.0912	3.792***	-0.341	-1.582***	-0.133	-0.498*	0.926***	-0.145	-0.889**	0.0886	-0.520**	-0.653*	-0.784***	-0.576***	0.407
	(0.822)	(0.351)	(0.147)	(0.528)	(0.375)	(0.529)	(0.197)	(0.287)	(0.276)	(0.447)	(0.416)	(0.171)	(0.209)	(0.393)	(0.188)	(0.171)	(0.409)
Head Sch	-0.587	-0.275	-0.558***	2.103***	-0.121	1.342**	-0.300	0.0292	-0.671**	-0.204	0.704	0.167	-0.175	0.129	0.0747	-0.610***	-0.610
	(0.918)	(0.392)	(0.164)	(0.589)	(0.419)	(0.590)	(0.220)	(0.320)	(0.309)	(0.499)	(0.465)	(0.190)	(0.233)	(0.439)	(0.210)	(0.191)	(0.456)
Head w age	-0.00822	0.626***	-0.0699	0.174	-0.0315	1.677***	-0.159	-0.765***	0.259	-0.813***	-0.968***	-0.591***	0.0901	-0.210	-0.149	-0.0320	1.102***
	(0.519)	(0.221)	(0.0927)	(0.333)	(0.237)	(0.334)	(0.124)	(0.181)	(0.175)	(0.282)	(0.263)	(0.108)	(0.132)	(0.248)	(0.119)	(0.108)	(0.258)
Hhold Emp	0.268	0.0887	-0.00529	-0.765***	0.224*	-0.0554	-0.102	-0.218**	-0.0628	-0.130	0.125	-0.0894	0.0669	0.134	0.199***	0.00181	0.226*
	(0.273)	(0.117)	(0.0488)	(0.176)	(0.125)	(0.176)	(0.0655)	(0.0953)	(0.0920)	(0.149)	(0.139)	(0.0567)	(0.0694)	(0.131)	(0.0626)	(0.0570)	(0.136)
Adult Age	1.627	-0.370	0.907**	-8.260***	0.233	2.254*	0.331	-0.0613	0.324	0.464	0.813	-0.252	0.406	1.034	-0.199	0.680	0.00803
	(2.094)	(0.893)	(0.374)	(1.345)	(0.956)	(1.347)	(0.502)	(0.730)	(0.705)	(1.139)	(1.061)	(0.435)	(0.532)	(1.003)	(0.479)	(0.437)	(1.042)
Child age	-0.693*	-0.372**	0.0878	2.461***	-0.323*	-0.931***	-0.00535	0.0739	-0.0605	0.0618	-0.136	0.0636	0.0117	-0.0355	-0.106	-0.0392	-0.125
	(0.378)	(0.161)	(0.0676)	(0.243)	(0.173)	(0.244)	(0.0907)	(0.132)	(0.127)	(0.206)	(0.192)	(0.0785)	(0.0962)	(0.181)	(0.0866)	(0.0789)	(0.188)
Head Age	-1.466	-0.467	-0.0989	10.37***	-0.688	-4.865***	0.988**	1.405**	0.283	-1.416	1.052	-0.0371	-0.524	-1.685**	0.0981	-0.444	-2.007**
	(1.780)	(0.760)	(0.318)	(1.144)	(0.813)	(1.146)	(0.427)	(0.621)	(0.599)	(0.969)	(0.902)	(0.369)	(0.452)	(0.852)	(0.408)	(0.371)	(0.885)
Most Edu	-0.422	0.886	0.699***	-0.745	-0.968	0.559	0.132	-0.935*	0.690	-1.075	-0.765	-0.783***	0.325	-0.522	0.0270	1.097***	0.857
	(1.410)	(0.602)	(0.252)	(0.906)	(0.644)	(0.907)	(0.338)	(0.492)	(0.475)	(0.767)	(0.715)	(0.293)	(0.358)	(0.675)	(0.323)	(0.294)	(0.701)
2.stratum	-0.781	-0.0353	0.130	-0.144	-0.440	0.868	-0.755**	-1.377***	-0.469	3.188***	-1.472**	1.092***	-0.144	0.0147	0.0622	0.392	-0.531
	(1.395)	(0.595)	(0.249)	(0.896)	(0.637)	(0.898)	(0.334)	(0.487)	(0.469)	(0.759)	(0.707)	(0.290)	(0.354)	(0.668)	(0.319)	(0.291)	(0.694)
3.stratum	-3.452	-1.526	-0.500	-2.182	-2.171	2.234	-1.498	-3.409**	2.490*	10.42***	-4.609**	1.446*	-1.546	0.192	4.135***	0.836	-1.447
	(4.233)	(1.806)	(0.756)	(2.719)	(1.932)	(2.724)	(1.015)	(1.477)	(1.424)	(2.303)	(2.145)	(0.879)	(1.076)	(2.027)	(0.969)	(0.883)	(2.105)
4.stratum	6.751***	-0.818	-0.395	-2.698**	-0.973	-1.061	-0.205	-0.913	0.412	0.930	1.230	0.0256	-0.136	-0.685	-0.547	-0.0292	-1.283
	(1.793)	(0.765)	(0.320)	(1.151)	(0.818)	(1.153)	(0.430)	(0.625)	(0.603)	(0.975)	(0.909)	(0.372)	(0.455)	(0.858)	(0.410)	(0.374)	(0.892)
5.stratum	4.150***	0.751**	0.240	-2.938***	-0.425	3.980***	-0.630***	0.0342	0.474	3.340***	-3.294***	0.762***	-0.672***	-0.914**	0.517***	-0.937***	-0.379
	(0.896)	(0.381)	(0.159)	(0.576)	(0.407)	(0.577)	(0.214)	(0.311)	(0.300)	(0.485)	(0.454)	(0.185)	(0.227)	(0.427)	(0.150)	(0.186)	(0.444)
6.stratum	6.136***	0.537	0.0967	-2.781***	-0.315	4.697***	0.347	0.00885	0.319	1.699***	-3.257***	0.275	-0.964***	-1.954***	Omitted	-0.964***	0.394
	(1.105)	(0.418)	(0.175)	(0.710)	(0.447)	(0.711)	(0.235)	(0.342)	(0.330)	(0.533)	(0.560)	(0.203)	(0.249)	(0.469)		(0.204)	(0.487)
7.stratum	4.991***	Omitted	Omitted	-2.641***	Omitted	4.305***	Omitted	Omitted	Omitted	Omitted	-2.732***	Omitted	Omitted	Omitted	0.402*	Omitted	Omitted
	(1.171)			(0.752)		(0.753)					(0.593)				(0.224)		
Cluster	-0.000351	-0.000321	8.99e-06	0.00189***	0.000239	-8.75e-05	-0.000402**	-0.00103***	0.000177	0.000430	-0.000318	-0.000149	0.000463**	0.00129***	0.000305	-0.000239	-0.00178***
	(0.000810)	(0.000345)	(0.000145)	(0.000520)	(0.000370)	(0.000521)	(0.000194)	(0.000282)	(0.000272)	(0.000441)	(0.000410)	(0.000168)	(0.000206)	(0.000388)	(0.000185)	(0.000169)	(0.000403)
Tobacco	-5.837	-2.723*	0.139	4.676*	5.820***	0.402	-0.493	-0.488	-0.834	1.160	0.288	-0.750	-0.445	0.599	-0.428	-1.257	0.548
	(3.739)	(1.595)	(0.668)	(2.401)	(1.706)	(2.405)	(0.896)	(1.304)	(1.258)	(2.034)	(1.895)	(0.776)	(0.950)	(1.790)	(0.856)	(0.780)	(1.859)
Cons	504.6*	226.5**	30.70	-265.3	-300.4**	-148.9	-73.30	-506.5***	595.5***	659.1***	-190.8	144.5***	97.51	-818.9***	-161.3***	79.15	-27.35
	(266.7)	(113.8)	(47.62)	(171.3)	(121.7)	(171.6)	(63.92)	(93.00)	(89.72)	(145.1)	(135.1)	(55.34)	(67.74)	(127.7)	(61.04)	(55.62)	(132.6)
Obs.	4,092	4,092	4,092	4,092	4,092	4,092	4,092	4,092	4,092	4,092	4,092	4,092	4,092	4,092	4,092	4,092	4,092
R-squared	0.248	0.005	-0.002	0.193	-0.004	0.080	0.015	-0.089	-0.040	0.147	0.013	-0.066	0.081	-0.144	-0.040	0.064	0.017

Table A 14: Three-Stage Least Squares, Bottom 50%

Variable	Food	Alc	Health	Sch	Clothing	Housing	Water	Electricity	Alt. Energ	Daily Tprt	Other Tprt	Equpt	Ent	Tele	Remit	Hse Care	Pers. Cre
d	-6.315***	8.989***	-0.113	-2.096**	-3.174**	0.571**	-0.0950	0.114	2.204***	-0.325	1.504**	-1.136***	0.174	-0.218	0.196	0.464***	-0.517
	(2.235)	(0.956)	(0.505)	(0.935)	(1.411)	(0.269)	(0.209)	(0.0825)	(0.774)	(0.227)	(0.635)	(0.373)	(0.134)	(0.338)	(0.226)	(0.155)	(1.318)
LnM	121.9	12.01	-24.29	-36.87	76.27	-5.299	-5.798	-0.699	12.44	-10.64	6.245	8.403	2.540	-48.22***	-5.151	-7.765	-105.8**
	(90.30)	(38.62)	(20.42)	(37.79)	(57.03)	(10.88)	(8.427)	(3.334)	(31.26)	(9.173)	(25.64)	(15.07)	(5.413)	(13.64)	(9.125)	(6.282)	(53.26)
LnM2	-5.659	-0.615	1.104	1.704	-3.383	0.219	0.270	0.0346	-0.525	0.500	-0.260	-0.293	-0.113	2.216***	0.253	0.364	4.607*
	(4.067)	(1.739)	(0.920)	(1.702)	(2.568)	(0.490)	(0.380)	(0.150)	(1.408)	(0.413)	(1.155)	(0.679)	(0.244)	(0.614)	(0.411)	(0.283)	(2.399)
Adult prop	-5.380*	2.127	-0.473	3.263**	4.006**	0.105	0.469	-0.0831	0.331	0.107	-1.505*	-0.0485	0.134	0.518	-0.501	-0.00480	-3.479*
	(3.175)	(1.358)	(0.718)	(1.329)	(2.005)	(0.383)	(0.296)	(0.117)	(1.099)	(0.323)	(0.902)	(0.530)	(0.190)	(0.480)	(0.321)	(0.221)	(1.873)
loghhsize	-2.866**	0.305	0.0263	2.983***	2.192**	-0.102	0.0399	-0.0597	-0.948**	-0.0833	-0.565	0.127	0.0436	-0.0192	-0.162	-0.194**	-0.830
	(1.350)	(0.577)	(0.305)	(0.565)	(0.852)	(0.163)	(0.126)	(0.0498)	(0.467)	(0.137)	(0.383)	(0.225)	(0.0809)	(0.204)	(0.136)	(0.0939)	(0.796)
Head Sch	0.247	-0.0462	0.00404	0.364	0.113	-0.0997	0.0454	0.0307	0.0938	0.0582	-0.670**	0.221	0.0487	0.0203	0.164	0.00402	-0.719
	(1.003)	(0.429)	(0.227)	(0.420)	(0.633)	(0.121)	(0.0936)	(0.0370)	(0.347)	(0.102)	(0.285)	(0.167)	(0.0601)	(0.151)	(0.101)	(0.0698)	(0.591)
Head w age	3.269*	1.034	-0.851**	-1.584**	-1.685	1.057***	-0.288	-0.0749	1.011	-0.238	-0.146	-0.795**	-0.0448	-0.645**	-0.0133	-0.0809	0.279
	(1.913)	(0.818)	(0.433)	(0.801)	(1.208)	(0.230)	(0.179)	(0.0706)	(0.662)	(0.194)	(0.543)	(0.319)	(0.115)	(0.289)	(0.193)	(0.133)	(1.128)
Hhold Emp	0.377	0.0333	0.0412	-0.876***	0.363	-0.0524	-0.0441	-0.0206	0.00809	-0.0191	-0.00695	0.0704	0.0227	-0.0511	0.0332	0.0563**	0.182
	(0.404)	(0.173)	(0.0914)	(0.169)	(0.255)	(0.0487)	(0.0377)	(0.0149)	(0.140)	(0.0411)	(0.115)	(0.0675)	(0.0242)	(0.0611)	(0.0409)	(0.0281)	(0.239)
Adult Age	3.846	-0.0246	0.690	-6.085***	0.853	-0.0956	0.119	-0.101	1.123	-0.0955	-1.011	1.042**	0.230	0.515	-0.253	-0.0507	-0.876
	(2.958)	(1.265)	(0.669)	(1.238)	(1.868)	(0.356)	(0.276)	(0.109)	(1.024)	(0.300)	(0.840)	(0.494)	(0.177)	(0.447)	(0.299)	(0.206)	(1.744)
Child age	-0.419	-0.0726	0.0859	2.093***	-0.785**	-0.0717	0.0465	0.00953	-0.229	0.0238	-0.173	-0.117	-0.0118	0.0177	-0.0603	0.0331	-0.300
	(0.510)	(0.218)	(0.115)	(0.214)	(0.322)	(0.0615)	(0.0476)	(0.0188)	(0.177)	(0.0518)	(0.145)	(0.0852)	(0.0306)	(0.0771)	(0.0516)	(0.0355)	(0.301)
Head Age	-2.080	-1.172	-0.493	8.790***	-3.334**	-0.553*	-0.119	0.147	-0.656	0.189	1.024	-0.720	-0.176	-0.613	0.156	0.143	-0.230
	(2.684)	(1.148)	(0.607)	(1.123)	(1.695)	(0.323)	(0.251)	(0.0991)	(0.929)	(0.273)	(0.762)	(0.448)	(0.161)	(0.405)	(0.271)	(0.187)	(1.583)
Most Edu	0.348	0.632	-0.462	0.295	-0.988	0.365**	-0.167	-0.0529	-0.638	-0.111	0.257	-0.776***	-0.0204	-0.0598	-0.171	-0.181*	2.207***
	(1.349)	(0.577)	(0.305)	(0.564)	(0.852)	(0.162)	(0.126)	(0.0498)	(0.467)	(0.137)	(0.383)	(0.225)	(0.0808)	(0.204)	(0.136)	(0.0938)	(0.795)
2.stratum	-0.172	-0.314	-0.267	0.0449	1.367	0.185	-0.158	-0.0268	-0.445	-0.179	-0.0521	-0.190	-0.00530	-0.167	0.0917	0.123	0.311
	(1.468)	(0.628)	(0.332)	(0.614)	(0.927)	(0.177)	(0.137)	(0.0542)	(0.508)	(0.149)	(0.417)	(0.245)	(0.0880)	(0.222)	(0.148)	(0.102)	(0.866)
4.stratum	7.913***	-1.282**	0.277	-0.260	-3.089***	-0.113	0.0504	0.0195	-0.337	0.0775	0.329	-0.786***	-0.00306	-0.0934	-0.164	0.106	-2.297***
	(1.191)	(0.509)	(0.269)	(0.498)	(0.752)	(0.143)	(0.111)	(0.0440)	(0.412)	(0.121)	(0.338)	(0.199)	(0.0714)	(0.180)	(0.120)	(0.0828)	(0.702)
5.stratum	2.650	1.951	-2.033	-2.579	2.457	-0.899	-0.831	0.261	-2.195	0.372	0.376	0.571	0.102	-1.012	0.131	0.209	-12.39***
	(5.697)	(2.436)	(1.725)	(3.193)	(4.819)	(0.686)	(0.532)	(0.210)	(1.972)	(0.579)	(2.167)	(1.273)	(0.457)	(1.152)	(0.576)	(0.396)	(4.500)
6.stratum	-12.53	0.929	Omitted	Omitted	Omitted	-2.992***	1.499*	0.0562	-0.570	-0.118	Omitted	Omitted	Omitted	Omitted	-0.261	-0.180	Omitted
	(9.305)	(3.979)				(1.121)	(0.868)	(0.344)	(3.221)	(0.945)					(0.940)	(0.647)	
7.stratum	Omitted	Omitted	-2.522	-3.648	3.612	Omitted	Omitted	Omitted	Omitted	Omitted	-0.448	0.379	0.00357	-1.893	Omitted	Omitted	-8.530
			(2.104)	(3.894)	(5.876)						(2.642)	(1.553)	(0.558)	(1.405)			(5.488)
Cluster	0.00163	-0.000594	-0.000171	0.000616	0.00255***	-0.000334**	0.000404***	3.14e-06	-0.00171***	-0.000140	-0.00103***	-0.000422*	7.02e-05	5.21e-05	-0.000238*	-0.000126	-0.00183**
	(0.00130)	(0.000558)	(0.000295)	(0.000546)	(0.000824)	(0.000157)	(0.000122)	(4.82e-05)	(0.000451)	(0.000132)	(0.000370)	(0.000218)	(7.82e-05)	(0.000197)	(0.000132)	(9.07e-05)	(0.000769)
Tobacco	-1.553	0.687	0.368	-1.488	2.875	-0.106	0.108	-0.0180	-1.849	-0.230	-0.258	1.164**	0.678***	-0.221	0.141	-0.0640	0.0618
	(3.483)	(1.490)	(0.788)	(1.458)	(2.200)	(0.420)	(0.325)	(0.129)	(1.206)	(0.354)	(0.989)	(0.581)	(0.209)	(0.526)	(0.352)	(0.242)	(2.054)
Cons	-598.1	-53.45	134.7	182.1	-408.3	34.03	30.76	3.469	-66.59	56.23	-33.44	-55.80	-14.54	261.2***	27.06	41.34	623.7**
	(500.9)	(214.2)	(113.2)	(209.6)	(316.3)	(60.34)	(46.74)	(18.49)	(173.4)	(50.88)	(142.2)	(83.58)	(30.02)	(75.65)	(50.62)	(34.85)	(295.4)
Obs.	3,212	3,212	3,212	3,212	3,212	3,212	3,212	3,212	3,212	3,212	3,212	3,212	3,212	3,212	3,212	3,212	3,212
R-squared	-0.001	0.058	-0.017	0.154	-0.070	0.094	0.038	0.000	-0.003	-0.027	0.003	-0.209	-0.004	-0.581	-0.003	-0.078	-0.085

Table A 15: Chi-square statistics for Durbin-Wu-Hausman test for exogeneity

	Urban	**Rural**	**Top 50%**	**Bottom 50%**
Food	**47.64**	**47.87**	**52.98**	**37.99**
Alcohol	**55.00**	**63.38**	**74.87**	**44.32**
Health	**24.90**	8.83	**26.72**	9.89
School	**29.79**	14.38	17.97	**34.07**
Clothing	**37.10**	**52.54**	**49.64**	**44.91**
Housing	**83.82**	**41.00**	**80.08**	**32.21**
Water	**45.64**	9.03	**58**	19.68
Electricity	**206.31**	**63.29**	**182.96**	**65.24**
Alt. Energy	**83.24**	**37.64**	**95.74**	**45.22**
Daily Transport	**30.31**	**41.48**	**42.53**	17.20
Other Transport	**32.27**	11.77	**30.88**	21.58
Equipment	**130.53**	**68.71**	**106.23**	**100.45**
Entertainment	14.12	16.44	13.68	15.52
Telephone	**70.30**	**27.34**	**65.64**	**54.53**
Remittance	**30.37**	**40.34**	**44.37**	17.38
House Care	**37.96**	**48.51**	**54.84**	**53.8**
Personal Care	**923.38**	**634.85**	**572.16**	**162.95**

Notes: The table shows chi-square statistics associated with a Durbin-Wu-Hausman test for exogeneity. The null hypothesis is that d, lnM and $(lnM)^2$ are exogenous. The hypothesis is rejected for most of the categories. Chi-square statistics whose p-values are small are in bold in the table.

Table A 16: Ordinary least squares (OLS) estimates for the coefficient on *d*, restricted sample

Coefficient on *d* in:	Full Sample	Urban	Rural	Top 50%	Bottom 50%
Food	-1.627***	-0.931	-2.085***	-0.163	-2.993***
	(0.480)	(0.636)	(0.716)	(0.623)	(0.771)
Alcohol	5.502***	4.667***	6.106***	4.780***	6.024***
	(0.194)	(0.246)	(0.298)	(0.257)	(0.334)
Health	-0.135	-0.073	-0.170	-0.204*	-0.109
	(0.101)	(0.109)	(0.167)	(0.112)	(0.178)
School	-1.594***	-1.504**	-1.683***	-1.526***	-1.493***
	(0.259)	(0.379)	(0.358)	(0.403)	(0.332)
Clothing	-0.394	-0.057	-0.700*	-0.213	-0.557
	(0.259)	(0.290)	(0.420)	(0.282)	(0.465)
Housing	-0.558***	-1.324***	-0.037	-1.428***	0.061
	(0.200)	(0.410)	(0.058)	(0.397)	(0.094)
Water	-0.140*	-0.232	-0.066*	-0.307**	0.002
	(0.081)	(0.167)	(0.037)	(0.148)	(0.074)
Electricity	-0.205**	-0.465**	-0.006	-0.407**	0.021
	(0.099)	(0.192)	(0.072)	(0.202)	(0.029)
Alt. Energy	0.124	0.235	0.055	-0.014	0.096
	(0.165)	(0.227)	(0.236)	(0.188)	(0.269)
Daily Tport	-0.057	-0.111	-0.053	-0.062	-0.035
	(0.174)	(0.307)	(0.182)	(0.334)	(0.080)
Other Tport	0.079	0.468*	-0.161	0.275	0.081
	(0.195)	(0.257)	(0.290)	(0.317)	(0.225)
Equipment	0.062	0.136	0.002	0.218*	-0.084
	(0.086)	(0.106)	(0.133)	(0.122)	(0.119)
Entertain.	0.127	0.150	0.086*	0.142	0.060
	(0.078)	(0.154)	(0.052)	(0.160)	(0.048)
Telephone	-0.181	-0.094	-0.217	-0.211	-0.166*
	(0.143)	(0.263)	(0.134)	(0.271)	(0.095)
Remittance	-0.133*	-0.035	-0.194*	-0.098	-0.158**
	(0.078)	(0.122)	(0.100)	(0.139)	(0.080)
House Care	0.033	-0.116	0.132**	-0.096	0.153***
	(0.066)	(0.122)	(0.062)	(0.129)	(0.053)
Personal	-1.060***	-1.105***	-0.974**	-0.994***	-1.043**
	(0.261)	(0.328)	(0.403)	(0.313)	(0.444)
Sample size	**8,555**	**4,545**	**4,010**	**4,092**	**3,212**

Notes: The results shown above are only for the coefficient on d in equation (2.1) for the exact same sample as that used for the Three stage least squares (3SLS). The OLS results presented in the table are qualitatively similar to those in Table 2.3. Standard errors are reported in parentheses. *, **, *** implies that the coefficient on d is statistically significant at the 10%, 5% and 1% significance respectively. Tport is an abbreviation for transport. Alt is an abbreviation for alternative, entertain is short for entertainment and personal is short for personal care.

CHAPTER 3: PRICE AND EXPENDITURE ELASTICITIES OF DEMAND FOR CIGARETTES IN UGANDA

3.1. Introduction

Uganda's smoking prevalence estimates are relatively high by African standards. According to the Tobacco Atlas, estimates for adult smoking prevalence for males and females were 17% and 2% respectively in 2013 (Eriksen, *et al.*, 2015). Comparable estimates for Africa place adult male and female smoking prevalence at 14% and 3% respectively (Blecher and Ross, 2013). Among Ugandan youths, smoking prevalence is even higher. The most recent Global Youth Tobacco Survey (GYTS) estimated youth prevalence among Ugandan males at 19% and that of females at 11% (Musoke, 2008). For Africa, comparable estimates of youth smoking prevalence were 9% for young males and 3% for young females (Blecher and Ross, 2013). There is some expectation that smoking prevalence and tobacco consumption in Uganda will increase in the coming years driven by strong economic growth and declining real prices of tobacco products. According to the World Bank, real GDP growth has averaged 7% per year over the period 2005 to 2014 (World Bank, 2015). The rate of inflation averaged 10% per year over the same period (ibid.). As a consequence, the real price of cigarettes declined by 3% between 2008 and 2012 (Eriksen, *et al.*, 2015). The combination of strong economic growth and relatively high inflation has likely made tobacco products in Uganda more affordable. Meanwhile, a recent joint assessment by the World Health Organization (WHO) and the Government of Uganda noted an increase in the incidences of heart disease, chronic obstructive pulmonary disease, strokes and cancer, diseases that are linked to the usage of tobacco products (WHO, 2012b).

Increasing excise taxes on tobacco products is the single-most effective policy tool for reducing the demand for tobacco products (Chaloupka and Warner, 2000; IARC, 2011). Unfortunately, excise taxes in Uganda are low and well below the WHO's recommended target. The percentage of the cigarette retail price that is due to the excise tax (known as the excise tax burden) was estimated at 25% in 2013 (WHO, 2013b). WHO recommends an excise tax burden of at least 70% of the retail price (WHO, 2010). We know from several studies that cigarettes (and all other tobacco products) are generally price inelastic. Available estimates put the price elasticity of demand for cigarettes in the range of -0.20 to -0.60 (Chaloupka and Warner, 2000; IARC, 2011; Guindon *et al.*, 2015). The implication of these estimates is that by increasing taxes on cigarettes, countries can decrease consumption and at the same time increase tax revenues.

To the best of my knowledge, there are no published estimates of the price elasticity of demand for cigarettes in Uganda. As a matter of fact, there are very few published estimates of the price elasticity of demand for cigarettes in Africa. The comprehensive survey of the literature on the price elasticity of demand for cigarettes by the International Agency for Research on Cancer (IARC, 2011) was only able to list two sub-Sahara African countries (South Africa and Zimbabwe) as having published estimates of the price elasticity of demand. Even though we expect cigarettes to be price inelastic in most countries, policymakers still demand local evidence before initiating policy changes. There is, therefore, a need to generate local evidence in Africa to support civil society's efforts in encouraging governments to increase the excise tax as part of a comprehensive strategy.

The literature has traditionally estimated price elasticities of demand using aggregate time series data on cigarette consumption and prices. This partly explains why there are only a handful of countries in Africa with estimates of the price elasticity of demand: many African countries do not maintain time series data of a sufficient time length to allow for consistent estimates. On the other hand, many countries often conduct household expenditure surveys or household budget surveys. Deaton (1988) proposed a method that uses data from household expenditure surveys to estimate price elasticities of demand of goods and services covered by the surveys. A small literature has recently emerged that uses Deaton's method in estimating price elasticities of demand for cigarettes in Low- and Middle-Income Countries (John, 2005; John, 2008a; Eozenou and Fishburn, 2009; Guindon *et al.*, 2011; Chen and Xing, 2011). The advantage of Deaton's method is that it adequately deals with concerns about simultaneity bias that often plague time series studies (Chaloupka and Warner, 2000). And the fact that sample sizes for household surveys often number in the thousands allows for the estimation of a richer set of elasticity estimates than would be the case with time series data.

This chapter uses Deaton's method along with the 2005 and 2009 rounds of the Uganda National Panel Survey (UNPS), a household expenditure survey, to estimate price and expenditure elasticities of demand for cigarettes in Uganda. I focus on cigarettes because the UNPS does not contain detailed enough information on other tobacco products. My estimates of the price elasticity and expenditure elasticity of demand for cigarettes in Uganda fall within the range of estimates in the literature. I find that cigarettes are price inelastic with demand expected to decline by 3% to 4%, at the very least, for every 10% increase in prices (that is, price elasticity of -0.30 to -0.40). I also find that cigarettes in Uganda are normal goods with demand expected to increase by about 1% for every 10% increase in household expenditure. These results are robust to the exclusion of outliers in the data. Unfortunately, I am unable to estimate elasticities for different sub-samples (the rich, poor, etc…) owing to the small number of households that report positive expenditures on cigarettes in the UNPS. This is a general shortcoming of Deaton's method. Guindon *et al.* (2011) go around this problem by estimating elasticities for sub-samples using a simple share-log functional form (Gibson and Rozelle, 2005). As I discuss later, the share-log functional form requires a large enough sample which Guindon *et al.* satisfy by pooling 9 rounds of India's massive National Sample Survey. Further, the share-log functional form does not address the issue of quality shading which is a real problem when using self-reported prices by households. That is, households are likely to trade down to cheaper brands in response to a price rise. Failure to account for quality shading can bias elasticity estimates. All in all, the results in this chapter suggest that Uganda, by increasing taxes on cigarettes, can reduce cigarette consumption and increase excise tax revenues at the same time.

The rest of this chapter is structured as follows. Section 3.2 reviews the literature. Section 3.3 discusses Deaton's method in some detail and Section 3.4 discusses the data, the Uganda National Panel Survey. Section 3.5 presents the main econometric results and presents results for the robustness tests. Section 3.6 discusses some limitations and 3.7 concludes.

3.2. Literature review

There are generally two approaches to estimating the price and income (or expenditure) elasticity of demand for cigarettes. One approach uses aggregate time series data while a second approach uses household-level data. Until about 25 years ago, estimates of elasticities of demand were based nearly exclusively on time series studies. The paucity of time series data in Low- and Middle-Income Countries (LMICs) has meant that most of the estimates of the price elasticity of demand have come from the developed world. The time series literature on cigarette demand is vast and has been ably surveyed by Chaloupka and Warner (2000) and by the International Agency for Research on Cancer's (IARC) 2011 *Handbook on the Effectiveness of Tax and Price Policies for Tobacco Control*. The studies reviewed by Chaloupka and Warner (2000) had price elasticities of cigarettes ranging from -0.30 to -0.50, centering around -0.40. The studies reviewed by IARC (2011) had price elasticity estimates in the -0.20 to -0.60 range. The IARC Handbook also reviewed studies estimating the income elasticity of demand and concluded that cigarettes are normal goods with income elasticities in the 0 to 0.60 range.

The majority of time series studies on the demand for cigarettes in Africa have been done on South Africa. These studies (Reekie, 1994; Van Walbeek, 1996; Economics of Tobacco Control Project, 1998; Van Walbeek, 2005; Boshoff, 2008) find long-run price elasticities in the range -0.16 to -1.52 and income elasticities ranging from 0.37 to 1.70. Only three studies, all time series, were identified by IARC (2011) has having been done on African countries other than South Africa. The Economics of Tobacco Control Project (2008) estimated a long-run price elasticity of demand for Zimbabwe of -0.85 and an income elasticity of 1.67. Aloui (2003) estimated a price elasticity of between -1.36 and -1.54 and income elasticity between 0.87 and 1.04 for Morocco. Hanafy *et al.* (2011) working with Egyptian data estimated a price elasticity equal to -0.47 and an income elasticity equal to 1.60. In a recent systematic review mainly focussing on time series studies conducted in Latin America and the Caribbean, Guindon *et al.* (2015) find that the price elasticity of demand for cigarettes is most likely below -0.50 (that is, below 0.50 in absolute terms) with a 95% confidence interval of -0.35 to -0.51.

A criticism of studies using aggregate time series data is that they do not adequately deal with simultaneity bias since aggregate consumption and aggregate price are determined at the same time. Some studies, usually from the United States, go around this by exploiting exogenous changes in state-level cigarette taxes (Keeler, *et al.*, 1993; Hu *et al.*, 1994, 1995; Sung *et al.*, 1994). Exogenous changes in tax rates are uncommon in most parts of the world and especially so in LMICs. Secondly, time series econometric techniques require sample sizes of appropriate length in order to deliver consistent estimates. However, most LMICs do not have time series data of a sufficient time span. These reasons are largely behind the growing interest in estimating price and income (or expenditure) elasticities of demand for LMICs using household level data. Concerns about simultaneity bias are less pronounced with household level data since each household is too small to influence the price on its own. Further, sample sizes for most of the surveys tend to number in the thousands allowing for the estimation of a richer set of elasticities. A number of studies over the last decade have made use of a technique developed by Deaton (1988) to estimate price and expenditure elasticities[27] of demand from household survey data for LMICs. Deaton's method relies on the empirical fact that prices of most goods in LMICs often vary across geographical space (across clusters) owing to the presence of significant transportation costs. Prices also vary across geographical space because markets are often localised in nature such that it is difficult to take advantage of arbitrage opportunities. In this way, price in Deaton's method largely varies exogenously in contrast to time series studies where price is

[27] Since Deaton's method uses data from expenditure surveys, the income elasticity is often referred to as an expenditure elasticity under this literature.

endogenously determined. Deaton's method has two additional attractive properties: (1) it adjusts price elasticity estimates for quality heterogeneity and (2) it accounts for measurement error which is a problem with household surveys. Below I review studies that have used Deaton's method to estimate price and expenditure elasticities of demand for cigarettes in some detail since this chapter contributes to this literature.

John (2005) was the first to use Deaton's method to estimate price and expenditure elasticities of demand for cigarettes. He used the 1999 to 2000 round of India's National Sample Survey and estimated price elasticities of -0.22 for urban India and -0.56 for rural India. His estimates for the expenditure elasticity of demand were 1.72 for urban India and 2.37 for rural India suggesting that cigarettes were luxury goods. John (2008a) used the same dataset and the same method as his 2005 study with the exception that he only considered households that reported positive cigarette consumption. His argument for excluding households with zero expenditure on cigarettes was based on his finding (John, 2008b) that smoking and non-smoking households were fundamentally different in that they had different preferences. That is, cigarettes were not an argument in the utility function of non-smoking households. By excluding households with zero expenditure, John (2008a) was therefore estimating conditional elasticities. His estimates of the price elasticity of demand for urban and rural India were respectively -0.19 and -0.34, slightly smaller (in absolute terms) than his 2005 estimates. His estimates for the expenditure elasticity were 1.59 and 2.37, almost identical to the 2005 findings. Eozenou and Fishburn (2009) used the 1998 Vietnam Living Standards Survey and Deaton's method and estimated price and expenditure elasticities of demand equal to -0.53 and 0.30 respectively. The expenditure elasticity estimate suggested that cigarettes were a normal good in Vietnam.

Guindon *et al.* (2011) used data from three rounds of India's National Sample Survey (1993/94, 1999/2000 and 2004/05) and Deaton's method to estimate price and expenditure elasticities of demand for cigarettes in India. In some specifications, particularly when treating districts as clusters, they found price elasticity estimates of cigarettes that were much higher (in absolute terms) than those estimated by John (2005, 2008a). On the other hand, when treating villages/urban blocks as clusters, they found price elasticity estimates that were not statistically different from zero. Their expenditure elasticity estimates were, however, in line with those in John (2005, 2008a). Estimating elasticities for subsamples (poor, rich, more educated, less educated, etc…) is problematic for subsamples with Deaton's method because the number of cigarette consuming households are often too few to start with. In order to estimate elasticities for subsamples in India, Guindon *et al.* pooled 9 rounds of India's National Sample Survey and used a simple share-log functional form (Gibson and Rozelle, 2005).[28] They found that households from lower social castes tended to respond more to changes in cigarette prices. They, however, did not find consistent patterns in differences in price elasticities across household income, educational levels and urban/rural strata.

Chen and Xing (2011) use Deaton's method and data from China's Urban Household Income and Expenditure Survey collected between 1999 and 2001. They find price elasticity estimates for cigarettes in the range -0.35 to -0.56 and expenditure elasticity estimates in the range 0.27 to 0.45. Chen and Xing check how their estimates compare with a two-part model that separates cigarette demand into smoking participation and smoking intensity. The two-part model delivers a higher price elasticity (-0.82) which Chen and Xing conclude is likely due to the failure of the two-part model to account for quality heterogeneity.

With the exception of some of the specifications in Guindon *et al.* (2011), the handful of studies using household-level data along with Deaton's method tend to arrive at price elasticity estimates for cigarettes in the range of -0.10 to -0.60. The studies all conclude that the consumption of cigarettes does increase (or does not decline) with a rise in household income (expenditure), implying that income elasticities are non-negative.

The International Tobacco Control (ITC) Policy Evaluation Project based at the University of Waterloo is currently collecting tobacco-dedicated household and individual-level data from Low-and Middle-Income Countries (LMICs). The advantage with the ITC surveys, unlike traditional income and expenditure surveys, is that they assure large sample sizes of smoking households and smoking individuals and they explicitly collect information, including price, on different tobacco brands. The small number studies that has so far been conducted on ITC surveys in LMICs (Nargis *et al.*, 2010;

[28] With the share-log functional form, the demand for cigarettes is a function of household expenditure, household characteristics and average cluster prices. The last point follows from the assumption that prices vary across geographical space. Unlike Deaton's method, the share-log functional form does not adjust the final price elasticity estimates for quality heterogeneity. As discussed below, failure to adjust for quality heterogeneity is likely to bias the elasticity estimates away from zero.

Nargis *et al.*, 2011; Huang *et al.*, 2015) also tend to find price elasticity estimates for cigarettes in the -0.10 to -0.60 range.

3.3. Method

This chapter uses the method proposed by Deaton (1988) and extended in Deaton (1989, 1990 and 1997) to estimate price and expenditure elasticities for cigarettes in Uganda. The method exploits the fact that prices of most goods in Low- and Middle-Income Countries (LMICs) vary across geographical space (and are fixed within clusters) owing to the presence of significant transportation costs and the fact that markets are usually small and isolated. This last point makes it difficult to take advantage of price disparities between markets. Since most household expenditure surveys in LMICs do not collect data on prices, Deaton proposes the use of what he calls "unit values" as a proxy. In the case of cigarettes, unit values, υ, are defined as the ratio of total household expenditure on cigarettes, to quantity of cigarettes purchased or

$$\upsilon_{ic} = \frac{x_{ic}}{q_{ic}} \qquad (3.1)$$

where υ_{ic}, x_{ic} and q_{ic} are respectively the unit value, expenditure and quantity of cigarettes in household i located in cluster c. Unit values, however, are not the same thing as prices. For one thing, unit values hide a great degree of quality heterogeneity whereas the classic treatment of demand concerns itself with homogeneous goods. With quality heterogeneity, households may respond to a price increase by shifting to a lower "quality" brand of cigarettes with a small decline in quantity. Deaton refers to this as "quality shading". Under classic demand theory, the only outcome of a rise in prices is a reduction in demand. With quality shading, the price elasticity of demand will be overestimated.[29] Secondly, unit values are not the same as prices because of measurement error. Households are unlikely to correctly recall the amount of money spent on cigarettes and/or the

[29] Recall that the formula for the price elasticity of demand, ε, is given as:

$$\varepsilon = \frac{\Delta q/q}{\Delta p/p}$$

where Δq and Δp are the changes in price and quantity respectively. With quality shading, the price does not change in full and so the reported change in price (which is really the reported change in unit value) is smaller (in absolute size) than it would be if we had actual price data. Since the denominator in the price elasticity formula is smaller than it would otherwise be, the estimated price elasticity is bigger in absolute size.

quantity consumed. In some cases, the survey enumerator might incorrectly capture this information. In such a case, the ratio of expenditure to quantity would result in a wrong price even if cigarettes were a homogeneous product (i.e. cigarettes were not susceptible to quality shading). The presence of measurement error will bias the estimate of the price elasticity of demand.[30]

By defining quality as “the value of a bundle of goods at fixed reference prices” (Deaton, 1997, p. 297) and assuming that preferences are separable over bundles of goods (Deaton and Muellbauer, 1980b), Deaton derives formulas that can be used to correct the final price elasticity estimates for quality shading.[31] In as far as measurement error is concerned, recall that prices are assumed fixed within clusters. In other words, all households within a cluster should report prices that converge on the “cluster price” (which could either be the average cluster price or the median cluster price). Any household reporting a price that is different from the cluster price should be as a result of measurement error. Quantifying these deviations from the cluster price allows us to correct our price elasticity estimates for measurement error (see below).

Deaton’s method proceeds in a series of steps. The first step involves checking whether prices (unit values in this case) vary across geographical space. This can be done in one of two ways: (1) using Analysis of Variance (ANOVA) to divide the total variation in unit values into “within cluster variation” and “between cluster variation” or (2) running a regression of unit values on cluster dummies. A large F statistic in either case leads to the conclusion that unit values vary across space. In the second step, one estimates “within cluster” regressions of the form:

$$lnv_{ic} = \lambda + \beta lnx_{ic} + \boldsymbol{\gamma Z_{ic}} + \psi ln\pi_c + e_{ic} \qquad (3.2) \text{ and}$$

$$w_{ic} = \alpha + \varepsilon lnx_{ic} + \boldsymbol{\delta Z_{ic}} + \theta ln\pi_c + (FE_c + u_{ic}) \qquad (3.3)$$

w_{ic} represents the share of cigarette expenditure in total household expenditure for household i in cluster c and lnv_{ic} is the log of the unit value, derived according to equation (3.1) for household i in

[30] The direction of the bias will depend on the nature of the correlation, if at all any, between measurement error in expenditure and measurement error in quantity (see Deaton, 1997, p. 292 – 293).

[31] A full discussion of this is beyond the scope of this chapter but is contained in Deaton (1997, p.296 – 299).

cluster c. lnx_{ic} is the log of total household expenditure over the relevant reference period. $\boldsymbol{Z_{ic}}$ is a vector of household specific characteristics including household size, household gender composition, gender of the household head, proportion of adults in the household and years of schooling of the household head. Other variables in $\boldsymbol{Z_{ic}}$ are the age of the household head and a dummy variable for whether the household head is formally employed or not. FE_c is a cluster fixed effect. I also include a year dummy in $\boldsymbol{Z}$ whenever I run equations (3.2) and (3.3) on the pooled sample. This captures any unobserved factors that are specific to different survey years.[32] u_{ic} and e_{ic} are the standard regression error terms. $ln\pi_c$ are the unobserved prices and consequently, equations (3.2) and (3.3) are estimated without them. The coefficients on the price terms can, however, be recovered (see below).

Equation (3.2), referred to as the "unit value" equation, allows us to check for the presence of quality effects in the unit value data. A positive and statistically significant relationship between household expenditure and unit values, after accounting for household characteristics, would suggest the presence of quality effects. That is, richer households report higher unit values primarily because they are buying cigarettes of a higher quality. Knowing the pattern of the quality effects (i.e. the magnitude of β), allows us to correct our final price elasticity estimates for quality shading. Equation (3.3), on the other hand, is a standard demand equation where the cigarette share is expressed as a function of household income (proxied by household expenditure), household characteristics and prices. Because of the assumption that prices are fixed within clusters and the fact that we do not have price data, prices are proxied by cluster fixed effects (FE_c). Further, the cluster fixed effects also allow us to hold constant cluster-level tastes and preferences. Similar tastes and preferences are to be expected for narrowly constructed clusters such as villages. Unlike the demand equation, the unit value equation (3.2) does not contain cluster-level fixed effects even though unit values are related to price. Adding a cluster-level fixed effect to equation (3.2) "would break the link between prices and unit values, would prevent [unit values] giving any useful information about [prices], and would thus remove any possibility of identification" (Deaton, 1997, p.295). Equations (3.2) and (3.3) also contain useful information about measurement error at the household level. The magnitude of the errors are captured by e_{ic} and u_{ic}, the regression error terms. The relationship between the two errors (as captured by, say, the covariance) is useful in correcting the final price elasticity estimates for measurement error (see more below).

[32] I run Deaton's method on three datasets: the 2005 edition of the Ugandan National Panel Survey (UNPS), the 2009 edition of the UNPS and on a pooled sample of the two surveys. See section 3.4 for more details on the data.

The third step involves stripping the household-level demand and unit values of the effects of household expenditure and household characteristics and then averaging across clusters. This step requires the following equations:

$$\widehat{y_c^1} = \frac{1}{n_c}\sum_{i=1}^{n_c}(lnv_{ic} - \hat{\beta} lnx_{ic} - \hat{\boldsymbol{\gamma}}\boldsymbol{Z_{ic}}) \qquad (3.4) \text{ and}$$

$$\widehat{y_c^2} = \frac{1}{n_c}\sum_{i=1}^{n_c}(w_{ic} - \hat{\varepsilon} lnx_{ic} - \hat{\boldsymbol{\delta}}\boldsymbol{Z_{ic}}) \qquad (3.5),$$

where n_c is the number of households in cluster c. $\widehat{y_c^1}$ and $\widehat{y_c^2}$ are the estimates of, respectively, cluster average unit value and cluster average demand after removing the effects of household expenditure and household characteristics (notice that $\widehat{y_c^1}$ and $\widehat{y_c^2}$ do not have the i subscript because they represent cluster averages). Recall that our identifying assumption is that prices vary at the cluster level (i.e. between clusters and not within clusters). Given this identifying assumption, price elasticities of demand can only be obtained by seeing how cluster-level demand responds to changes in cluster-level prices. This leads to the fourth step which involves regressing cluster-level demand, $\widehat{y_c^2}$, on cluster-level unit values, $\widehat{y_c^1}$. The coefficient on $\widehat{y_c^1}$ in such a regression can alternatively be obtained by dividing the covariance between $\widehat{y_c^2}$ and $\widehat{y_c^1}$ by the variance of $\widehat{y_c^1}$. That is $\hat{\phi}$, the estimate of the coefficient on y_c^1, is obtained by

$$\hat{\phi} = \frac{Cov\left(\widehat{y_c^2}, \widehat{y_c^1}\right) - \frac{\widehat{\sigma^{12}}}{n_{size}}}{Var\left(\widehat{y_c^1}\right) - \frac{\widehat{\sigma^{11}}}{n_{size}}} \qquad (3.6),$$

where n_{size} is the average cluster size (households) in the sample; $\widehat{\sigma^{12}}$ is the estimate of the covariance of the errors in equations (3.2) and (3.3); $\widehat{\sigma^{11}}$ is the variance of the errors in equation (3.2). Equation (3.6) is a standard errors-in-variables regression where the covariance and variance of errors is used to correct for measurement error. Notice that the correction factors for measurement error become small as the average cluster size becomes large (i.e. as $n_{size} \rightarrow \infty$).

The fifth and final step in Deaton's method applies quality correction formulas in obtaining the estimate of the price elasticity of demand, $\widehat{\varepsilon_P}$, as follows:

$$\widehat{\varepsilon_P} = \left(\frac{\hat{\theta}}{\bar{w}}\right) - \hat{\psi} \qquad (3.7),$$

where $\bar{w}$ is the average share of total household expenditure dedicated to cigarettes in the sample. $\hat{\psi}$ and $\hat{\theta}$, the estimates of the coefficients on the unobserved price terms in equations (3.2) and (3.3) respectively, are recovered as follows:

$$\hat{\psi} = 1 - \frac{\hat{\beta}(\bar{w} - \hat{\theta})}{\hat{\varepsilon} + \bar{w}} \qquad (3.8) \text{ and}$$

$$\hat{\theta} = \frac{\hat{\phi}}{1 + (\bar{w} - \hat{\phi})\hat{\zeta}} \qquad (3.9) \text{ with}$$

$$\hat{\zeta} = \frac{\hat{\beta}}{\hat{\varepsilon} + \bar{w}(1 - \hat{\beta})} \qquad (3.10).$$

$\hat{\beta}$ is the estimate of the coefficient on total household expenditure in equation (3.2), the within-cluster unit value equation, and $\hat{\varepsilon}$ is the estimate of the coefficient on total household expenditure in equation (3.3), the within-cluster demand equation. $\hat{\phi}$ is the errors-in-variables estimate of the coefficient of a regression of cluster-level demand on cluster-level unit value.

Deaton also proposes the following formula for obtaining the estimate of the expenditure elasticity of demand, $\widehat{\varepsilon_I}$:

$$\widehat{\varepsilon_I} = 1 + \left(\frac{\hat{\varepsilon}}{\bar{w}}\right) - \hat{\beta} \qquad (3.11).$$

Deaton's method is not without its critics. Gibson and Rozelle (2004 and 2005) show that using unit values as a substitute for actual price yields biased estimates for the price elasticity of demand even after correcting for quality effects and measurement error. In one study, Gibson and Rozelle (2005) compare price elasticity estimates from unit values with estimates from actual prices obtained from a market survey and find that unit values give biased results. The criticisms notwithstanding, unit values present the most cost-effective way of saying something about the effect of price on demand. An alternative would be to conduct a detailed price survey (such as Gibson and Rozelle describe in their 2005 paper) for each commodity and map these prices to each household. If done properly, such an exercise would go a long way in addressing some of the biases in Deaton's method. But this would be many times more costly than using Deaton's method. Alternatively, one can use time series data for a country's aggregate consumption of cigarettes and some measure of aggregate prices to estimate price elasticities such as in Van Walbeek (2005). Unfortunately most sub-Saharan African countries do not have time series data of a sufficient time span to allow for consistent estimates of the price elasticity of demand. Further, time series estimates of the demand for cigarettes are likely to suffer from reverse causality and simultaneity bias (Chaloupka and Warner, 2000). Deaton's method overcomes these problems in large part by relying on an exogenous driver of prices (transportation costs) and the fact that each household's demand is too small on its own to affect overall demand.

3.4. Data

The data for this chapter come from the 2005 and 2009 editions of Uganda's National Panel Survey (UNPS), a nationally representative survey. The UNPS is conducted by the Uganda Bureau of Statistics with technical assistance from the World Bank. Households that were surveyed in 2005 were revisited in 2009. Additional panel studies have since been conducted in 2010 and 2011. This chapter does not use data from the 2010 and 2011 rounds because the latter two panel surveys did not collect sufficiently detailed information on the quantities of cigarettes purchased by households. Specifically, information on whether the unit of purchase was a stick, packet or bundle was missing from the 2010 and 2011 rounds of the survey. Deaton's method requires that the units for the quantity variable be the same.[33]

[33] One can convert non-standard units (say, sticks) into standard units (say, packs) if the survey is explicit in distinguishing reported units.

A two stage sampling design was used to select the 2005 sample. In the first stage, 783 Enumeration Areas (EAs) were drawn with probability proportional to size. In the second stage, simple random sampling (SRS) was used to select 3,123 households. The 2009 panel survey aimed at revisiting all the households that were interviewed in 2005 and achieved a sample size of 2,975 households.[34] Since EAs are the smallest geographical unit in the survey and the use of SRS presumes that households within an EA are similar, I take EAs to represent clusters.

The Uganda National Panel Survey collects an exhaustive list of information on the socio-economic characteristics of households. This includes information on household expenditure patterns, household composition patterns, and the occupational and educational status of household members. Information on the agricultural activities engaged in by households is also collected given that Uganda, like most African countries, has a prominent agricultural sector. In the household expenditure module of the survey, which is this chapter's primary interest, households are required to recall expenditure on various goods and services. The recall period ranges from 7 days for cigarettes to one year for household expenditures on education, health and clothing to name a few. In this chapter, I convert all expenditures into weekly expenditures.[35]

The expenditure module collects information on the quantity and total amount paid for cigarettes over the last 7 days. The quantities are recorded either as sticks or packets. I convert sticks into packets by dividing them by 20, which is the number of cigarettes in a pack in Uganda. The module also collects information on "other tobacco". Unfortunately, there is not much additional information on what constitutes other tobacco let alone a careful distinction of the units in which quantities are collected. Because of this, I only focus on cigarettes in estimating the price elasticity of demand. Further, I cannot estimate cross price elasticities between cigarettes and other tobacco because of a lack of standardization in the units of the latter. Another unfortunate aspect of the Uganda National Panel Survey is that it does not distinguish between zero expenditures and expenditure information that is

[34] Additional information on sample selection and other aspects of the survey can be accessed here: http://econ.worldbank.org/WBSITE/EXTERNAL/EXTDEC/EXTRESEARCH/EXTLSMS/0,,menuPK:3359053~pagePK:64168427~piPK:64168435~theSitePK:3358997,00.html

[35] Expenditures reported as recalled over the previous 30 days are converted to weekly expenditures by multiplying them by 7/30. Expenditures recalled over the previous year are converted to weekly expenditures by multiplying them by 7/365.

simply missing. In other words, a household's information on cigarette expenditure is left out of the survey if (1) the household reported zero expenditure on cigarettes in the previous week or (2) the household was never asked this question or (3) the respondent did not know the answer to the question. Notice that case (1) coincides with a scenario where the household consumes cigarettes but did not do so in the previous week because, say, the price was high (i.e. a corner solution). It also coincides with a scenario where the household does not consume any cigarettes at all. On the other hand, cases (2) and (3) are scenarios where information is simply missing. Because the survey is not explicit about distinguishing zero expenditures from missing expenditures, I choose to only use households that reported positive cigarette expenditures in estimating the price elasticity of demand. In other words, I estimate conditional price elasticities of demand.[36, 37, 38]

In estimating the price elasticity of demand, I run two specifications. The first specification estimates separate price elasticities of demand for the 2005 and 2009 rounds of the Ugandan National Panel Survey. In the second specification, I pool the 2005 and 2009 datasets and estimate a single price elasticity of demand. That is, I treat the 2005 and 2009 panels as if they were separate cross-sectional surveys. To allow for comparisons across years, I express all unit values and total expenditures in terms of 2010 prices using Uganda's Consumer Price Index (CPI).[39]

[36] An alternative approach would be to assume that all households without information on cigarette expenditure reported zero consumption in the previous period. In other words, to assume away missing information. This would not be advisable given that households with genuinely missing information are unlikely to be trivial for a good like cigarettes. Respondents to household surveys in Low- and Middle-Income Countries (LMICs) are usually female spouses who might not know if their husbands smoke (especially for "light" smokers) and if they do, may not feel comfortable in disclosing this information due to the socially undesirable nature of smoking in Africa.

[37] If the survey were explicit in distinguishing zeroes from missing information, one could use the method proposed by Vermeulen (2003) to test for whether reported zeroes are a result of corner solutions or a result of households not being consumers of cigarettes.

[38] John (2008a), using Deaton's method, also estimates elasticities for smoking households only.

[39] I obtain the CPI from the World Development Indicators provided by the World Bank. Available here: http://data.worldbank.org/data-catalog/world-development-indicators

Table 3.1: Summary statistics from the Uganda National Panel Survey (UNPS)

Variable	2005	2009	Pooled Sample
Percentage of households with positive cigarette expenditure	9%	7%	8%
Average cigarette share in total household expenditure	8.86%	7.67%	8.40%
Real average unit value per pack	1,952.11	1,056.26	1,605.40
Average unit value per pack (USD equivalent)	0.83	0.45	0.68
Average weekly quantity of cigarettes (in packs)	1.45	1.55	1.49
Ave. real h/hold expenditure (last 7 days in Ugandan Shillings)	34,598.08	32,784.07	33,896.02
Ave. household expenditure (last 7 days USD equivalent)	14.72	13.95	14.42
Average household size	5.58	6.33	5.87
Average age of household head in years	41.13	43.73	42.15
Average proportion of male heads	86.81%	90.17%	88.12%
Average proportion of adults	54.58%	46.88%	51.59%
Average proportion of males in household	53.27%	52.00%	52.63%
Average years of schooling of household head	6.55	5.81	6.26
Average proportion of heads with some employment	89.74%	84.43%	87.73%
Total number of clusters	178	121	299
Total number of households	274	173	447
Average number of households per cluster	1.54	1.43	1.49

Notes: Summary statistics for the relevant variables from the 2005 and 2009 Uganda National Panel Survey (UNHS) and the pooled sample. Adults are those household members who are 18 years or older.

Table 3.1 reports summary statistics for the relevant variables from the data. I report but do not discuss the summary statistics for the pooled sample because in most cases, the variables for the pooled sample are the averages of the 2005 and 2009 datasets.

8% of the sample of households from 2005 reported positive expenditures on cigarettes in the last 7 days. For 2009, this number was 7%. The share of the weekly budget allocated to cigarette expenditure was, on average, 8.86% in 2005 and 7.67% in 2009. Although Uganda's cigarette expenditure share is relatively high, it falls within the range identified by John (2008b) particularly for LMICs. The reported average unit value per pack in 2010 Ugandan Shillings was 1,952 (83 US cents) in 2005 and 1,056 (45 US cents) in 2009.[40] To the extent that unit values are a good proxy of actual prices, then the foregoing would suggest that cigarettes became cheaper, in real terms, by about 46% over the four year period in question. The average weekly household expenditure expressed in 2010 Uganda shillings was 34,598 (USD15) in 2005 and declined slightly by 5% to 32,784 (USD14) in 2009. The rest of the

[40] I convert 2010 Uganda shillings into US dollars by using the 2010 end of year exchange rate between the shilling and the dollar available from www.oanda.com.

summary statistics in Table 3.1 mainly pertain to the $\boldsymbol{Z}$ vector of control variables used in the regressions (see equations 3.2 and 3.3 above).

3.5. Econometric results

3.5.1. Main results

In this section of the chapter, I present and discuss the results of implementing Deaton's method. Recall from Section 3.3 that the main identifying assumption behind the method is that prices vary across geographical space. The validity of this assumption can be tested using Analysis of Variance (ANOVA) techniques. I report the results of the ANOVA exercise in Table 3.2.

Table 3.2: Testing the spatial variation hypothesis

2005 sample				2009 sample				Pooled sample			
F statistic	p value	R-squared	n	F statistic	p value	R-squared	n	F statistic	p value	R-squared	n
1.29	0.08	0.70	274	1.12	0.33	0.72	173	1.26	0.06	0.72	447

Notes: The F statistic and the p-value are associated with the null hypothesis of no spatial variation in unit values. The hypothesis is rejected in the 2005 and pooled samples but not in the 2009 sample. The R-squared measures the proportion of variation in prices taking place between clusters. n is the total number of households.

According to the results of the ANOVA procedure, most of the variation in unit values takes place between clusters. The R-squared, which measures the proportion of total variation in unit values between clusters, is at least 70% in all three data specifications. The F statistics associated with a null hypothesis of no spatial variation are all large with the exception of the 2009 sample. Even with the 2009 sample, the R-squared is 72%. In large part, the assumption of spatially varying prices is confirmed by the data. John (2005) and John (2008a) also find that cluster effects explain at least 70% of the variation in unit values in India.[41]

The next step in Deaton's method involves running the within-cluster unit value and budget share regressions (as per equations 3.2 and 3.3). The results for the unit value regression for all three

[41] I am unable to compare with more studies as the results of the spatial variation hypothesis are often not reported.

samples are reported in Table 3.3. The results in the table show that quality effects are present in the data: households with higher household expenditure tend to report higher unit values controlling for other household characteristics and controlling for year effects in the pooled sample. The expenditure elasticities of quality range between 0.11 and 0.23. That is, reported unit values rise by between 1% and 2% for every 10% increase in household expenditure. These expenditure elasticities of quality compare favourably with those estimated for cigarettes by John (2005), John (2008a), Eozenou and Fishburn (2009), Guindon *et al.* (2011) and Chen and Xing (2011). These studies find expenditure elasticities of quality in the range of 0.11 to 0.40.

Table 3.3: Results of the unit value regression (equation 3.2)

Variables	(2005) lnv	(2009) lnv	(Pooled) lnv
lnx	0.234***	0.115**	0.176***
	(0.051)	(0.048)	(0.036)
Size	-0.042	-0.010	-0.019
	(0.124)	(0.119)	(0.088)
Adults	-0.203	0.159	-0.071
	(0.295)	(0.300)	(0.213)
Males	0.261	0.131	0.239
	(0.216)	(0.223)	(0.159)
Education	-0.143*	0.108	-0.023
	(0.080)	(0.074)	(0.055)
Age	-0.015	-0.409**	-0.123
	(0.153)	(0.166)	(0.114)
Gender	0.217	0.218	0.225*
	(0.163)	(0.183)	(0.123)
Work	-0.144	0.101	-0.000
	(0.141)	(0.118)	(0.094)
Year			-0.191***
			(0.061)
Constant	4.957***	6.602***	5.490***
	(0.692)	(0.739)	(0.511)
No. of households	233	147	380
R-squared	0.115	0.126	0.110

Notes: Results of the regression of the log of unit value (lnv) on the log of household expenditure (lnx) and other household characteristics (see equation 3.2). Household size (Size), education of household head (Education) and age of household head (Age) are in natural logarithms. Adults refers to the proportion of adults in a household and adults are defined as aged 18 years or older. Males is the proportion of males in a household. Gender is a dummy variable which takes on the value of 1 if the household head is male and zero if they are female. Work is a dummy variable which takes on the value of 1 if the household head is employed and zero otherwise. Year takes on the value of 1 if the household was surveyed in 2005 and zero if surveyed in 2009. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

The other variables in Table 3.3 are not statistically significant with the exception of the education level of the household head in the 2005 sample and age for the household head in the 2009 sample. Other variables that are statistically significant are the gender dummy and the year dummy in the pooled sample. Households in the 2005 sample with heads who are more educated tend to report smaller unit values (although the effect is only significant at the 10% level). For 2009, households with heads who are older tend to report smaller unit values. And in the pooled sample, the gender of the head of household matters for the reported unit values. Households headed by males tend to report higher unit values controlling for other characteristics. The year dummy in the pooled sample is negative, agreeing with the observation from Table 3.1 that reported unit values declined in real terms between 2005 and 2009.

Table 3.4: Results of the budget share regression (equation 3.3)

Variables	(2005) w	(2009) w	(Pooled) w
lnx	-0.056***	-0.065***	-0.056***
	(0.017)	(0.023)	(0.013)
Size	0.002	0.039	0.002
	(0.031)	(0.043)	(0.023)
Adults	0.008	0.092	0.011
	(0.072)	(0.103)	(0.054)
Males	0.013	0.010	0.001
	(0.059)	(0.068)	(0.042)
Education	-0.001	-0.012	-0.007
	(0.020)	(0.025)	(0.015)
Age	0.028	-0.077	0.009
	(0.044)	(0.072)	(0.035)
Gender	-0.038	-0.108*	-0.050*
	(0.037)	(0.056)	(0.029)
Work	0.037	0.058	0.044*
	(0.037)	(0.039)	(0.026)
Year			-0.320***
			(0.107)
Constant	0.533***	0.963***	0.618***
	(0.193)	(0.292)	(0.148)
No. of households	233	147	380
R-squared	0.866	0.909	0.876

Notes: Results of the regression of the cigarette budget share (w) on the log of household expenditure (lnx) and other household characteristics (see equation 3). Household size (Size), education of household head (Education) and age of household head (Age) are in natural logarithms. Adults refers to the proportion of adults in a household and adults are defined as aged 18 years or older. Males is the proportion of males in a household. Gender is a dummy variable which takes on the value of 1 if the household head is male and zero if they are female. Work is a dummy variable which takes on the value of 1 if the household head is employed and zero otherwise. Year takes on the value of 1 if the household was surveyed in 2005 and zero if surveyed in 2009. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Cluster fixed effects are suppressed for space reasons but are jointly statistically significant at the 1% level for the 2005 and pooled samples and at 10% for the 2009 sample.

Table 3.4 reports the results of running the budget share regression (equation 3.3) on household expenditure alongside other household characteristics, cluster fixed-effects and year effects (for the pooled sample). In Table 3.4, there is a negative and statistically significant relationship between household expenditure and the share of the household budget allocated to cigarettes. The magnitudes of the coefficients on the log of household expenditure are similar in all three data specifications. The cigarette budget share tends to fall as household expenditure rises. The cluster fixed effects, whose results are not presented in Table 3.4 for purposes of space, are all jointly statistically significant at the 1% level for the 2005 and pooled samples and at the 10% level for the 2009 sample. This means that unobserved but fixed factors at the cluster level (prices, tastes, preferences, etc…) matter for cigarette demand. The only other variables that are significant in Table 3.4 are the gender variable (2009 and pooled sample) and the work dummy (pooled sample). For the 2009 and pooled samples, households headed by males tend to allocate smaller budget shares to cigarettes. On the other hand, in the pooled sample, households with an employed household head tend to allocate more of the household's budget to cigarettes.

The next step in Deaton's method involves stripping unit values and the budget shares of the effects of household expenditure and household characteristics and averaging by cluster, as per equations (3.4) and (3.5). Afterwards, I obtain an estimate of $\hat{\phi}$, the coefficient in an errors-in-variables regression of cluster-level demand on cluster-level unit value, as per equation (3.6). $\hat{\phi}$ is corrected for measurement error using the residuals from the within-cluster unit value and within-cluster budget share regressions as described in equation (3.6). The final step in Deaton's method involves applying equation (3.7) along with the quality-correction formulas in equations (3.8), (3.9) and (3.10) to obtain an estimate of the price elasticity of demand. The price elasticity estimates for Uganda for the 2005, 2009 and pooled sample along with bootstrapped standard errors[42] are presented in Table 3.5.

[42] Since the formulae in equations (3.7), (3.8), (3.9), (3.10) and (3.11) are not Stata commands, the standard errors have to be obtained via the method of the bootstrap. For the price elasticity of demand, this is done by running the cluster-level regressions 1000 times and in each instance using the formulae in equations (3.7), (3.8), (3.9) and (3.10) to compute the price elasticity. That is, I end up with 1000 estimates of the price elasticity. The standard error is then calculated as "half the length of the interval that is symmetric around the bootstrapped mean [of the 1000 estimates], and that contains 68.3% of the bootstrapped estimates" (Deaton, 1997, p317). Since the inputs into calculating the expenditure elasticity of demand are all within-cluster coefficients (see equation 3.11), within-cluster regressions (3.2) and (3.3) are run 1000 times in obtaining bootstrapped standard errors.

Table 3.5: Estimates of the price elasticity of demand for cigarettes in Uganda

	(2005)	(2009)	(Pooled sample)
$\widehat{\varepsilon_P}$	-0.326***	-0.258***	-0.405***
	[0.021]	[0.011]	[0.048]
	(-0.368 , -0.284)	(-0.280 , -0.235)	(-0.499 , -0.311)
No. of households	233	147	380
No. of clusters	184	130	314

Notes: Estimates of the price elasticity of demand for cigarettes in Uganda for the 2005, 2009 and pooled samples. Bootstrapped standard errors are in square brackets. 95% confidence intervals are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

The estimates in Table 3.5 show that cigarettes are price inelastic in Uganda. The estimates for the price elasticity of demand for the three data specifications are all statistically significant at the 1% level. For 2005, the price elasticity is estimated at -0.33 while for the 2009 sample the estimate is at -0.26. The pooled sample has a slightly more responsive elasticity at -0.41. Taken together, these results imply that the demand for cigarettes in Uganda is expected to decline, at the very least, by between 3% and 4% for every 10% rise in prices.[43] These results fall within the range of estimates for studies that use Deaton's method (-0.10 to -0.60) and studies using aggregate time series data (-0.20 to -0.60).

The results presented in Table 3.5 are conditional price elasticities and, therefore, only measure price responsiveness among households whose members continue to smoke. We, however, know that a price rise encourages some smokers to quit as well as deters young people from starting to smoke in the first place (Forster and Jones, 2001; Tauras *et al.*, 2001; Nicolas, 2002; Cawley *et al.*, 2004; Franz, 2008). Therefore, the total price elasticity of demand is made up of the conditional elasticity (as

[43] Ideally, the three price elasticity estimates in Table 3.5 should roughly be the same. For instance, the estimate of the price elasticity from the pooled sample should not be that different to the estimates of the two constituent samples. However, inputs into the formulae for calculating the price elasticity of demand (equations 3.7 to 3.10) are sample specific and might therefore influence the final estimate. For instance, the average cigarette budget share, $\bar{w}$, which features in all of the formulae in equations 3.7 to 3.10, is equal to 0.089, 0.077 and 0.084 for the 2005, 2009 and pooled samples respectively. Such small differences are likely to influence, in a small way, the final estimates.

estimated in Table 3.5) and the elasticity associated with cessation and initiation (sometimes referred to as the participation elasticity). The magnitude of the participation elasticity is often equal to that of the conditional elasticity so that the total elasticity is twice the conditional elasticity (see Table 5.1 in IARC, 2011). This implies that the total price elasticity of demand for cigarettes in Uganda is likely to be higher (in absolute size) than the results presented in Table 3.5.

Table 3.6 presents estimates of the expenditure elasticity of demand using equation (3.11) and coefficient estimates from Tables 3.3 and 3.4. The estimates in Table 3.6 show that cigarettes are normal goods in Uganda. That is, cigarette demand does not, at the very least, decline with an increase in household income (expenditure).

Table 3.6: Estimates of expenditure elasticities of demand for cigarettes in Uganda

	(2005)	**(2009)**	**(Pooled sample)**
$\hat{\varepsilon}_I$	0.132 [0.338] (-0.531 , 0.796)	0.043 [0.539] (-1.014 , 1.100)	0.157 [0.250] (-0.333 , 0.647)
No. of households	233	147	380

Notes: Estimates of the expenditure elasticity of demand for cigarettes in Uganda for the 2005, 2009 and pooled samples. Bootstrapped standard errors are in square brackets. 95% confidence intervals are in parentheses. Since the expenditure elasticity of demand is estimated at the household level (see equation 3.11), I only report the number of households.

3.5.2. Robustness

The estimates of the price elasticity of demand for cigarettes presented above are likely to be influenced by extreme unit values given that the sample size is relatively small. To check the robustness of my estimates, I follow Guindon *et al.* (2011) and exclude unit values that are greater than 2.5 standard deviations and 5 standard deviations from their respective means. (Given the similarity of the results in Tables 3.5 and 3.6, I only perform the robustness test on the pooled sample). The first-stage within-cluster regressions for the robustness tests are presented in Appendix B in Table B1 and Table B2. The expenditure elasticities of quality are fairly similar to those in Table 3.3: reported unit values rise by 0.15 for every 1% increase in household expenditure. Similarly, the coefficients on household expenditure in the budget share equation in Table B2 are similar to those in Table 3.4.

Table 3.7 contains the estimates of the price elasticity of demand for cigarettes for the robustness tests. I have also included in Table 3.7, estimates for the pooled sample taken from Table 3.5. The second column (column 2) excludes unit values that deviate by more than 5 standard deviations from the mean while the third column (column 3), being more stringent, excludes unit values that deviate by more than 2.5 standard deviations from the mean. The price elasticity estimates in columns 2 and 3 are statistically significant at the 1% level and are similar to the estimate from the pooled sample taken from Table 3.5. Table 3.8 presents the results of the robustness test for the expenditure elasticity (I have also included the estimates from the pooled sample from Table 3.6). The results in Table 3.8 are similar to those in Table 3.6.

Table 3.7: Estimates of price elasticities of demand for cigarettes in Uganda (Robustness check)

	(Pooled sample)	(2)	(3)
$\widehat{\varepsilon_P}$	-0.405***	-0.387***	-0.395***
	[0.048]	[0.068]	[0.082]
	(-0.499, -0.311)	(-0.521 , -0.253)	(-0.555 , -0.235)
No. of households	380	378	375
No. of clusters	314	312	312

Notes: Estimates of the price elasticity of demand for cigarettes in Uganda for the pooled sample and for the pooled sample with extreme unit values excluded. Column 1 (pooled sample) includes all unit values. In column 2, all unit values that are equal to or greater than 5 standard deviations from the mean are excluded. In column 3, all unit values that are equal to or greater than 2.5 standard deviations from the mean are excluded. Bootstrapped standard errors are in square brackets. 95% confidence intervals are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 3.8: Estimates of the expenditure elasticity of demand for cigarettes in Uganda (Robustness check)

	(Pooled sample)	(2)	(3)
$\widehat{\varepsilon_I}$	0.157	0.186	0.199
	[0.250]	[0.235]	[0.250]
	(-0.333, 0.647)	(-0.274 , 0.646)	(-0.292 , 0.690)
No. of households	380	378	375

Notes: Estimates of the expenditure elasticity of demand for cigarettes in Uganda for the pooled sample and for the pooled sample with extreme unit values excluded. Column 1 (pooled sample) includes all unit values. In column 2, all unit values that are equal to or greater than 5 standard deviations from the mean are excluded. In column 3, all unit values that are equal to or greater than 2.5 standard deviations from the mean are excluded. 95% confidence intervals are in parentheses. Since the expenditure elasticity of demand is estimated at the household level (see equation 3.11), I only report the number of households. None of the estimates are statistically significant at the conventional levels.

3.6. Limitations

This chapter does not present estimates of price and expenditure elasticities of demand for different subsamples of the Ugandan population. Policymakers are, for instance, interested in learning how responsive poorer households are to prices and expenditure vis-à-vis richer households, or in learning the degree of responsiveness to price and expenditure changes of urban households relative to rural households. Unfortunately, the small number of households reporting positive expenditure on cigarettes in the Uganda National Panel Survey (UNPS) makes analysis of subsamples problematic.[44] One way of going around this is to combine several rounds of the UNPS so as to boost the sample size, similarly to what Guindon *et al.* (2011) did with several rounds of India's National Sample Survey. However, there have only been 4 rounds of the UNPS to date and the cigarette expenditure data in the last two rounds was not collected at a sufficiently detailed level to allow for the calculation of unit values. Secondly, this chapter does not provide elasticity estimates for other tobacco products such as leaf or loose tobacco and it does not provide cross-price elasticities between cigarettes and other tobacco products. The consumption of other tobacco products is likely to be substantial, particularly among the poor, given that Uganda is a tobacco growing country. Unprocessed tobacco products are likely to find their way into the market. Knowing the price and income responsiveness of demand for products such as leaf tobacco is likely to be important to policymakers. Unfortunately, as stated in the data section, the UNPS does not collect detailed enough information on other tobacco products to allow for the estimation of these elasticities.

Secondly, this chapter has provided elasticity estimates only for households that reported positive expenditures on cigarettes. That is, the estimates in this chapter are conditional price and expenditure elasticities. Policymakers, in wanting to know the overall impact of tax measures, are often interested in the overall price and expenditure elasticities of demand. As detailed in the data section, I am unable to include households with zero expenditure in the analysis because the UNPS does not distinguish between households reporting zeroes and those whose expenditure information on cigarettes is missing.

[44] For instance I attempted to estimate price elasticities of demand for upper/lower income households and for urban/rural households using the pooled sample and found elasticities that go against theoretical expectations and elasticity estimates found in the literature. This is probably due to the small sample sizes that make the estimates easily susceptible to outliers in the data.

The hope is that future editions of the Uganda National Panel Survey will collect detailed enough information on household expenditures on not only cigarettes but on other tobacco products as well.

Finally, the small number of households per cluster means that the elasticity estimates presented in this chapter likely suffer from two sources of bias. Firstly, the reported unit values are likely endogenous because they are obtained from a few households in each cluster. A second source of bias is measurement error given that the measurement error correction formula in equation 3.6 requires that clusters contain many households.

3.7. Summary and conclusion

The aim of this chapter was to estimate price and expenditure elasticities of demand for cigarettes in Uganda. In doing so, I used the Uganda National Panel Survey (UNPS) and exploited the empirical fact that unit values, a measure of prices, varied exogenously across geographical space in Uganda. The last point allowed me to control for possible endogeneity in the cigarette demand equation.

I focussed on cigarettes because the first two rounds of the UNPS did not collect detailed enough information on other tobacco products to allow for the estimation of their elasticities using Deaton's method, which was the method I used in the chapter. For example, the survey was not explicit about the units in which other types of tobacco products were purchased. Using standardised units is critical with Deaton's method. Further, I was only able to estimate conditional elasticities because the UNPS does not distinguish between households reporting zero expenditure on cigarettes and those whose information is missing. Failure to separate zero expenditures from expenditure information that is missing is likely to result in biased estimates especially if the pattern of missing values is systematic. Lastly, sample size restrictions made it difficult to estimate separate elasticities for different sub-samples of Uganda's population.

I found that cigarette demand was price inelastic in Uganda with an elasticity estimate ranging between -0.3 to -0.4. This means that the quantity of cigarettes demanded is expected to decline by between 3% and 4%, at the very least, every time prices rise by 10%. This estimate of the price elasticity of demand falls within the range of estimates found in the literature.

The price elasticity estimate for Uganda is likely to be a lower bound estimate (in absolute terms) given we know that conditional elasticities of demand are smaller (in absolute terms) than total elasticities of demand. The total elasticity incorporates, in addition to reductions in demand, the fact that some smokers are likely to quit and some young people are unlikely to initiate when prices rise. The literature tends to find that the total price elasticity of demand is, on average, twice (in absolute terms) the conditional elasticity of demand (IARC, 2011).

My estimates of the price elasticity of demand suggest that Uganda can reduce cigarette consumption by increasing excise taxes on cigarettes. Further, and given that the reduction in demand is expected to be proportionately smaller than the increase in prices, the country can increase excise tax revenues at the same time. The important point to note is that the tax increases have to be real, inflation-adjusted, tax increases.

APPENDIX B

Table B 1: Unit value regression (equation 3.2) for robustness tests

Variables	(1) lnv	(2) lnv
lnx	0.152***	0.148***
	(0.031)	(0.027)
Size	0.011	-0.022
	(0.077)	(0.067)
Adults	0.116	0.054
	(0.186)	(0.162)
Males	0.152	0.146
	(0.138)	(0.121)
Education	0.001	0.024
	(0.048)	(0.042)
Age	-0.181*	-0.216**
	(0.099)	(0.086)
Gender	0.213**	0.173*
	(0.107)	(0.092)
Work	-0.022	0.028
	(0.081)	(0.071)
Year	-0.145***	-0.102**
	(0.054)	(0.047)
Constant	5.801***	5.964***
	(0.446)	(0.388)
No. of households	378	375
R-squared	0.118	0.142

Notes: Results of the regression of the log of unit value (lnv) on the log of household expenditure (lnx) and other household characteristics for the pooled sample only. In column 1, all unit values that are equal to or greater than 5 standard deviations from the mean are excluded. In column 2, all unit values that are equal to or greater than 2.5 standard deviations from the mean are excluded. Household size (Size), education of household head (Education) and age of household head (age) are in natural logarithms. Adults refers to the proportion of adults in a household and adults are defined as aged 18 years or older. Males is the proportion of males in a household. Gender is a dummy variable which takes on the value of 1 if the household head is male and zero if they are female. Work is a dummy variable which takes on the value of 1 if the household head is employed and zero otherwise. Year takes on the value of 1 if the household was surveyed in 2005 and zero if surveyed in 2009.Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table B 2: Budget share regressions (equation 3.3) for robustness tests

Variables	(1) w	(2) w
lnx	-0.056***	-0.055***
	(0.013)	(0.013)
Size	0.002	-0.000
	(0.023)	(0.023)
Adults	0.011	0.013
	(0.054)	(0.055)
Males	0.001	0.004
	(0.042)	(0.042)
Education	-0.007	-0.007
	(0.015)	(0.015)
Age	0.009	0.003
	(0.035)	(0.037)
Gender	-0.050*	-0.053*
	(0.029)	(0.029)
Work	0.044*	0.054*
	(0.026)	(0.028)
Year	-0.145***	-0.102**
	(0.054)	(0.047)
Constant	0.618***	0.627***
	(0.148)	(0.153)
No. of households	378	375
R-squared	0.876	0.879

Notes: Results of the regression of the cigarette budget share (w) on the log of household expenditure (lnx) and other household characteristics for the pooled sample only. In column 1, all unit values that are equal to or greater than 5 standard deviations from the mean are excluded. In column 2, all unit values that are equal to or greater than 2.5 standard deviations from the mean are excluded. Household size (Size), education of household head (Education) and age of household head (age) are in natural logarithms. Adults refers to the proportion of adults in a household and adults are defined as aged 18 years or older. Males is the proportion of males in a household. Gender is a dummy variable which takes on the value of 1 if the household head is male and zero if they are female. Work is a dummy variable which takes on the value of 1 if the household head is employed and zero otherwise. Year takes on the value of 1 if the household was surveyed in 2005 and zero if surveyed in 2009. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Cluster fixed effects are suppressed for space reasons but are jointly statistically significant at the 1% level for the two specifications of the data.

CHAPTER 4: EVALUATING SOUTH AFRICA'S TOBACCO CONTROL INITIATIVE: A SYNTHETIC CONTROL APPROACH

4.1. Introduction

South Africa has since 1994 aggressively and consistently increased the excise tax on cigarettes so as to meet and maintain a total tax burden (including Value Added Tax) of 50% of the average retail selling price. The target was met in 1997 and revised upwards to 52% in 2004. The tax rises have translated into substantial increases in the inflation-adjusted retail selling prices of cigarettes. For instance, the average real price per pack of cigarettes increased by 110% between 1994 and 2004 and by 190% if one extends the period to 2012 (see Figure 4.1). The increase in prices has coincided with substantial declines in prevalence and consumption. Van Walbeek (2005) estimated that prevalence declined from 31% of the adult population in 1993 to 24% in 2003 while aggregate cigarette consumption and per capita consumption declined by 32% and 46% respectively over the same period. Real excise tax revenues increased during this time in spite of the fact that cigarette consumption was declining. In 2012 rands, real annual tax revenues increased from R3billion in 1994 to R11billion in 2012, an increase of 270%.[45]

Declines in prevalence and consumption were well underway by the time the tax increases began in 1994 (Van Walbeek, 2002; 2005). In the absence of a credible counterfactual (a what-if scenario), the impact of taxes on consumption and prevalence is likely to be overstated. The literature evaluating the impact of South Africa's aggressive tobacco control efforts is not very extensive.

This chapter uses a transparent data-driven technique, the Synthetic Control method developed by Abadie and Gardeazabal (2003) and extended in Abadie *et al.* (2010), to create a credible counterfactual of cigarette consumption in South Africa from 1994 to 2004. The counterfactual is constructed as a weighted average of the per capita cigarette consumption of countries similar to South Africa that did not initiate large-scale tobacco control measures over the period 1994 to 2004. Using this counterfactual, I am able to estimate a "treatment effect" of South Africa's tax increases on cigarette consumption. I find that per capita cigarette consumption would not have continued declining in the absence of the consistent tax and price rises that began in 1994. Specifically, I estimate

[45] Tax revenue data are taken from the National Treasury of South Africa and CPI from Statistics South Africa.

a treatment effect of 36% by 2004. That is, per capita cigarette consumption in 2004 was 36% lower than it would have been had the government not consistently increased excise taxes in the preceding years.

The rest of this chapter is structured as follows: Section 4.2 provides some background to South Africa's tobacco control measures. Section 4.3 reviews the literature evaluating tobacco control measures in South Africa and in other parts of the world. Section 4.4 describes the Synthetic Control method in some detail and Section 4.5 describes the data. Section 4.6 discusses the selection of the control countries (what I call the donor pool) while Section 4.7 presents the main results and conducts placebo tests. I present the results of the robustness tests in Section 4.8 while Section 4.9 discusses what implications, if any, illicit trade has for my estimates of the treatment effect. Section 4.10 concludes.

4.2. Tobacco Control in South Africa

Prior to 1994, South Africa did not consciously target the consumption of tobacco products on public health grounds. According to Van Walbeek (2005), the relegation of public health concerns in tobacco tax policy was likely due to the cordial relations that existed between the tobacco industry and the National Party, the party that ruled South Africa from 1948 to 1994. The end result was that the real tax on cigarettes, the main tobacco product in South Africa, declined by 70% between 1961 and 1990 (ibid.). Coincidentally, per capita cigarette consumption increased by 60% from 50 packs in 1961 to 80 packs in 1991 (ibid.).

In the 1980s and early 1990s, the medical research community (Yach, 1982) and the South African Medical Research Council (1988, 1992) published research showing that tobacco consumption imposed a net cost on the country. For instance, the 1992 study by the South African Medical Research Council (SAMRC) estimated the costs of tobacco consumption at 1.82% of GDP against benefits of 0.49% of GDP (SAMRC, 1992). The publicity generated by these studies rallied the public health community and civil society behind the common goal of getting the South African government to take tobacco control seriously. The momentum that had built up during the 1980s and early 1990s, along with the impending change of government, culminated in the passing of the Tobacco Products Control

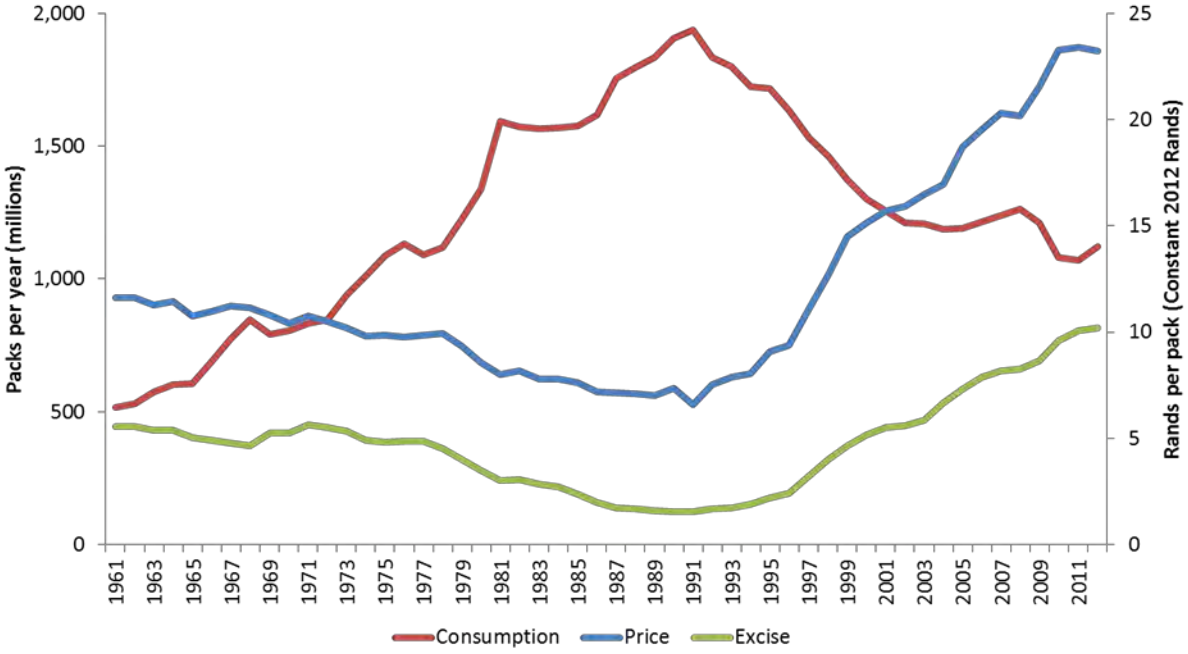
Act of 1993 by Parliament.[46] The big turning point, however, came in 1994 when the new African National Congress-led government announced that the government would target a tax burden on cigarettes (including Value Added Tax) of 50% of the retail price to be phased in over a number of years (Republic of South Africa, 1994). As a result, 1994, 1995 and 1996 saw excise tax increments of respectively 25%, 25% and 18% (Republic of South Africa, 1994, 1995, 1996). In 1997, the Minister of Finance announced a large increase of 52% in the excise tax on cigarettes, a move that was expected to bring the total tax burden (including Value Added Tax) to 50% of the average retail selling price (Republic of South Africa, 1997). From 1997, the annual increases on excise taxes on cigarettes have, therefore, been predictable in order to maintain the stipulated tax burden.[47] In 2004, the total tax burden was revised upwards to 52% of the average retail selling price (Republic of South Africa, 2004).

South Africa's aggressive excise tax policy since 1994 has translated into substantial increases in the real price of cigarettes. From 1994 to 2012, the average real price per pack of cigarettes increased by 190% (see Figure 4.1). Between 1994 and 2004, which is the period I evaluate in this chapter, the increase in the real price per pack was 110%. This is in stark contrast to the period before 1994 which saw considerable declines in the real price of cigarettes. It is this unprecedented increase in real cigarette prices, beginning in 1994, whose impact on consumption I seek to evaluate in this chapter.

[46] Saloojee (1994), Malan and Leaver (2003) and Van Walbeek (2005) contain detailed accounts of the events and debates leading up to the adoption of the Tobacco Products Control Act of 1993.

[47] Because the industry responds by increasing retail prices, the tax burden is always slightly less than the government's target (see Van Walbeek, 2005, 2006).

Figure 4.1: Trends in the excise tax per pack of cigarettes, real price per pack of cigarettes and consumption of cigarettes in packs, South Africa 1960 to 2012



Notes: Based on data from the National Treasury of South Africa and Statistics South Africa.

4.3. Literature review

The literature evaluating the impact of South Africa's tax increases since 1994 on prevalence and consumption is not very extensive. Van Walbeek (2002, 2005) investigated the impact of the tax increases on prevalence and consumption by fitting a linear trend to the All Media and Products Survey (AMPS), which is a commercially generated dataset. He estimated that smoking prevalence in South Africa declined from 31% of the adult population in 1993 to 24% in 2003. He also found that African and Coloured population groups experienced the biggest declines in prevalence over the same period. In terms of consumption, Van Walbeek (2005) found that aggregate consumption declined by 32% over the period 1993 to 2004 whereas per capita consumption declined by 46%. Boshoff (2008) estimated a vector autoregression (VAR) in an attempt to assess the relative importance of price changes, income changes and general health awareness in influencing cigarette demand over the period 1996 to 2006. He estimated demand elasticities for price, income and health awareness and found that all three factors were important in influencing cigarette demand over the period 1996 to 2006. Bosch and Koch (2014) examined the regressivity of a large tax increase in 2006 by exploiting the fact that data collection for the Income and Expenditure Survey coincided with the tax increase.

They found that cigarette taxes became less regressive after the tax increase. Other work has instead focussed on estimating the impact of the tax increases on illicit trade (Blecher, 2010, 2011; Van Walbeek, 2014) and on the impact of the Tobacco Products Control Amendment Act of 1999 on restaurant revenues (Blecher, 2006; Van Walbeek *et al.*, 2007).

An implicit assumption in the South African literature that evaluates tobacco control measures is that pre-intervention trends in, say, prevalence and consumption would have continued in the absence of the intervention. Given this, impact can be assessed by comparing present day consumption or prevalence with the magnitudes of these variables before the onset of treatment. That pre-intervention trends would have continued without treatment is not directly evident for consumption and prevalence. The two variables were already declining by the time the government introduced its new tax policy on tobacco products (Van Walbeek, 2002, 2005). Any evaluation that does not attempt to create a counterfactual is likely to give a biased estimate of the impact of a particular tobacco control measure.[48]

Internationally, researchers in the United States (US) have made progress in evaluating tobacco control initiatives by using counterfactuals. Warner (1977) was one of the first US studies to do this. He estimated a regression of per capita cigarette consumption on cigarette price and other covariates using time series data from before the Surgeon-General's 1964 report and the ensuing anti-smoking campaigns. He then used the estimated coefficients from the pre-treatment period to predict what per capita cigarette consumption would have been like in the absence of the campaigns. The treatment effect was then calculated as the difference between predicted consumption and actual consumption in the treatment period. Warner's conclusion was that per capita cigarette consumption would have been 30% higher in 1975 had the anti-smoking measures not happened. Implicit in Warner's study was the assumption that the pre-intervention regression coefficients remained stable even after the intervention. That this assumption was unlikely to hold, in general settings, was pointed out by Lucas (1976) in his important critique of econometric models of policy evaluation.

More recent work in the US has focussed on developing methods of conducting policy evaluations that avoid some aspects of Lucas's critique. For instance, Fichtenberg and Glantz (2000) evaluated

[48] Koch and Tshiswaka-Kashalala (2008, p2) make a similar point.

California's tobacco control programme by comparing rates of change in per capita cigarette consumption in California against rates of change in the rest of the US. They found that after the introduction of the programme in 1989, California's rate of decline in per capita cigarette consumption exceeded that of the rest of the US by 2.72 packs per year. A critique of the method in Fichtenberg and Glantz (2000) is that treatment effects were underestimated since the rest of the US included states that, alongside California, had also implemented some tobacco control measures.[49] The method in Abadie *et al.* (2010), which I describe fully below, attempts to correct for this shortcoming by comparing California to only those states that did not implement large-scale tobacco control measures after 1989.

4.4. Method

This chapter uses the method developed by Abadie and Gardeazabal (2003) and extended further in Abadie *et al.* (2010) to evaluate South Africa's tobacco control policies from 1994 to 2004. The method involves estimating South Africa's counterfactual cigarette consumption trend line following the consistent hikes in cigarette excise taxes that began in 1994. In other words, the method involves creating a synthetic South Africa, a country that looks like South Africa in all relevant respects except for the tax hikes. The observed outcome variable for the "real" South Africa is then compared to the outcome variable for the synthetic South Africa. In this section I discuss in some detail the formal aspects of the method.

4.4.1. Identification

Suppose I have $J + 1$ regions and region 1 experiences a policy change and is therefore referred to as the "treated" region. The remaining J regions do not experience the policy change and since I use these regions to construct a counterfactual scenario for the treated country, I collectively refer to them as the "donor" pool. The policy change happens at time period T_0 where $1 \leq T_0 < T_0 + P$ with P being the number of time periods after treatment. In the case of South Africa, $P = 10$ and $T_0 = 1994$ (Below I motivate why I choose to end the evaluation 10 years after 1994). The outcome variable of interest is Y_{it} with $i = 1,2, \dots, J + 1$ and $t = 1, \dots, T_0 + P$. For any region i and time period t, I can define Y_{it}^I and Y_{it}^N. Y_{it}^I is the observed outcome variable and Y_{it}^N is the outcome variable in the absence

[49] For example, Alaska, Hawaii, Maryland, Michigan, New Jersey, New York and Washington had raised their state cigarette taxes by at least 50 US cents over the period 1989 to 2000 (Abadie *et al.*, 2010).

of treatment (the superscripts I and N are chosen to represent respectively “intervention” and “no intervention”). That is, Y_{it}^N is unobserved after T_0 but is equal to Y_{it}^I before T_0. Given this, I can then define the treatment effect of the policy change, α_{it}, as:

$$\alpha_{it} = Y_{it}^I - Y_{it}^N \qquad (4.1)$$

for $t = T_0 + 1, \ldots, T_0 + P$. The complication is that Y_{it}^N is unobserved for all $t > T_0$. In order to estimate the effect of the policy change, I need to estimate Y_{it}^N after treatment. Suppose Y_{it} evolves according to the equation

$$Y_{it} = \lambda_t + \boldsymbol{\theta}_t \boldsymbol{Z_i} + \boldsymbol{\delta}_t \boldsymbol{\mu}_i + \varepsilon_{it} \qquad (4.2)$$

where λ_t is some factor common to all regions, $\boldsymbol{Z_i}$ is a vector of observed factors and $\boldsymbol{\mu}_i$ is a vector of unobserved factors that have an impact on Y_{it}. $\boldsymbol{\theta}_t$ and $\boldsymbol{\delta}_t$ are the unknown time varying parameters associated with $\boldsymbol{Z_i}$ and $\boldsymbol{\mu}_i$ respectively.[50] ε_{it} is the unobserved error term with mean zero. Given a donor pool and a $J \times 1$ vector of weights $\boldsymbol{W} = (w_2, \ldots, w_{J+1})'$ such that $w_j \geq 0$ and $w_2 + w_3 + \cdots + w_{J+1} = 1$, I can always construct for any i

$$\sum_{j=2}^{J+1} w_j Y_{jt} = \lambda_t + \boldsymbol{\theta}_t \sum_{j=2}^{J+1} w_j \boldsymbol{Z_j} + \boldsymbol{\delta}_t \sum_{j=2}^{J+1} w_j \boldsymbol{\mu}_j + \sum_{j=2}^{J+1} w_j \varepsilon_{jt} \qquad (4.3).$$

That is, I can always express the outcome variable of a treated region as a weighted average of the regions in the donor pool. For $i = 1$ (i.e. the treated country), Abadie and Gardeazabal (2003) and Abadie *et al.* (2010) show that there exists a $J \times 1$ vector of weights $\boldsymbol{W}^* = (w_2^*, \ldots, w_{J+1}^*)'$ with $w_2^* + w_3^* + \cdots + w_{J+1}^* = 1$ and $w_j^* \geq 0$ such that

[50] Notice that $\boldsymbol{Z_i}$ and $\boldsymbol{\mu_i}$ do not have time subscripts. We can think of their values as fixed over short periods of time but still allow for their effects, via $\boldsymbol{\theta}_t$ and $\boldsymbol{\delta}_t$ respectively, to vary across time. The method also allows for more general specifications of $\boldsymbol{Z_i}$ and $\boldsymbol{\mu_i}$ with time subscripts.

$$\sum_{j=2}^{J+1} w_j^* Y_{j1} = Y_{11}$$

$$\sum_{j=2}^{J+1} w_j^* Y_{j2} = Y_{12}$$

.

. (4.4)

.

$$\sum_{j=2}^{J+1} w_j^* Y_{jT_0} = Y_{1T_0} \quad and$$

$$\sum_{j=2}^{J+1} w_j^* \boldsymbol{Z_j} = \boldsymbol{Z_1}$$

That is, I can always exactly recreate the pre-treatment characteristics of the treated region using only the donor pool and the weights in $\boldsymbol{W}^*$.[51] Since the factors in $\boldsymbol{\mu}_i$ are unobserved, I cannot create their empirical counterparts in equation (4.4). However, if the set of equations in (4.4) hold exactly, then

$$\sum_{j=2}^{J+1} w_j^* \boldsymbol{\mu_j} = \boldsymbol{\mu_1} \qquad (4.5)$$

also holds (Abadie *et al.*, 2010, p495). Having recreated the pre-treatment characteristics of the treated country using the donor pool, I can then use the same linear combination of regions to trace out the time path of the outcome variable after treatment. This time path is the outcome variable I would have observed for the treated region in the absence of treatment (the counterfactual). The difference between the counterfactual trend line and the actual trend line is then an estimate of the treatment effect. Formally, given equations (4.4) and (4.5), the treatment effect estimator for $i = 1$ is

[51] Appendix B of Abadie *et al.* (2010) contains the mathematical proofs related to this point.

$$\widehat{\alpha_{1t}} = Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt} \qquad (4.6)$$

for $t > T_0$ and $j = 2,3, \dots, J$.[52]

The treatment effects estimator in equation (4.6) is a generalized version of the standard difference-in-difference estimator (Angrist and Pischke, 2009, p227 – 243). Whereas the standard difference-in-difference estimator assumes that the effects of the unobserved factors are fixed and therefore can be "differenced" out, (4.6) allows for them to be time varying. This is an attractive property given that the impact of most factors is likely to change over time as opposed to remaining fixed. In addition, the treatment effects estimator in equation (4.6) is a dynamic estimator that gives us the treatment effect at each point in time after treatment. The standard difference-in-difference estimator only gives a static average treatment effect. Further, Abadie *et al.* (2015) show that the Synthetic Control estimator in (4.6) is related to the standard regression estimator in the sense that both apply the idea of weights that sum to one. The only difference is that the Synthetic Control estimator restricts the weights to be non-negative, whereas the regression estimator places no such restriction on the weights. Not placing this restriction allows regression to perfectly fit a counterfactual even when the data does not allow for one. In more technical terms, regression allows extrapolation outside the support of the data whereas the synthetic control estimator can only perfectly fit a counterfactual if the data allows it to do so. Extrapolating from outside of the support of the data makes regression susceptible to the problem of "extreme counterfactuals" (King and Zeng, 2006).

The equations in (4.4) are unlikely to hold exactly in practise. It is, therefore, desirable to get as close approximations to these equations as possible. One of the ways of assuring this is to have a donor pool of regions that share a "common support" with the treated region. In other words, the outcome variable for the regions in the donor pool should be influenced by the same factors as the outcome variable for the treated region. That is, the outcome variable for both types of regions should evolve according to equation (4.2). Secondly, the treated region should be contained within the set of all linear combinations of the donor pool. This is technically known as the "convex hull" requirement (King and Zeng, 2006). These two conditions essentially require the treated region to not be too

[52] For $t < T_0$, $\widehat{\alpha_{1t}} = 0$. That is , before treatment, the treatment effect is zero.

extreme relative to the regions in the donor pool. In any case, the degree of pre-treatment discrepancy between the treated country and its synthetic counterpart can be assessed by calculating the Root Mean Square Error (RMSE) as:

$$RMSE = \left(\frac{1}{T_0} \sum_{t=1}^{T_0} (Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt})^2 \right)^{\frac{1}{2}} \qquad (4.7).$$

A large RMSE would suggest a poor pre-treatment fit between the treated region and its synthetic counterpart. Using the Synthetic Control Method in this situation would not be advisable.

$\boldsymbol{W}^*$ (the vector of optimum weights) is chosen as the solution to the following constrained optimization problem:

$$\min_{\boldsymbol{W} \in \boldsymbol{\mathcal{M}}} \| \boldsymbol{X}_1 - \boldsymbol{X}_0 \boldsymbol{W} \| = \sqrt{(\boldsymbol{X}_1 - \boldsymbol{X}_0 \boldsymbol{W})' \boldsymbol{V} (\boldsymbol{X}_1 - \boldsymbol{X}_0 \boldsymbol{W})} \text{ such that } w_j \geq 0 \text{ and } w_2 + w_3 + \cdots + w_{J+1} = 1 \qquad (4.8)$$

where $\boldsymbol{X}_1$ is a matrix of pre-intervention characteristics of the treated region (including Y_{1t} and $\boldsymbol{Z_1}$) and $\boldsymbol{X}_0$ is a matrix of the same pre-intervention characteristics for the regions in the donor pool. $\boldsymbol{\mathcal{M}}$ is the set of all vectors satisfying the requirement that their elements sum to one and are non-negative[53] and $\boldsymbol{V}$ is some diagonal matrix whose diagonal elements weight factors in $\boldsymbol{Z_1}$ according to how well they predict the outcome variable Y_{it}. The problem in (4.8) seeks to minimize, by selecting $\boldsymbol{W}^*$, a measure of distance between the treated region and the donor pool.[54] The minimization problem in (4.8) can be solved numerically in Stata using the *Synth* routine.[55]

---

[53] For instance, $\boldsymbol{\mathcal{M}}$ might contain a vector with the following elements (1 0 0 … 0) or another vector with elements (0.5 0 0 … 0.5) and so on.

[54] Recall that ‖ ‖ is the Euclidean norm or Euclidean metric, a distance function.

[55] Available from Jens Hainmueller's website at http://web.stanford.edu/~jhain/synthpage.html

4.4.2. Inference

In order to ensure that the treatment effect identified in equation (4.6) is not due to random chance, Abadie *et al.* (2010, 2015) suggest inferential techniques based on the idea of placebo tests. They suggest constructing synthetic counterparts for all the regions in the donor pool, one at a time, and for each region estimating a treatment effect according to equation (4.6). This exercise results in the construction of an empirical distribution of treatment effects similar to the student's t distribution. The identified effect for the treated region is statistically significant (i.e. not due to chance) if the probability of obtaining an effect as large as that of the treated region, in the empirical distribution of treatment effects, is small. In other words, the effect for the treated region is statistically significant if the number of donor regions that show a treatment effect, even after receiving a placebo, is small.[56]

4.4.3. Implementation

In terms of implementing the method for South Africa, I follow the approach in Abadie *et al.* (2010). Y_{it}, the outcome variable, is cigarette consumption per capita (in sticks). The vector $\boldsymbol{Z_1}$ comprises of the standard predictors of cigarette demand found in the literature (Chaloupka and Warner, 2000; IARC, 2011). The variables in $\boldsymbol{Z_1}$, include the following : the real price of a pack of cigarettes, real Gross Domestic Product per capita (real GDP per capita), alcohol consumption per capita (expressed in litres of pure alcohol) and the proportion of adults in the total population. Alcohol consumption is included in the analysis given that recent work has shown that there is a positive and statistically significant association between cigarette and alcohol consumption (Dierker *et al.*, 2006; Cohn *et al.*, 2015). $\boldsymbol{Z_1}$ also includes lagged values of per capita cigarette consumption to capture some aspect of habit formation (Warner, 1977; Chaloupka, 1991). The data sources for all these variables are discussed in detail in Section 4.5 below.

My choice of conducting the evaluation over the period 1994 to 2004 is due to the World Health Organization's Framework Convention on Tobacco Control (FCTC) which came in to effect in 2005. As discussed in Chapter 1, the treaty encourages countries to implement a wide array of tobacco control measures. I, therefore, expect that most of the countries in my donor pool began, from 2005 onwards, to think seriously about tobacco control, a situation that might result in a downward bias in my

[56] This idea is borrowed from medical trials, where patients receiving a placebo are not expected to show results that are similar to patients receiving the actual drug, if the drug is effective.

treatment effect estimates. Further, Abadie *et al.* (2010, 2015) consider a ten year period to be a sufficient timespan to properly evaluate the effects of a policy change.[57]

The Synthetic Control method has gained prominence after being favourably reviewed by Imbens and Wooldridge (2009) in their extensive survey of the impact evaluation literature. It has been used to assess episodes of economic liberalization across the world (Billmeier and Nannicini, 2013), to quantify the economic costs of conflict in Spain (Abadie and Gardeazabal, 2003) and the economic effects of reunification in Germany (Abadie *et al.*, 2015). From a public health perspective, the method has been used to evaluate California's Tobacco Control Programme (Abadie *et al.*, 2010), to quantify the health benefits of the liberalization of the sex trade in the US state of Rhode Island (Cunningham and Shah, 2014) and to estimate the effect of bar closing times on traffic accidents in the United Kingdom (Green *et al.*, 2014).

4.5. Data

The data used in this chapter come from a number of sources. Data on the outcome variable, cigarette consumption per capita (in sticks), come from the World Cigarette Report published by the ERC Group (ERC, 2010). The ERC Group is an independent research company that compiles market intelligence data on an annual basis on a number of products, including cigarettes. The country coverage of the World Cigarette Report is extensive and also contains complete time series on cigarette consumption from 1990 to 2009. Consumption data from the report has been used previously by Blecher (2011) to investigate the impact of advertising bans on cigarette consumption.[58]

[57] In their 2010 paper on California's tobacco control initiative, Abadie *et al.* evaluate the initiative's effect for the period running from 1989 to 2000. In their 2015 paper on the economic effects of reunification on West Germany's economy, Abadie *et al.* conduct the evaluation over the period 1990 to 2000.

[58] An alternative data source for consumption is the Tobacco Country Profiles available from the World Health Organization (WHO) at http://www.who.int/tobacco/surveillance/policy/country_profile/en/. Unfortunately, and as noted by Blecher (2011, p139), the Tobacco Country Profiles do not contain complete consumption series for the time periods that I am interested in.

Cigarette price data is from the Economist Intelligence Unit's (EIU's) Worldwide Cost of Living Survey. The survey has been collecting cigarette price data alongside the price of other goods and services for 140 cities since 1990.[59] For cigarettes, prices are collected semi-annually from supermarkets, medium-priced retailers and more expensive specialty stores for two brands: Marlboro (or the nearest international equivalent) and the cheapest local brand (or the cheapest brand in the absence of a local brand). I follow Blecher and Van Walbeek (2004, 2009) and Blecher (2008, 2011) and use the price of a pack of the cheapest brand. This is because the cheapest brand is usually the most popular brand in a country and consequently its price is the most representative. The price data is expressed in constant 2000 US dollars using the United States Consumer Price Index City Average for All Items (United States Department of Labour).[60] A drawback of using the EIU price data is that cigarette prices are only collected from a few cities (sometimes only a single city) within a country. This might reduce the representativeness of the price data.

GDP per capita and data on the proportion of adults (16 to 64 years) in the population come from the World Bank's World Development Indicators database.[61] GDP per capita is expressed in constant 2000 US dollars. Finally, data on alcohol consumption per capita (in litres of pure alcohol) comes from the World Health Organization's Global Information System on Alcohol and Health.[62]

4.6. Selection of the Donor Pool

The validity of the Synthetic Control method relies on the selection of a donor pool that meets the following set of criteria: (i) the common support requirement, (ii) the convex hull requirement and (iii) regions in the donor pool should not have experienced treatment during the relevant time period. In selecting an appropriate donor pool, I begin by addressing the third requirement and then work backwards to (i) and (ii).

[59] For more see: http://www.eiu.com/handlers/PublicDownload.ashx?fi=data-section/worldwide-cost-of-living.pdf&mode=m

[60] Available at www.bls.gov

[61] Available at http://data.worldbank.org/data-catalog/world-development-indicators

[62] Available at http://apps.who.int/gho/data/node.main.GISAH

In order to select a donor pool consisting of untreated countries, I rely on the work on cigarette affordability by Blecher and Van Walbeek (2004, 2009). Blecher and Van Walbeek propose a measure of cigarette affordability, the Relative Income Price (RIP), which is calculated as the ratio of the cost of 100 packs of cigarettes in a country to that country's real GDP per capita. A declining RIP means that cigarettes are becoming more affordable while a rising RIP signifies declining affordability. In their 2009 paper, Blecher and Van Walbeek were able to classify 77 countries according to whether they experienced increasing affordability or declining affordability over the period 1990 to 2006. These were countries for which the authors were able to obtain complete and comparable data on real cigarette prices and real GDP per capita over the period 1990 to 2006. The authors identified 37 countries where cigarettes became more affordable over the period 1990 to 2006.[63] For 20 out of the 37 countries, the increase in affordability occurred because of a decrease in the real price of cigarettes coupled with an increase in real GDP per capita. For the remaining 17 countries, the increase in affordability was due to real GDP per capita growing faster than the increase in real prices.

I opt to use the increase in affordability over the period 1990 to 2006 as a proxy for the absence of treatment. That is, I regard countries whose affordability increased on average over this period as not having enacted significant tobacco control measures. This is obviously the case for the 20 countries where affordability increased as a result of declining real cigarette prices. I contend, however, that even for the remaining 17 countries where affordability increased due to real incomes growing faster than real prices, a conclusion of the absence of treatment is a reasonable one to make. This is because effective tobacco control measures require (i) real tax/price increases and (ii) real tax/price increases that grow faster than the rate of growth in incomes (WHO, 2010; IARC, 2011). I also recognise that the Relative Income Price (RIP) might have some shortcomings in identifying whether a country has instituted tobacco control measures or not. For instance, a country may have adopted a wide set of tobacco control measures such as advertising bans and/or clean indoor air policies but neglected to significantly increase real cigarette prices. My measure of treatment would consign this country to the pool of potential donor countries in spite of its tobacco control efforts. In as much as I recognise that tobacco control measures constitute more than just tax/price measures, the tobacco control literature recognises the primacy of tax/price policies in curbing demand (Chaloupka and Warner, 2000; IARC, 2011). In any case, I would consider my estimates of the treatment effect to be lower bound estimates

[63] See Figure 4 in Blecher and Van Walbeek (2009).

if the donor pool had some countries whose treatment status was misclassified in the manner suggested above.

An alternative approach would be to determine treatment status based on the Tobacco Country Profiles available from WHO.[64] Unfortunately, the country profiles are often not clear as to whether the listed tobacco control measures have been implemented effectively or not. Further, the country profiles often provide the analyst with lots of room for discretion in classifying treatment status. On the other hand, the Relative Income Price (RIP) measures outcomes and not the intent of treatment. Secondly, the RIP, in using a rigid decision rule, leaves the analyst with little room for discretion and in this way limits errors due to misclassification. Lastly, the procedure of assigning treatment based on the RIP is transparent, a hallmark of the Synthetic Control method.

My criterion for identifying treatment correctly classifies many of the countries that are known for having instituted significant tobacco control measures over the period 1990 to 2006. For example, South Africa, the country of interest in this chapter, is correctly classified as treated since its Relative Income Price (RIP) increased (i.e. affordability declined) on average over the period 1990 to 2006. Thailand, a country whose positive experience with tobacco control is often held up as a model for other Low- and Middle-Income Countries (Levy *et al.*, 2008; Sangthong *et al.*, 2012), is also classified as having undergone treatment. Most of the developed countries, whose tobacco control efforts predate the 1990s, are also classified correctly as treated. On the other hand, the list of untreated countries consists mainly of Low- and Middle-Income Countries (LMICs), an expected outcome given these countries' slow progress in implementing effective tobacco control measures over the period 1990 to 2004 (Jha and Chaloupka, 2000). The full list of treated and untreated countries from Blecher and Van Walbeek's 2009 paper are contained in Table C1 in Appendix C.

Having identified the potential donor pool, I need to ensure that the common support and convex hull requirements are met. The two requirements are readily satisfied by excluding from my potential donor pool in Table C1 countries that are dissimilar to South Africa in some fundamental way. One of the most transparent ways of ensuring this is to use the World Bank's Country Classification System

[64] Available at http://www.who.int/tobacco/surveillance/policy/country_profile/en/

based on per capita income.[65] I rely on Blecher and Van Walbeek's (2009) usage of the Classification System as it stood at the time of writing their paper and exclude from the donor pool all high income countries.[66] These countries are often perceived as being structurally different in many respects to Low- and Middle-Income countries such that including them in the donor pool would risk violation of the convex hull and common support requirements. Lastly, I drop from the potential donor pool countries without a complete set of data for all variables over the period 1990 to 2004.[67] The final donor pool consists of 24 countries which are listed in Table 4.1.

Table 4.1: Donor Pool

Donor Pool	
Argentina	Morocco
Brazil	Pakistan
Chile	Panama
China	Peru
Colombia	Philippines
Costa Rica	Romania
Cote d'Ivoire	Senegal
Ecuador	Sri Lanka
Egypt	Tunisia
India	Uruguay
Indonesia	Vietnam
Jordan	
Malaysia	

Notes: List of untreated countries from Table C1 that are not high income countries and have a complete set of data over the period 1990 to 2004.

The final donor pool consists of countries that are often thought of as South Africa's peers. The list contains Latin American, sub-Saharan African, North African and South-East Asian countries. The donor pool also contains three countries from the BRICS group (Brazil, India and China).[68] The BRICS countries are often thought of collectively as the vanguard of emerging economies.

[65] Available at http://data.worldbank.org/news/2015-country-classifications

[66] This results in the exclusion of Kuwait, Bahrain, Czech Republic, Ireland, Denmark, Greece, Finland, Luxembourg and Norway.

[67] This results in the exclusion of Bangladesh, Croatia, Iran and Serbia and Montenegro.

[68] BRICS stands for Brazil, Russia, India, China and South Africa. Russia is not in the donor pool as it was classified as treated according to criterion outlined in Section 4.6.

4.7. Main results

This section presents the main results of implementing the Synthetic Control method for South Africa using the donor pool listed in Table 4.1.

4.7.1. Treatment effects

Table 4.2 presents the results of the solution to the minimization problem stated in equation (4.8). According to Table 4.2, synthetic South Africa is a linear combination of 27.6% of Argentina, 47.6% of Brazil, 14.6% of Chile, 0.7% of Romania and 9.4% of Tunisia. In other words, this combination of countries with their respective weights, produces the lowest pre-treatment root mean square error (RMSE) between the actual South Africa and its synthetic counterpart. The pre-treatment RMSE between the actual South Africa and its synthetic counterpart obtained by applying the weights in Table 4.2 is 0.144. That is, on average, the pre-treatment difference between South Africa and synthetic South Africa for the outcome variable is about one-tenth of a per capita cigarette. The optimal weights in Table 4.2 show that synthetic South Africa is mostly made up of Latin American countries (with a combined weight of 90%) with Brazil being the most important.

Table 4.2: Synthetic weights

Country	Weight
Argentina	0.276
Brazil	0.476
Chile	0.146
China	0
Colombia	0
Costa Rica	0
Ecuador	0
Egypt	0
India	0
Indonesia	0
Cote d'Ivoire	0
Jordan	0
Malaysia	0
Morocco	0
Pakistan	0
Panama	0
Peru	0
Philippines	0
Romania	0.007
Senegal	0
Sri Lanka	0
Tunisia	0.094
Uruguay	0
Vietnam	0

Notes: The table shows the vector of optimal weights, $\boldsymbol{W}^*$, obtained as the solution to the problem in equation (4.8).

In Table 4.3, I compare the average pre-treatment characteristics, the predictors in $\boldsymbol{Z_1}$, for South Africa with its synthetic counterpart using the weights in Table 4.2. The table shows that synthetic South Africa resembles the actual South Africa in most of the pre-treatment characteristics. The only variable whose pre-treatment average differs between South Africa and its synthetic counterpart is alcohol consumption per capita: South Africa's average is somewhat higher than its synthetic counterpart. This is due to the fact that South Africa's alcohol consumption per capita is "extreme" relative to the countries in the donor pool. In other words, there is no linear combination of countries in the donor pool than can perfectly reproduce South Africa's alcohol consumption profile (i.e. in terms of alcohol consumption, South Africa is unlikely to be in the convex hull of the donor pool). Having one or two predictors that differ in magnitude between the treated country and its synthetic counterpart is typical of the Synthetic Control method as the treated country is likely to have some "extreme" predictors.[69]

Table 4.3: Average pre-treatment characteristics for South Africa and Synthetic South Africa

	South Africa	Synthetic South Africa
Log of GDP per capita	8.44	8.28
Price of cigarette pack in USD	0.89	1.15
Pure alcohol consumption (in litres)	9.05	7.11
Proportion of adults in population	58.62	60.95
Consumption (1992)	947	945.58
Consumption (1990)	1010	1008.86

Notes: Average pre-treatment characteristics for South Africa and synthetic South Africa. Obtained by applying the weights in Table 4.2 to the pre-treatment characteristics of the donor pool. Alcohol consumption is in litres of pure alcohol per capita.

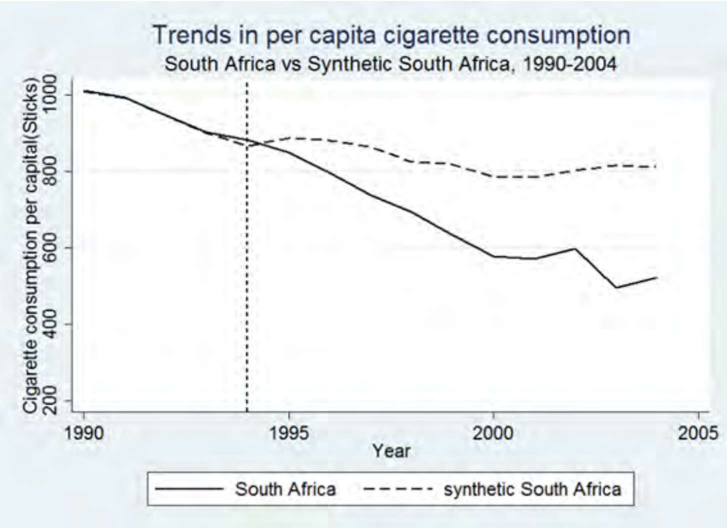
Having shown that synthetic South Africa largely matches actual South Africa in its pre-treatment characteristics (as evidenced in Table 4.3 and by the pre-treatment RMSE), I can now use synthetic South Africa to estimate the treatment effect of the policy change. Figure 4.2 plots cigarette consumption per capita for South Africa and synthetic South Africa over the period 1990 to 2004. The vertical distance between the two lines is the estimate of the treatment effect (see equation 4.6). As one would expect, there is hardly any treatment effect before 1994 as the two lines are indistinguishable from one another. The last point is another way of judging the success of the Synthetic Control method in reproducing South Africa's pre-treatment characteristics.

After the onset of treatment in 1994, the two lines in Figure 4.2 begin to diverge with South Africa's consumption line being everywhere lower than synthetic South Africa's consumption line. South Africa's per capita cigarette consumption declines throughout the treatment period whereas synthetic

[69] In their study assessing the economic costs of reunification on West Germany's economy, Abadie *et al.* (2015) were unable to find a linear combination of donor countries that reproduces West Germany's average pre-treatment inflation rate. This is because West Germany had a very low inflation rate in the pre-treatment period compared to the OECD countries which form the donor pool in their study. Similarly, Abadie and Gardeazabal (2003) in their study of the economic costs of conflict in Spain were unable to reproduce the Basque region's pre-treatment industrial share as a percentage of total production. This is because the Basque region, which is the treated region in their study, had a very high pre-treatment industrial share relative to the rest of Spain.

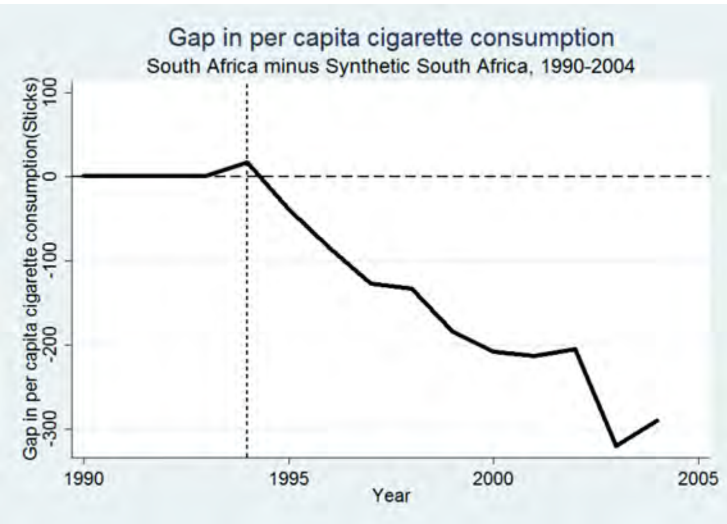
South Africa's trend line initially rises and eventually stabilises at around 800 cigarettes per capita from the year 2000.

Figure 4.2: Cigarette consumption per capita, South Africa vs Synthetic South Africa



Notes: The figure shows the trend lines in per capita consumption of cigarettes for South Africa and its synthetic counterpart. As is clear in the figure, the two lines are indistinguishable before the onset of treatment in 1994 but diverge after treatment.

Figure 4.3: Treatment effect



Notes: The figure shows the gap in per capita cigarette consumption between South Africa and synthetic South Africa over the period 1990 to 2004. The gap is calculated using equation (4.6). As can be seen from the figure, the treatment effect (the gap) is on average zero between 1990 and 1993. Thereafter, it is negative which means that synthetic South Africa has a higher cigarette consumption per capita than actual South Africa during the entire treatment period.

One of the factors that might explain why per capita cigarette consumption stopped declining for synthetic South Africa after 1994 is the performance of the economy. The literature on the demand for cigarettes in South Africa tends to find a positive income elasticity of demand (Reekie, 1994; Van Walbeek, 1996; Economics of Tobacco Control Project, 1998; Van Walbeek, 2005; Boshoff, 2008). That is, on average and *ceteris paribus*, cigarette demand tends to rise with an increase in incomes and tends to fall with a decrease in incomes. Between 1980 and 1994, South Africa's real GDP per capita declined at the average rate of 1% per year.[70] On the other hand, between 1994 and 2004, real GDP per capita increased at the rate of 2% per year. Therefore, the decline in consumption that was already underway by 1994 would likely have stopped, in the absence of tax increases, simply because incomes began to rise. On the other hand, per capita consumption for synthetic South Africa stabilized as opposed to increasing after 1994 likely because other factors such as increased health awareness were also at play (Boshoff, 2008).

Figure 4.3 presents another way of visualizing the treatment effect. The line in the figure measures the cigarette consumption gap between South Africa and its synthetic counterpart (Table C2 in Appendix C provides actual estimates of the treatment effect). Between 1990 and 1993, the treatment effect is approximately zero. By 1995, the first year after treatment begins, South Africa's per capita cigarette consumption is 38 cigarettes less than its synthetic counterpart (or 4% below). The treatment effect increases with each additional year the authorities raise excise taxes on cigarettes so that by 2004, South Africa's per capita cigarette consumption is about 290 cigarettes less than its synthetic counterpart. That is, South Africa's per capita cigarette consumption is 36% lower than where it would have been had treatment not began in 1994.

4.7.2. Placebo tests

The treatment effects from Section 4.7.1 might have been produced by random chance in which case they would not be statistically significant. To confront this assertion, I use the inferential techniques suggested by Abadie *et al.* (2010, 2015) and described in Section 4.4.2. I place South Africa in the donor pool and subject each of the countries in Table 4.1 to the same synthetic control routine as I did for South Africa. This exercise results in a distribution of effects against which South Africa's treatment effects can be compared. South Africa's treatment effects would be statistically significant (i.e. not

[70] Obtained from the World Bank's Development Indicators database available at: http://data.worldbank.org/data-catalog/world-development-indicators (accessed October 2015).

due to random chance) if the probability of obtaining a treatment effect as large as South Africa's, in the distribution of treatment effects, were small. These are called placebo tests because I do not expect many of the untreated countries in Table 4.1 to have treatment effects as large as those observed for the treated country. Figures 4.4 to 4.7 present the results of running the placebo tests. I also include in the figures South Africa's treatment effect from Figure 4.3.

Figure 4.4 presents the treatment effects for all 25 countries. In the figure, most of the countries have treatment effects that are greater than zero or equal to zero over the period 1995 to 2004 (recall from equation 4.6 that a successful treatment results in a negative difference between a country's cigarette consumption per capita and its synthetic counterpart in the treatment period). South Africa's treatment effect appears unusual in the figure although it is matched by Brazil's treatment effect (Brazil's treatment effect is the other line that is also everywhere less than zero). Brazil's pre-treatment fit, with a RMSE of 95, is however poor making it a bad comparison for South Africa which has a pre-treatment RMSE of 0.14. Looked at differently, Brazil's pre-treatment fit is about 600 times poorer than South Africa's pre-treatment fit. Consequently, in Figure 4.5 I do not present the treatment results of countries whose pre-treatment RMSEs are greater than 500 times South Africa's pre-treatment RMSE. This results in the exclusion of four countries.[71] South Africa's unusual treatment effect is now visible. By 2004, no other country has a treatment effect as large as South Africa's. The probability of obtaining a treatment effect as large as South Africa's is 1/21 = 4.76%, which is less than the 5% level used in standard tests of statistical significance. Figures 4.6 and 4.7 continue the exercise of not presenting the treatment results of countries with poor pre-treatment fits. Figure 4.6 excludes countries with a pre-treatment RMSE that is 100 times greater than South Africa's.[72] Figure 4.7 excludes countries with a pre-treatment RMSE that is 50 times greater than South Africa's.[73] The unusual nature of South Africa's treatment effect is now more evident in figures 4.6 and 4.7. The probability of obtaining an effect as large as South Africa's in Figure 4.6 is 1/14 = 7% whereas in Figure 4.7 the probability is 1/10 = 10%. Both probabilities are small given the number of countries in Figures 4.6 and 4.7. Cunningham and Shah (2014) and Dube and Zipperer (2015) make the point that a 10%

[71] Brazil (RMSE = 95), China (RMSE = 281), Romania (RMSE = 139) and Tunisia (RMSE = 123).

[72] In addition to the countries in footnote 71, the following countries are also excluded: Argentina (RMSE = 17), Colombia (RMSE = 32), Costa Rica (RMSE = 39), Egypt (RMSE = 23), India (RMSE = 23), Jordan (RMSE = 28) and Vietnam (RMSE = 26).

[73] In addition to the countries excluded in footnotes 71 and 72, Figure 4.7 excludes Chile (RMSE = 12), Pakistan (RMSE = 9), Panama (RMSE = 13) and Philippines (RMSE = 10).

level is actually a stringent threshold for making inference under the Synthetic Control method given that donor pools usually contain a small number of countries.

Figure 4.4: Placebo test 1

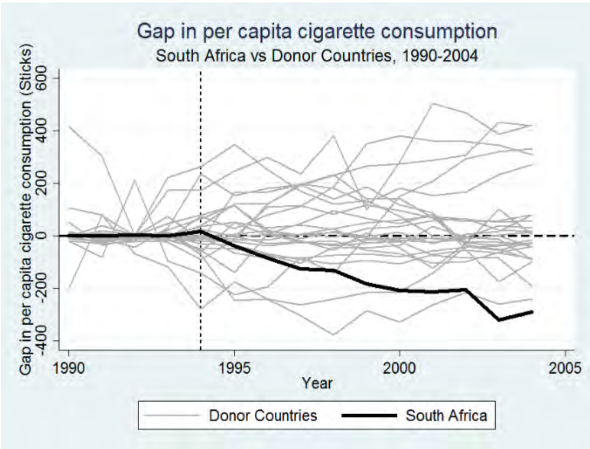


Figure 4.5: Placebo test 2

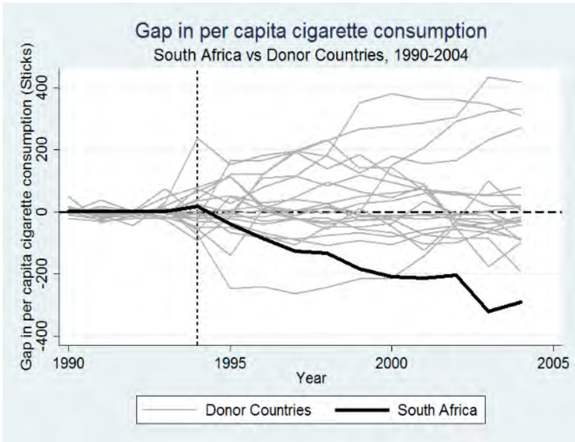


Figure 4.6: Placebo test 3

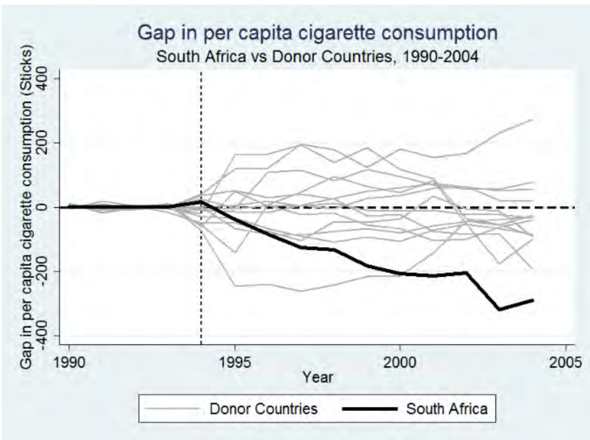
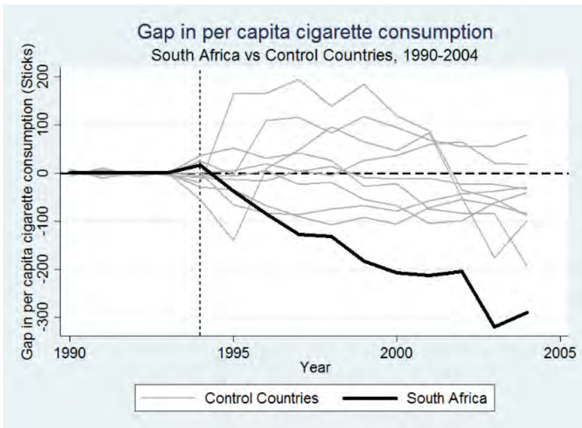


Figure 4.7: Placebo test 4



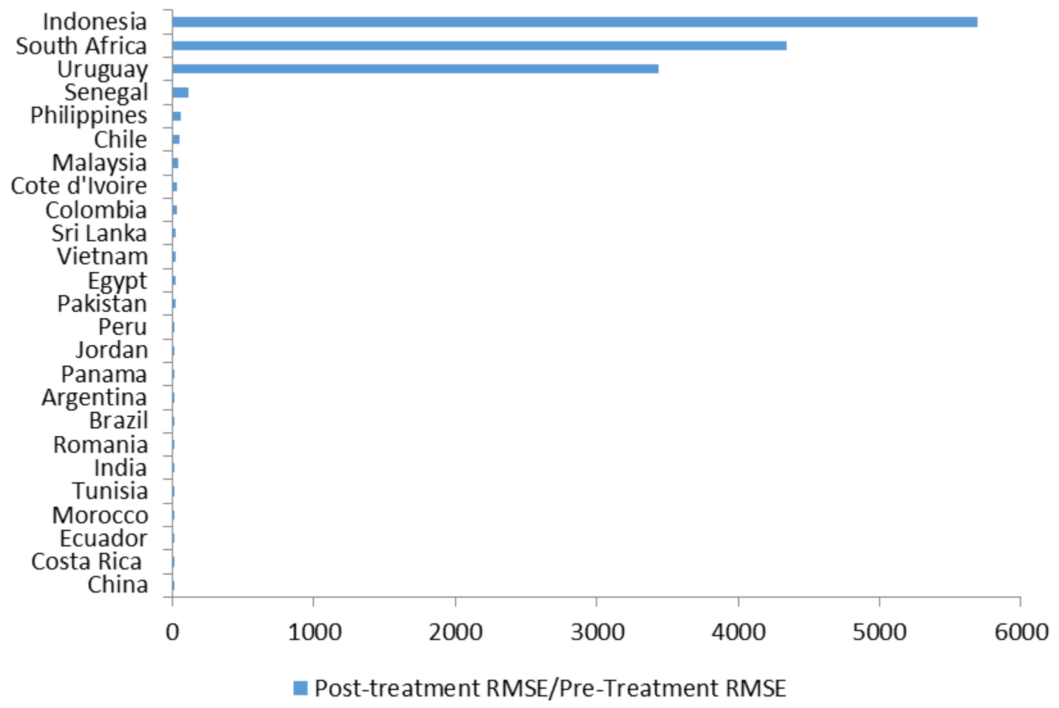
Another way of presenting the results of the placebo tests is to divide each country's post-treatment RMSE by its pre-treatment RMSE and then to rank the ensuing ratios for all countries. This is attractive because it avoids the arbitrary RMSE cut-offs that I used in Figures 4.4 to 4.7 and at the same time penalises countries with large treatment effects but poor pre-treatment fits (like Brazil).[74] Figure 4.8 presents the results of this ranking exercise for all the 25 countries in Table 4.1. In the figure, the ratio for most countries is so small that it is not even visible in the figure (the actual ratios are reported in Table C3 in Appendix C). On the other hand, at about 5, 000, the magnitude of South Africa's ratio is large and is only surpassed by Indonesia's ratio. The results from Indonesia's placebo test cannot, however, be regarded as a successful treatment. This is evident in Figure C1 in Appendix C which plots Indonesia's treatment effect against South Africa's. Indonesia's treatment effect is mostly positive over the period 1995 to 2004 implying that its per capita cigarette consumption is mostly greater than synthetic Indonesia's consumption, a situation that can hardly be described as a successful treatment. Indonesia's unusually high ratio in Figure 4.8 is the result of a very low pre-treatment RMSE relative to South Africa and the fact that the calculation of the post-treatment RMSE does not distinguish between negative and positive treatment effects.[75] The Indonesian case notwithstanding, the probability of obtaining a ratio as large as South Africa's in Figure 4.8 is 2/25 = 8% which is small given the number of countries ("sample size").[76]

[74] This ratio is similar to the t statistic used in standard inferential methods. A large t statistic is obtained whenever the identified effect is large relative to the standard error. The pre-treatment RMSE, in my case, plays the role of a standard error while the post-treatment RMSE plays the role of the identified effect.

[75] The RMSE formula squares and sums over the deviations (which are essentially the treatment effects). See the RMSE formula in equation (4.7).

[76] If I only consider successfully treated countries, then this probability reduces to 1/25 = 4%.

Figure 4.8: Ranking of treatment effects



Notes: The figure shows rankings of ratios of post-treatment root mean square errors (RMSE) to pre-treatment RMSEs for the countries in Table 4.1 plus South Africa.

4.8. Robustness

This section tests the robustness of my treatment effect estimates from Section 4.7. Firstly, I check whether the treatment effects are sensitive to the composition of the donor pool. I do this by excluding, one at a time from the donor pool, the countries in Table 4.2 that have positive donor weights and re-estimating the treatment effect. This is done so as to guard against the possibility that my estimated effects are being driven by a single donor country with a positive weight. Secondly, I vary the timing of the onset of treatment to account for any delays in the implementation of the policy.

Figures 4.9 to 4.13 present the results of successively excluding from the donor pool countries which earlier had positive weights. The pattern of the trajectories of synthetic South Africa is similar across the five figures and, more importantly, similar to the pattern in Figure 4.2. By 2004, the five figures all show a counterfactual consumption level of around 800 cigarettes per capita which was what I found in Figure 4.2. Table 4.4 compares the actual treatment effect estimates of the robustness tests with the main results from Section 4.7. The treatment effects are presented as annual percentage deviations from their respective counterfactual trend lines. Column (2) shows the main results while columns (3) to (7) show the results from excluding, one at a time, donor countries with positive weights from the donor pool. The treatment effect estimates by 2004 are similar across columns (2) to (7). By 2004, all specifications report a treatment effect of at least 30%. My treatment effects estimates are, therefore, not disproportionately influenced by the composition of the donor pool.

Figure 4.9: Excluding Argentina

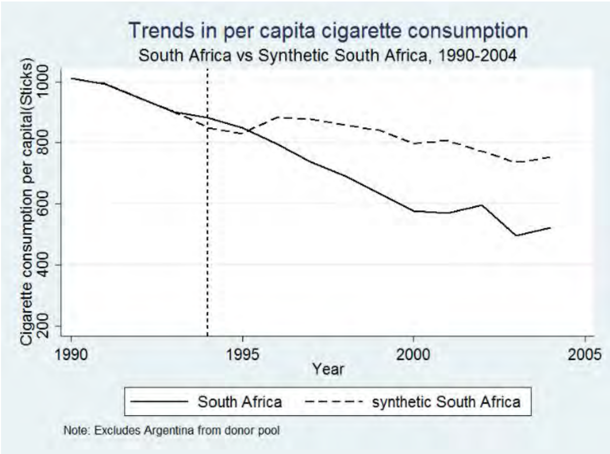


Figure 4.10: Excluding Brazil

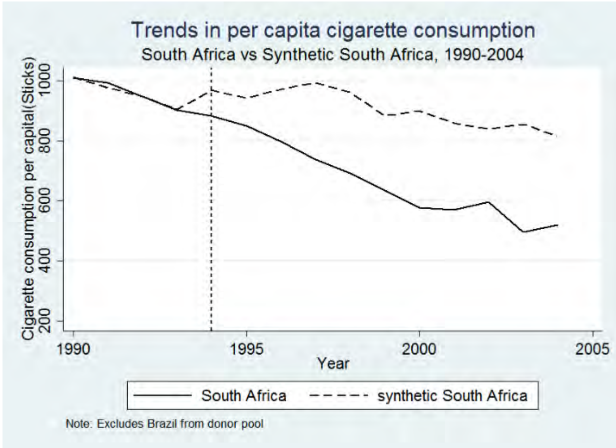


Figure 4.11: Excluding Chile

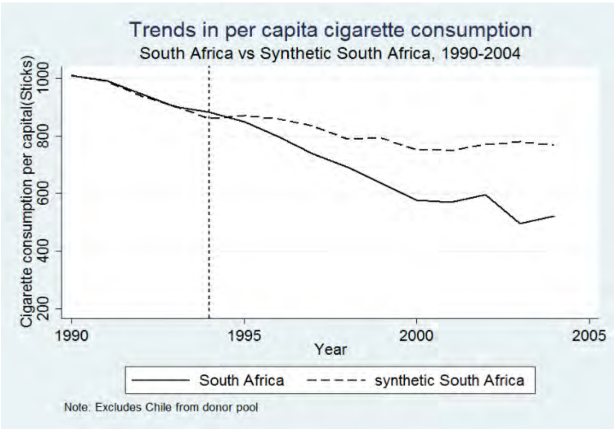


Figure 4.12: Excluding Romania

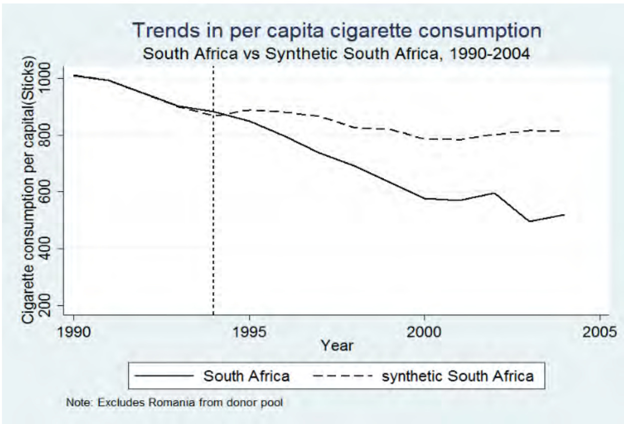


Figure 4.13: Excluding Tunisia

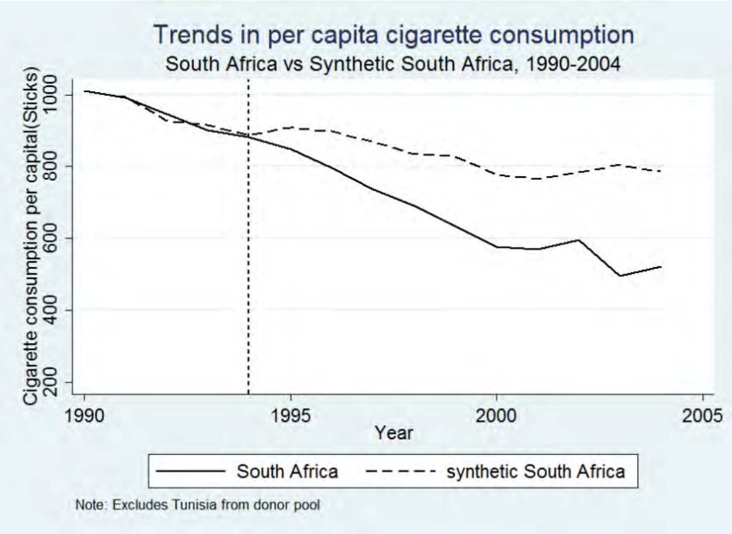


Figure 4.14: Treatment beginning in 1995

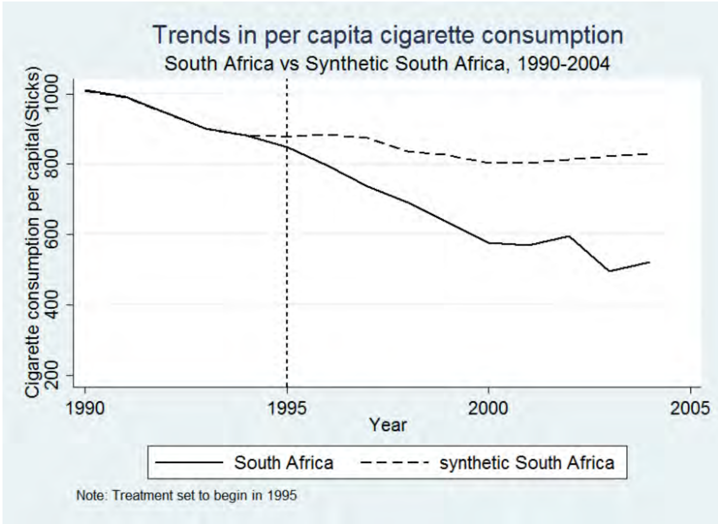


Table 4.4: Treatment effects (in %) associated with robustness tests

Year (1)	Main Results (2)	Excluding Argentina (3)	Excluding Brazil (4)	Excluding Chile (5)	Excluding Romania (6)	Excluding Tunisia (7)	Treatment from 1995 (8)
1990	0.11	0.00	-0.05	-0.03	-0.10	0.00	0.09
1991	0.11	-0.07	1.74	0.36	0.09	-0.03	0.33
1992	0.15	0.03	-0.06	0.85	0.01	2.04	0.11
1993	0.12	-0.14	-0.23	-0.22	0.05	-1.52	0.08
1994	1.97	4.02	-8.74	2.75	1.72	-0.44	0.10
1995	-4.31	2.18	-9.88	-2.56	-4.38	-6.63	-3.28
1996	-9.60	-9.80	-17.99	-7.30	-9.72	-11.32	-10.08
1997	-14.7	-16.02	-25.77	-11.64	-14.94	-15.26	-15.70
1998	-16.07	-19.32	-27.98	-12.32	-16.26	-17.14	-17.29
1999	-22.49	-24.60	-28.16	-19.95	-22.78	-23.48	-23.21
2000	-26.51	-27.63	-35.85	-23.27	-26.72	-25.75	-28.24
2001	-27.22	-29.40	-33.64	-23.96	-27.25	-25.61	-29.06
2002	-25.54	-22.69	-28.83	-22.50	-25.54	-23.86	-26.59
2003	-39.26	-32.71	-42.09	-36.48	-39.34	-38.51	-39.84
2004	-35.75	-30.78	-36.11	-32.11	-35.84	-33.73	-37.19

Notes: The numbers in columns (3) to (8) are treatment effects in percentages associated with the six tests for robustness. The numbers represent annual percentage deviations from their respective counterfactual trend lines. Column (2) reports the main results from Section 4.7. In column (3), Argentina is excluded from the donor pool, column (4) excludes Brazil, column (5) excludes Chile, column (6) excludes Romania and column (7) excludes Tunisia. Column (8) presents results for treatment beginning in 1995 as opposed to 1994.

The final robustness check allows for the possibility that treatment did not begin in earnest in 1994. This is likely to have been the case if the initial tax increase was small relative to the ones in later years or if tobacco companies did not immediately pass-on, in full, the 1994 tax increase.[77] Figure 4.14 and the last column of Table 4.4 (column 8) show treatment effect estimates under the assumption that treatment implementation was delayed by at least a year (i.e. started in 1995). In Figure 4.14, the pattern of the counterfactual trend line is very similar to the one in Figure 4.2 and similar to the ones in Figures 4.9 to 4.13. In the figure, counterfactual cigarette consumption per capita is also around 800 cigarettes by 2004. The treatment effect by 2004 is also similar to the treatment effects obtained for the main result (column 2 of Table 4.4) and for the other donor pool specifications (columns 3 to 7).

[77] Although the available evidence shows that tobacco companies immediately passed-on to consumers some of the tax rise (Van Walbeek, 2006), I nonetheless confront the possibility that full treatment was delayed.

4.9. The impact of illicit trade on the treatment effect

The argument is often made that an aggressive excise tax policy, such as the one that South Africa has been implementing since 1994, might translate into an increase in the market for illicit cigarettes. If this is the case, then the treatment effect estimates from Section 4.7 might be overestimated. Blecher (2010, 2011) has provided some estimates of the size of South Africa's illicit market over most of the period that I study in this chapter. Using several data sources, he obtained an estimate of the illicit market that was implied by smoking prevalence and legal consumption data. For 2004, which is the cut-off point in my evaluation, Blecher estimated an illicit market of between 5% and 12% of the total market.

Using legal consumption data for 2004 from Table C2 in Appendix C (see column 2) and Blecher's estimates of the illicit market in 2004, I can obtain an estimate of the total market (legal and illegal cigarettes) for South Africa. My estimates suggest that the total market for cigarettes in 2004 was somewhere between 548 and 592 cigarettes per capita.[78] Comparing these estimates to synthetic South Africa's estimate for per capita consumption in 2004 (column 3 in Table C2) results in a treatment effect of between 27% and 32%. That is, the treatment effect estimates, when one takes into account the size of the illicit market, are not very different from the main treatment effect estimate of 36% for 2004. In any case, the 27% estimate of the treatment effect, corresponding to Blecher's upper bound estimate of the illicit market share, can be taken to be a lower bound estimate of the treatment effect.

Subsequently, Van Walbeek (2014) has also attempted to measure the size of South Africa's illicit market for cigarettes. He uses a method that compares predicted percentage changes in total consumption with actual changes in legal consumption. If predicted changes in total consumption are greater than actual changes in legal consumption, then the share of the illicit market is growing and vice versa.

[78] In Table C2, the figure for legal cigarette consumption per capita in 2004 is 521 cigarettes. A 5% illicit market share implies that the legal market was 95% of the total market. Similarly, an estimate of 12% of the illicit market implies that the legal market was 88% of the total market.

Between 1995 and 2004, Van Walbeek's estimates suggest that the share of the market that was due to illicit cigarettes remained virtually unchanged.[79] Unfortunately, his method does not allow for me to obtain a treatment effect that takes into account the illicit market. This is because he estimates percentage changes in the share as opposed to providing estimates of the actual share. However, given the consensus that the illicit market share was very low when the new tax policy started (Blecher, 2010, 2011), Van Walbeek's estimates suggest a small illicit market share over the period 1995 to 2004. This implies that my main treatment effect of 36% by 2004 is, therefore, not incredibly overestimated.

4.10. Summary and conclusion

South Africa has consistently increased the excise tax on cigarettes since 1994 largely on public health grounds. In increasing the tax, the government has sought to maintain a total tax burden of at least 52% of the average retail selling price (the target was initially set at 50%). This has resulted in substantial increases in the real price of cigarettes. For instance, between 1994 and 2004, the average real price per pack increased by 110%.

The main focus of this chapter was to evaluate the impact on consumption of this unprecedented increase in the price of cigarettes. I argued in the chapter that comparing current cigarette consumption to cigarette consumption before 1994 was likely to overstate the impact of the tax rises. This is because consumption had already started declining by the time the government's policy of raising taxes began.

The challenge in conducting impact evaluations is to create a credible counterfactual of what would have happened to cigarette consumption in the absence of the tax rises. This chapter, therefore, used the Synthetic Control method to create such a counterfactual for South Africa. The counterfactual was created as a linear combination of the per capita cigarette consumption of countries similar to South Africa that did not engage in large-scale tobacco control initiatives over the period 1994 to 2004. Using this counterfactual, I found that South Africa's cigarette consumption per capita would not have continued declining in the absence of the tax rises. Specifically, I found that cigarette consumption

[79] Van Walbeek's estimates of the change in the illicit market share range from an average decline of 2 percentage points to an average increase of 2 percentage points over the period 1995 to 2004.

would have stabilized at around 800 cigarettes per capita from the year 2000. Further, I found that by 2004, South Africa's per capita cigarette consumption was 36% lower than it would have been had the tax rises not happened.

South Africa's successful experience with tobacco control holds many lessons for countries, particularly those in Africa, that are trying to forestall an impending tobacco epidemic. South Africa's experience shows that significant public health dividends can be obtained by consistently increasing the real tax on cigarettes.

APPENDIX C

Table C 1: Treated and Untreated Countries

Treated		Untreated	
Australia	New Zealand	Argentina	Ireland
Austria	Nigeria	Bahrain	Jordan
Azerbaijan	Papua New Guinea	Bangladesh	Kuwait
Belgium	Paraguay	Brazil	Luxembourg
Cameroon	Poland	Chile	Malaysia
Canada	Portugal	China	Morocco
France	Russia	Colombia	Norway
Gabon	Saudi Arabia	Costa Rica	Pakistan
Germany	Singapore	Cote d'Ivoire	Panama
Guatemala	**South Africa**	Croatia	Peru
Hong Kong	Spain	Czech Rep	Philippines
Hungary	Sweden	Denmark	Romania
Iceland	Switzerland	Ecuador	Senegal
Israel	Thailand	Egypt	Serbia & Montenegro
Italy	Turkey	Finland	Sri Lanka
Japan	U.A.E	Greece	Tunisia
Kenya	United Kingdom	India	Uruguay
Korea, Rep.	United States	Indonesia	Vietnam
Mexico	Venezuela	Iran	
Netherlands	Zimbabwe		

Notes: Treated countries are those whose Relative Income Prices (RIPs) increased on average over the period 1990 to 2006 (i.e. where affordability declined). Untreated countries are those whose RIPs declined on average over the same period (i.e. where affordability increased). The information on RIPs is taken from Blecher and Van Walbeek (2009).

Table C 2: Actual estimates of treatment effects

Year	South Africa (Consumption Sticks p.c.)	Synthetic South Africa (Consumption Sticks p.c.)	Treatment Effect (Sticks p.c.)	Treatment Effect (%)
1990	1010	1008.86	1.14	0.11%
1991	993	991.93	1.07	0.11%
1992	947	945.58	1.42	0.15%
1993	901	899.94	1.06	0.12%
1994	883	865.95	17.05	1.97%
1995	849	887.26	-38.26	-4.31%
1996	796	880.52	-84.52	-9.60%
1997	737	864.11	-127.11	-14.71%
1998	692	824.46	-132.46	-16.07%
1999	634	817.92	-183.92	-22.49%
2000	577	785.12	-208.12	-26.51%
2001	570	783.16	-213.16	-27.22%
2002	597	801.77	-204.77	-25.54%
2003	495	814.90	-319.90	-39.26%
2004	521	810.86	-289.86	-35.75%

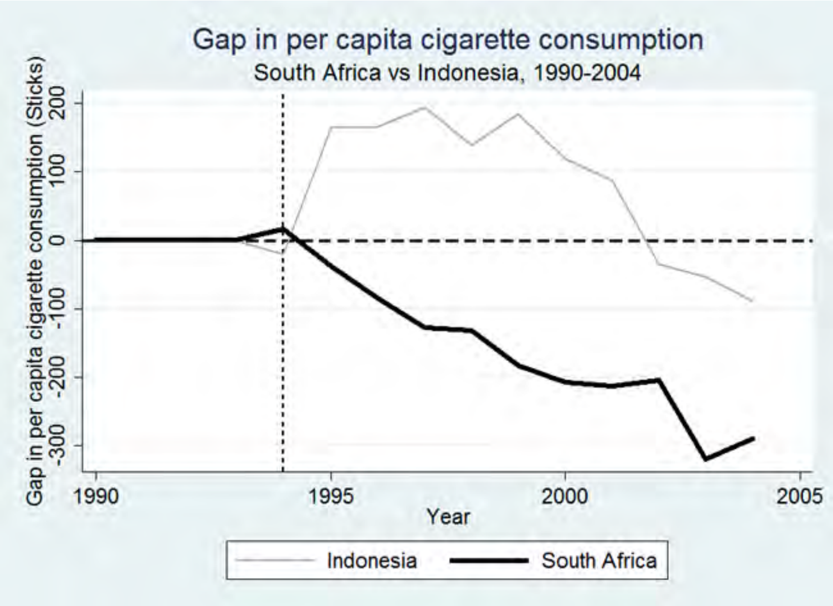
Notes: Treatment effects in the fourth column obtained by using equation (4.6). The last column presents treatment effects as a percentage difference. The consumption numbers for synthetic South Africa are obtained by applying the weights in Table 4.2 to the cigarette consumption numbers of the donor countries in Table 4.1. Cigarette consumption data is from ERC Group (2010).

Table C 3: Ranking of ratios of post-treatment RMSE to pre-treatment RMSE

Rank	Country	Post-treatment RMSE / pre-treatment RMSE
1	Indonesia	5695
2	South Africa	4341
3	Uruguay	3436
4	Senegal	112
5	Philippines	58
6	Chile	49
7	Malaysia	37
8	Cote d'Ivoire	28
9	Colombia	27
10	Sri Lanka	26
11	Vietnam	23
12	Egypt	23
13	Pakistan	17
14	Peru	16
15	Jordan	16
16	Panama	15
17	Argentina	13
18	Brazil	9
19	Romania	8
20	India	3
21	Tunisia	3
22	Morocco	2
23	Ecuador	2
24	Costa Rica	2
25	China	2

Notes: The Table shows a ranking of ratios of post-treatment RMSE to pre-treatment RMSE for all 25 countries in Table 4.1 plus South Africa. The data in this table is used to create Figure 4.8.

Figure C 1: Treatment effect, South Africa vs. Indonesia



Notes: Comparison of treatment effects between Indonesia and South Africa.

CHAPTER 5: CONCLUSION

This thesis set out to contribute to the literature on the economics of tobacco control with particular focus on sub-Saharan African countries. In so doing, I engaged with three broad but related themes in the literature: (i) the opportunity costs of tobacco expenditure within households, (2) the estimation of price elasticities of demand for cigarettes and (3) evaluating countries that have engaged in large-scale tobacco control initiatives. In what follows, I briefly summarise the main findings and contributions of my thesis with respect to each of the themes. Thereafter, I discuss some ideas for future research.

5.1. Main findings and contributions of the thesis

In Chapter 2 of the thesis, I set out to investigate whether tobacco expenditure displaced other goods and services within Zambian households. In so doing, the chapter made two contributions to the literature. Firstly, the chapter used expenditure data from a sub-Saharan African country where a majority of the households were poor and, therefore, likely to be income constrained. Aside from work by Koch and Tshiswaka-Kashalala (2008), who did similar work on South Africa, this is the only other study to investigate this type of crowding out within Africa.

The second contribution of Chapter 2 was more methodological. Previous work has used the method of instrumental variables to overcome the problem of reverse causality and/or simultaneity bias within demand systems. That is, the decision to spend on other goods and services is just as likely to be influenced by tobacco expenditure as tobacco expenditure is likely to be influenced by the expenditure on other goods and services. The instrumental variable of choice in the literature for tobacco expenditure has been the adult sex ratio. I argued in the chapter that the adult sex ratio was unlikely to satisfy the exclusion restriction. That is, the assumption of no correlation between the instrumental variable and the error term. This is because the adult sex ratio, being an aspect of household structure, is likely on its own to influence how households allocate expenditure between different goods and services. A plausible approach would be to use the adult sex ratio as an instrumental variable but allow for its correlation with the error term. This was the approach I adopted in the chapter.

Even after allowing for the correlation of the adult sex ratio with the error term, I still confirmed many of the findings in the literature. For instance, I found that food and schooling were displaced by tobacco expenditure in most households. Other expenditure categories that were displaced included clothing, expenditure on transportation and on the maintenance of equipment. Lastly, and unlike previous literature, I did not find that tobacco expenditure led households to spend more on alcohol. My conclusion was that the positive association between tobacco and alcohol found in previous research was likely a correlation rather than a causal relationship.

Traditionally, the costs of tobacco consumption have been thought of only in terms of mortality and morbidity. The findings from Chapter 2 show that a broader accounting of tobacco's toll in Zambia should include, for instance, the costs associated with under-nutrition and under-investment in education by households.

Even though we expect cigarettes to be price inelastic in most contexts, there is a demand for local evidence before initiating any meaningful policy changes. Outside of South Africa, there are currently few published estimates of the price elasticity of demand for tobacco products in Africa. Chapter 3, therefore, sought to provide such estimates for cigarettes in Uganda. The chapter made use of a method developed by Deaton (1988) that uses expenditure data and exploits the fact that prices of most goods and services tend to vary across geographical space in Low- and Middle-Income Countries (LMICs). The main finding from Chapter 3 was that cigarettes were price inelastic in Uganda with a price elasticity of demand in the -0.3 to -0.4 range. That is, each 10% rise in the price of cigarettes was expected decrease the quantity of cigarettes demanded by between 3% and 4%. These estimates are within the range found in the literature.

Given that cigarettes are price inelastic in Uganda, the government can reduce cigarette consumption and at the same time increase revenues by raising the excise tax on cigarettes. The important point is that the excise tax changes have to be real (i.e. inflation adjusted) for their impacts to be meaningful. Methodologically, the work in Chapter 3 shows that the plethora of expenditure surveys across Africa can be used to estimate elasticities of demand for many more countries.

South Africa is an example of an African country that has since 1994 substantially increased the real excise tax on cigarettes largely on public health grounds. Consequently, the real average price per pack of cigarettes increased by 190% between 1994 and 2012. We know from the literature reviewed in Chapter 3 that cigarette consumption is expected to decline whenever prices rise. We are, however, also interested in finding out what the trajectory of cigarette consumption would have been if the tax changes had not occurred. That is, we are interested in finding out the counterfactual. Counterfactuals are important in policy evaluations because they guard against the pitfall of wrongly attributing to a policy an outcome that would have happened anyway.

In Chapter 4, created such a counterfactual for South Africa using the Synthetic Control method first proposed by Abadie and Gardeazabal (2003). The counterfactual consumption trend line was estimated as a weighted average (or linear combination) of the trend lines of countries similar to South Africa that did not engage in large-scale tobacco control initiatives over the period 1990 to 2004. The counterfactual showed that per capita cigarette consumption would not have continued declining in the absence of the consistent tax rises that began in 1994. Specifically, the chapter found that consumption would have stabilized at around 800 cigarettes per capita from the year 2000. Lastly, I found that by 2004, South Africa's cigarette consumption per capita was 36% lower than that it would have been if the tax rises had not occurred.

South Africa's experience with tobacco control shows that significant public health dividends can be obtained by consistently increasing the real excise tax on cigarettes.

5.2. Ideas for future research

This section discusses some ideas for future research.

One of the main findings of this thesis is that tobacco expenditure displaces food consumption within Zambian households. An implication of this is that children in households with at least one smoker are likely to suffer from nutritional deficiencies when compared with children from non-smoking households. This hypothesis can readily be tested by comparing Z scores (or height-for-age measures) for children in the two types of households. Many African countries, including Zambia, regularly collect anthropometric data (including Z scores for children) as part of the Demographic and Health Surveys (DHS). Data from the DHS can readily be used to conduct such analyses.

Finally, the Synthetic Control method for creating counterfactuals can be used to conduct many types of policy evaluations in tobacco control. For instance, the method can go further in South Africa and provide an estimate of lung cancer cases averted since the government started raising excise taxes on cigarettes. The method can also be used to evaluate new approaches in tobacco control that have not yet been widely adopted. Australia's enactment of plain packaging legislation in 2012 is a case in point.

BIBLIOGRAPHY

Abadie, A. & Gardeazabal, J., 2003. "The economic costs of conflict: A case study of the Basque Country". *American Economic Review*, 93(1): 113 – 132

Abadie, A., Diamond, A. & Hainmueller, J., 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the effect of California's Tobacco Control Program". *Journal of the American Statistical Association*, 105(490): 493 – 505

Abadie, A., Diamond, A. & Hainmueller, J., 2015. "Comparative Politics and the Synthetic Control Method". *American Journal of Political Science*, 59(2): 495 – 510

Aloui, O., 2003. "Analysis of the economics of tobacco control in Morocco". HNP Discussion Paper Series, Economics of Tobacco Control Paper No. 16. Washington, DC: The World Bank

Andersen, S., Harrison, G.W., Morten, I.L. & Rutstrom, E.E., 2010. "Preference heterogeneity in experiments: comparing field and laboratory". *Journal of Economic Behaviour and Organization,* 73: 209 – 224

Angrist, J.D. & Pischke, J., 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press

Aragon, F.M. & Rud, J.P., 2015. "Polluting industries and agricultural productivity: Evidence from mining in Ghana". *Economic Journal*, Forthcoming

Averett, S.L. & Smith, J.K., 2014. "Financial hardship and obesity". *Economics and Human Biology*, 15: 201 – 212

Banks, J., Blundell, R. and Lewbel, A., 1997. "Quadratic Engel curves and consumer demand". *Review of Economics and Statistics*, 79(40): 527 – 539

Becker, G., 1974. "A theory of social interactions". *Journal of Political Economy*, 82(6): 1063 – 1094

Becker, G., 1981. *A treatise on the family.* Cambridge: Harvard University Press

Bilano, V., Gilmour, S., Moffiet, T., d'Espaignet, E.T., Stevens, G.A., Commar, A., Tuyl, F., Hudson, I. & Shibuya, K., 2015. "Global trends and projections for tobacco use, 1990 – 2025: an analysis of smoking indicators from the WHO Comprehensive Information Systems for Tobacco Control". *The Lancet*, 385(9972): 966 – 976

Billmeier, A. & Nannicini, T., 2013. "Assessing Economic Liberalization episodes: A Synthetic Control Approach". *Review of Economics and Statistics*, 95(3): 983 – 1001

Bishop, J.A., Liu, H. & Meng, Q., 2007. "Are Chinese smokers sensitive to price?". *China Economic Review*, 18(2): 113 – 121

Blecher, E.H. & Van Walbeek, C.P., 2004. "An international analysis of cigarette affordability". *Tobacco Control,* 13(4): 339 – 346.

Blecher, E.H., 2006. "The effects of the Tobacco Products Control Amendment Act of 1999 on restaurant revenues in South Africa: A panel data approach". *South African Journal of Economics*, 74(1): 123 – 130

Blecher, E., 2008. "The impact of tobacco advertising bans on consumption in developing countries". *Journal of Health Economics*, 27(4): 930 – 942

Blecher, E.H. & Van Walbeek, C.P., 2009. "Cigarette affordability: An update and some methodological comments". *Tobacco Control*, 18: 167 – 175

Blecher, E., 2010. "A mountain or a molehill: is the illicit trade in cigarettes undermining tobacco control policy in South Africa?". *Trends in Organized Crime*, 13(4):299 – 315

Blecher, E.H., 2011. *The economics of tobacco control in Low- and Middle-Income Countries*. PhD Thesis. University of Cape Town

Blecher, E.H. and Ross, H., 2013. *Tobacco use in Africa: Tobacco Control through Prevention*. American Cancer Society

Block, S. & Webb, P., 2009. "Up in smoke: Tobacco use, expenditure on food, and child malnutrition in Developing Countries". *Economic Development and Cultural Change*, 58(1): 1 – 23

Bosch, A. & Koch, S.F., 2014. "Using a natural experiment to examine tobacco tax regressivity". ERSA Working Paper 434

Boshoff, W.H., 2008. "Cigarette demand in South Africa over 1996 to 2006: the role of price, income and health awareness". *South African Journal of Economics*, 76: 1 – 16

Busch, S.H., Jofre-Bonet, M, Falba, T.A. & Sindelar, J.L., 2004. "Burning a hole in the budget: Tobacco spending and its crowd-out of other goods". *Applied Health Economics and Health Policy*, 3(4): 263 – 272

Cawley, J., Markowitz, S. & Tauras, J., 2004. "Lighting up and slimming down: the effects of body weight and cigarette prices on adolescent smoking initiation". *Journal of Health Economics*, 23(2): 293 – 311

Chaloupka, F., 1991. "Rational addictive behaviour and cigarette smoking". *Journal of Political Economy*, 99(4): 722 – 742

Chaloupka, F. & Warner K.E., 2000. "The economics of smoking". In J. Newhouse & D. Cutler (Eds.), *Handbook of Health Economics* (pp. 1539 – 1567). Amsterdam: North-Holland

Chabris, C., Laibson, D., Morris, C., Schuldt, J. & Taubinsky, D., 2008. "Individual laboratory-measured discount rates predict field behaviour". *Journal of Risk and Uncertainty*, 37(2): 237 – 269

Chelwa, G., 2012. "The Tobacco Story in Zambia: A demand, supply and tax analysis". Unpublished manuscript. University of Cape Town

Chelwa, G. & Van Walbeek, C., 2014. "Assessing the causal impact of tobacco expenditure on household spending patterns in Zambia". ERSA Working Paper 453

Chen, Y. & Xing, W., 2011. "Quantity, quality, and regional price variation of cigarettes: Demand analysis based on a household survey in China". *China Economic Review*, 22(2): 221 – 232

Cohn, A., Brandon, T., Armeli, S., Ehlke, S. & Bowers, M., 2015. "Real-time patterns of smoking and alcohol use: an observational study protocol of risky-drinking smokers". *Addiction.*

Craig, J., 2000. "Evaluating privatization in Zambia: a tale of two processes". *Review of African Political Economy*, 27(85): 357 – 366

Cunningham, S. & Shah, M., 2014. "Decriminalizing indoor prostitution: Implications for sexual violence and public health". NBER Working Paper No. 20281

Deaton, A. & Muellbauer, J., 1980a. "An Almost Ideal Demand System". *American Economic Review*, 70(3): 312 – 326

Deaton, A. and Muellbauer, J., 1980b. *Economics and Consumer Behaviour*. New York: Cambridge University Press

Deaton, A., 1988. "Quality, Quantity, and Spatial Variation of Price". *American Economic Review*, 78(3):418 – 430

Deaton, A., 1989. "Household Survey Data and Pricing Policies in Developing Countries". *World Bank Economic Review*, 3(2):183 – 210

Deaton, A., Ruiz-Castillo, J. & Thomas, D., 1989. "The influence of household composition on household expenditure patterns: Theory and Spanish evidence". *Journal of Political Economy*, 97(1): 179 – 200

Deaton, A, 1990. “Price elasticities from survey data: extensions and Indonesian results”. *Journal of Econometrics*, 44(3):281-309

Deaton, A. & Grimard, F., 1992. “Demand analysis for tax reform in Pakistan*”. LSMS Working Paper* 85. Washington: The World Bank

Deaton, A., 1997. *Analysis of household surveys*. Baltimore: Johns Hopkins University Press

Dierker, L., Lloyd-Richardson, E., Stolar, M., *et al.*, 2006. “The proximal association between smoking and alcohol use among first year college students”. *Drug and Alcohol Dependence*, 81: 1 – 9

Doss, C., 2013. “Intrahousehold bargaining and resource allocation in developing countries”. *World Bank Research Observer*, 28(1): 52 – 78

Dube, A. & Zipperer, B., 2015. “Pooling multiple case studies using Synthetic Controls: An application to Minimum Wage policies”. IZA Discussion Paper Number 8944

Duflo, E., 2003. “Grandmothers and granddaughters: Old age pensions and intrahousehold allocation in South Africa”. *World Bank Economic Review*, 17: 1 – 25

Economics of Tobacco Control in South Africa Project, 1998. “The economics of tobacco control in South Africa.” Report submitted to the International Tobacco Initiative. Cape Town, Applied Fiscal Research Center, University of Cape Town

Efroymson, D., Ahmed, S., Townsend, J., Alam, S.M., Dey, A.R., Saha, R., Dhar, B., Sujon, A.I., Ahmed K.U. & Rahman, O., 2001. “Hungry for tobacco: an analysis of the economic impact of tobacco consumption on the poor in Bangladesh”. *Tobacco Control*, 10: 212 – 217

Eozenou, P. & Fishburn, B., 2009. “Price elasticity estimates for cigarette demand in Vietnam”. DEPOCEN Working Paper Series No. 2009/05

ERC Group, 2010. *World Cigarettes*. Newmarket: ERC Group

Eriksen, M., Mackay, J., Schluger, N., Gomeshtapeh, F.I. and Drope, J., 2015. *The Tobacco Atlas*. 5th edition. American Cancer Society

Fersterer, J. & Winter-Ebmer, R., 2003. “Smoking, discount rates, and returns to education”. *Economics of Education Review,* 22(6): 561 – 566

Fichtenberg, C. & Glantz, S., 2000. “Association of the California Tobacco Control Program with declines in cigarette consumption and mortality from heart disease”. *New England Journal of Medicine*, 343(24): 1772 – 1777

Forster, M. & Jones, A.M., 2001. “The role of tobacco taxes in starting and quitting smoking: duration analysis of British data”. *Journal of the Royal Statistical Society*, 164(3): 517 – 547

Franz, G.A., 2008. “Price effects on the smoking behaviour of adult age groups”. *Public Health*, 1343 – 1348.

Gibson, J. & Rozelle, S., 2004. “An empirical test of methods for estimating price elasticities from household survey data”. University of Waikato working paper.

Gibson, J. & Rozelle, S., 2005. “Prices and unit values in poverty measurement and tax reform analysis”. *World Bank Economic Review*, 19(1): 69 – 97

Gitter, S. & Barham, B., 2008. “Women’s power, conditional cash transfers, and schooling in Nicaragua”. *World Bank Economic Review*, 22: 271 – 290

Green, C.P., Heywood, J.S. & Navarro, M., 2014. “Did liberalising bar hours decrease traffic accidents?”. *Journal of Health Economics*, 35: 189 – 198

Gruber, J. & Koszegi, B., 2001. “Is addiction “rational”? Theory and evidence”. *Quarterly Journal of Economics*, 116(4): 1261 – 1303

Gruber, J. & Koszegi, B., 2004. “Tax incidence when individuals are time-inconsistent: the case of cigarette excise taxes”. *Journal of Public Economics*, 88(9 – 10): 1959 – 1987

Gruber, J.H. & Mullainathan S., 2005. “Do cigarette taxes make smokers happier?”. *B.E. Journal of Economic Analysis & Policy*, 5(1): 1935 – 1682

Guindon, E.G., Nandi, A., Chaloupka, F. & Jha, P., 2011. “Socioeconomic differences in the impact of smoking tobacco and alcohol prices on smoking in India”. NBER Working Paper Number 17580

Guindon, E.G., Paraje, G.R. & Chaloupka, F.J., 2015. “The impact of prices and taxes on the use of tobacco products in Latin America and the Caribbean”. *American Journal of Public Health*, 105(3): e9 – e19

Hanafy, K., Saleh, A.S.E., Elmallah M.E.B.E. *et al.*, 2011. “The economics of tobacco taxation in Egypt”. Paris: International Union Against Tuberculosis and Lung Disease

Harrison, G.W., Lau, M.L. & Williams, M.B., 2002. “Estimating individual discount rates in Denmark: A field experiment”. *American Economic Review,* 92(5): 1606 – 1617

Harrison, G.W., Morten, I.L., Rutstrom, E.E. & Sullivan, M.B., 2005. “Eliciting risk and time preferences using field experiments: some methodological issues”. *Field Experiments in Economics,* 10: 125 – 218

Harrison, G.W., Lau, M.I. & Rutstrom, E.E., 2010. "Individual discount rates and smoking: Evidence from a field experiment in Denmark". *Journal of Health Economics,* 29(5): 708 – 717

Heckman, J.J., 1978. "Dummy endogenous variables in a simultaneous equation system". *Econometrica,* 46(6): 931 – 959

Hill, C., 2013. "Cancer prevention and screening". *Bulletin of Cancer*, 100: 547 – 554

Holford, T.R., Meza, R., Warner, K.E., Meernik, C., Jeon, J., Moolgavkar & Levy, D.T., 2014. "Tobacco control and the reduction in smoking-related premature deaths in the United States, 1964-2012". *Journal of the American Medical Association*, 311(2): 164 – 171

Hoddinott, J. & Haddad, L., 1995. "Does female income share influence household expenditures? Evidence from Cote d'Ivoire". *Oxford Bulletin of Economics and Statistics*, 57: 77 – 96

Hu T.W., Bai J., Keeler T.E., Barnett P.G. & Sung H.Y., 1994. "The impact of California's Proposition 99, a major anti-cigarette law, on cigarette consumption". *Journal of Public Health Policy*, 15(1): 26 – 36

Hu T.W., Sung H.Y. & Keeler T.E., 1995. "Reducing cigarette consumption in California: tobacco taxes vs an anti-smoking media campaign". *American Journal of Public Health*, 85(9): 1218 – 1222

Huang, J., Zheng, R., Chaloupka, F.J., Fong, G.T. & Jiang, Y., 2015. "Differential responsiveness to cigarette price by education and income among adult urban Chinese smokers". *Tobacco Control*, 1 – 7

Imbens, G.W. & Wooldridge, J.M., 2009. "Recent developments in the econometrics of program evaluation". *Journal of Economic Literature*, 47(1): 5 – 86

International Agency for Research on Cancer (IARC), 2008. *Handbook on methods for evaluating tobacco control policies*. Geneva: World Health Organization

International Agency for Research on Cancer (IARC), 2011. *Handbook on the Effectiveness of Tax and Price Policies for Tobacco Control.* Geneva: World Health Organization

Jha, P. & Chaloupka, F.J. (Eds), 2000. *Tobacco control in developing countries*. New York: Oxford University Press

Jha, P., Musgrove, P., Chaloupka, F. & Yurekli A., 2000. "The economic rationale for intervention in the tobacco market". In P. Jha & F. Chaloupka (Eds.), *Tobacco control in developing countries* (pp. 153 – 174). New York: Oxford University Press

Jha, P. & Peto, R., 2014. “Global effects of smoking, of quitting, and of taxing tobacco”. *New England Journal of Medicine*, 370: 60 – 68

John, R.M., 2005. “Price elasticity estimates for tobacco and other addictive goods in India”. Indira Gandhi Institute of Development Research, Working Paper Number WP-2005-003

John, R.M., 2008a. “Price elasticity estimates for tobacco products in India”. *Health Policy and Planning*, 23: 200 – 209

John, R.M, 2008b. “Crowding out effect of tobacco expenditure and its implications on household resource allocation in India”. *Social Science and Medicine*, 66: 1356 – 1367

John, R.M., Ross, H. & Blecher, E., 2012. “Tobacco expenditure and its implications for household resource allocation in Cambodia”. *Tobacco Control*, 21(3): 341 – 346

Kang, H.Y., Kim, H.K., Park, T.K., Jee, S.H., Nam, C.M., & Park, H.W., 2003. “Economic burden of smoking in Korea”. *Tobacco Control*, 12: 37 – 44

Keeler T.E, Hu T.W, Barnett P.G. & Manning W.G., 1993. “Taxation, regulation and addiction: a demand function of cigarettes based on time-series evidence”. *Journal of Health Economics*, 12(1): 1 – 18

King, G. & Zeng, L., 2006. “The dangers of extreme counterfactuals”. *Political Analysis*, 14: 131 – 159

Koch, S.F. & Tshiswaka-Kashalala, G., 2008. “Tobacco substitution and the poor”. University of Pretoria Working Paper Series no. 2008 – 32

Kortelainen, M. & Saarimaa, T., 2015. “Do urban neighbourhoods benefit from homeowners? Evidence from housing prices”. *Scandinavian Journal of Economics*, 117(1): 28 – 56

Lance, P.M., Akin, J.S., Dow, W.H. & Loh, C.P., 2004. “Is cigarette smoking in poorer nations highly sensitive to price? Evidence from Russia and China”. *Journal of Health Economics*, 23(1): 173 – 189

Levine, P.B., Gustafson, T.A. & Velenchik, A.D., 1997. “More bad news for smokers? The effects of cigarette smoking on wages”. *Industrial Relations & Labour Review,* 50(3): 493 – 509

Levy, D.T., Benjakul, S., Ross, H. & Ritthiphakdee, B., 2008. “The role of tobacco control policies in reducing smoking and deaths in a middle income nation: results from the Thailand SimSmoke simulation model”. *Tobacco Control*, 17: 53 – 59

Liu, Y., Rao, K., Hu T.W., Sun, Q & Mao, Z., 2006. “Cigarette smoking and poverty in China”. *Social Science and Medicine,* 63: 2784 – 2790

Lucas, R.E., 1976. "Econometric policy evaluation: A critique". In K. Brunner & A.H. Meltzer (Eds.), *The Phillips Curve and Labour Markets, Carnegie-Rochester Conferences on Public Policy*. Amsterdam: North Holland

Lundberg, S., Pollak, R. & Wales, T., 1997. "Do husbands and wives pool their resources? Evidence from the United Kingdom Child Benefit". *Journal of Human Resources*, 32: 463 – 480

Lye, J.N. & Hirschberg, J., 2004. "Alcohol consumption, smoking and wages". *Applied Economics*, 36(16): 1807 – 1817

Malan, M. & Leaver, R., 2003. "Political change in South Africa: New tobacco control and public health policies". In De Beyer & Waverly Brigden (Eds.), *Tobacco control policy: Strategies, successes and setbacks* (pp. 121 – 151). Washington and Ottawa: World Bank and Research for International Tobacco Control

Mathers, C.D. & Loncar, D., 2006. "Projections of global mortality and burden of disease from 2002 to 2030". *PLoS Medicine*, 3(11): e442

Max, W., Rice, D.P., Sung, H.Y., Zhang, X. & Miller, L., 2004. "The economic burden of smoking in California". *Tobacco Control,* 13: 264 – 267

Musoke, F., 2008. *The Uganda Global Youth Tobacco Survey Report: Tobacco Control Policy Implications*. Geneva: World Health Organization

Muuka, G.N., 1997. "Wrong-footing MNCs and local manufacturing: Zambia's 1992 – 1994 structural adjustment program". *International Business Review*, 6(6): 667 – 687

Nargis, N., Rutbah, U.H. and Fong, G., 2010. "Taxation of tobacco products in Bangladesh: Findings from the 2009 ITC Bangladesh survey." ITC Project Working Paper Series. University of Waterloo

Nargis, N., Rutbah, U.H., Hussain, A.K.M., Ashiquzzaman, S.M., Fong, G.T. and Huq, I., 2011. "Pricing and taxation of tobacco products in Bangladesh: Findings from Wave 1 and Wave 2 of the ITC Bangladesh survey". ITC Working Paper Series. University of Waterloo

Nevo, A. & Rosen, A.M., 2012. "Identification with imperfect instruments". *Review of Economics and Statistics*, 94(3): 659 – 671

Nicolas, A.L., 2002. "How important are tobacco prices in the propensity to start and quit smoking? An analysis of smoking histories from the Spanish National Health Survey". *Health Economics*, 11: 521 – 535

Oberg, M., Jaakola, M., Woodward, A., Peruga, A. & Pruss-Ustun, A., 2010. "Worldwide burden of disease from exposure to second-hand smoke: a retrospective analysis of data from 192 countries". *The Lancet*, 377(9760): 139 – 146

Pampel, F. C., 2008. "Tobacco use in sub-Saharan Africa: Estimates from the Demographic Health Surveys". *Social Science and Medicine*, 66: 1772 – 1783

Peterson, D.E., Zeger, S.L., Remington, P.L. & Anderson, H.A., 1992. "The effect of state cigarette tax increases on cigarette sales, 1955 to 1988". *American Journal of Public Health*, 82(1): 94 – 96

Pollak, R.A., 1969. "Conditional demand functions and consumption theory". *Quarterly Journal of Economics*, 83(1): 60 – 68

Pu, C., Lan, V, Chou, Y-J. & Lan, C., 2008. "The crowding-out effects of tobacco and alcohol where expenditure shares are low: Analysing expenditure data for Taiwan". *Social Science and Medicine*, 9: 1979-1989

Republic of South Africa, 1994. *Budget review*. Available at http://www.treasury.gov.za/documents/national%20budget/Budget%20Review%201994.pdf (Accessed November, 2014)

Republic of South Africa, 1995. *Budget review.* Available at http://www.treasury.gov.za/documents/national%20budget/Budget%20Review%201995.pdf (Accessed November 2014)

Republic of South Africa, 1996. *Budget review*. Available at http://www.treasury.gov.za/documents/national%20budget/Budget%20Review%201996.pdf (Accessed November 2014)

Republic of South Africa, 1997. *Budget speech by the Minister of Finance.* Available at: http://www.treasury.gov.za/documents/national%20budget/1997/speech/speech.pdf (Accessed November 2014)

Republic of South Africa, 2004. *Budget speech by the Minister of Finance.* Available at: http://www.treasury.gov.za/documents/national%20budget/2004/speech/Speech.pdf (Accessed November 2014)

Reekie, W.D., 1994. "Consumers' surplus and the demand for cigarettes." *Managerial and Decision Economics*, 15: 223 – 234

Reinhold, S. & Woutersen, T., 2011. “Endogeneity and imperfect instruments: Estimating bounds for the effect of early childbearing on high school completion”. Mimeo

Sahn, D.E & Stifel, D., 2003. “Exploring alternative measures of welfare in the absence of expenditure data”. *Review of Income and Wealth*, 49(4): 463 – 489

Saloojee, Y., 1994. “Tobacco control: A case study”. In D. Yach and Harrison (Eds.), *Proceedings of the All Africa Conference on Tobacco or Health*. Harare

Samuelson, P.A., 1956. “Social indifference curves”. *Quarterly Journal of Economics*, 70(1): 1 – 22

Sangthong, R., Wichaidit, W. & Ketchoo, C., 2012. “Current situation and future challenges of tobacco control policy in Thailand”. *Tobacco Control*, 21: 49 – 54

Senauer, B., Garcia, M. & Jacinto, E., 1988. “Determinants of the intrahousehold allocation of food in the rural Philippines”. *American Journal of Agricultural Economics*, 70: 170 – 180

South African Medical Research Council (SAMRC), 1988. *Smoking and health in South Africa: The need for action.* Pretoria: SAMRC

South African Medical Research Council (SAMRC), 1992. *Smoking in South Africa: Health and economic impact*. Pretoria: SAMRC

Stock, J.H., Wright, J.H. & Yogo, M., 2002. “A survey of weak instruments and weak identification in Generalized Method of Moments”. *Journal of Business and Economics Statistics*, 20(4): 518 – 529

Sung, H.Y, Hu T.W. & Keeler T.E., 1994. “Cigarette taxation and demand: an empirical model”. *Contemporary Economic Policy*, 12(3): 91 – 100

Takada, H., Ullah, A. & Chen, Y-M., 1995. “Estimation of the Seemingly Unrelated Regression model when the error covariance matrix is singular”. *Journal of Applied Statistics,* 22(4): 517 – 530

Tauras, J.A., O’Malley, P.M. & Johnston, L.D., 2001. “Effects of price and access laws on teenage smoking initiation: A national longitudinal analysis”. NBER Working Paper Number 8331

Thomas, D., 1990. “Intra-household resource allocation: An inferential approach”. *Journal of Human Resources*, 25: 635 – 664

Thomas, D., 1994. “Like father, like mother, like daughter: Parental resources and child height”. *Journal of Human Resources*, 29(4): 950 – 988

United Nations, 2012. *World Population Prospects: The 2012 Revision, Volume I*. United Nations, Department of Economics and Social Affairs, Population Division

United States Department of Health and Human Services, 2014. *The health consequences of smoking – 50 years of progress: A report of the Surgeon General*. Atlanta: Centers for Disease Control and Prevention.

Van Walbeek, C.P., 1996. "Excise taxes on tobacco: how much scope does the government have?". *South African Journal of Economics*, 64: 21 – 42

Van Walbeek, C., 2002. "Recent trends in smoking prevalence in South Africa: Some evidence from AMPS data". *South African Medical Journal*, 92(6): 468 – 472

Van Walbeek, C., 2005. *The economics of tobacco control in South Africa*. PhD Thesis. University of Cape Town

Van Walbeek, C., 2006. "Industry responses to the tobacco excise tax increases in South Africa". *South African Journal of Economics*, 74(1): 110 – 122

Van Walbeek, C., Blecher, E. & Van Graan, M., 2007. "Effects of the Tobacco Products Control Amendment Act of 1999 on restaurant revenues in South Africa – a survey approach". *South African Medical Journal*, 97(3): 208 – 211

Van Walbeek, C., 2014. "Measuring changes in the illicit cigarette market using government revenue data: the example of South Africa". *Tobacco Control*, 23: e69 – e70

Varian, H., 1992. *Microeconomic Analysis*. New York: W.W. Norton & Company

Vermeulen, F., 2003. "Do smokers behave differently? A tale of zero expenditures and separability concepts". *Economics Bulletin,* 4(6): 1 – 7

Viscusi, W.K. & Hersch, J., 2001. "Cigarette smokers as job risk takers". *Review of Economics and Statistics*, 83(2): 269 – 280

Wang, H., Sindelar, J.L. & Busch, S.H., 2006. "The impact of tobacco expenditure on household consumption patterns in rural China". *Social Science and Medicine,* 62(6): 1414 – 1426

Warner, K.E., 1977. "The effects of the anti-smoking campaign on cigarette consumption". *American Journal of Public Health*, 67(7): 645 – 650

Warner, K.E., 1990. "Tobacco taxation as health policy in the Third World". *American Journal of Public Health*, 80(5): 529 – 531

Wooldridge, J.M., 2002. *Econometric Analysis of Cross Section and Panel Data.* Cambridge, Massachusetts: MIT Press

World Bank, 2015. *World Development Indicators*. Available at data.worldbank.org

World Health Organization (WHO), 2005. *Framework Convention on Tobacco Control*. Geneva: World Health Organization

World Health Organization (WHO), 2008. *Report on the global tobacco epidemic: the MPOWER package*. Geneva: World Health Organization

World Health Organization (WHO), 2010. *Technical manual on tobacco tax administration*. Geneva: World Health Organization

World Health Organization (WHO), 2012a. *Global report on mortality attributable to tobacco*. Geneva: World Health Organization

World Health Organization (WHO), 2012b. *Joint capacity assessment on the implementation of effective tobacco control policies in Uganda*. Geneva: World Health Organization

World Health Organization (WHO), 2013a. *Report on the global tobacco epidemic*. Geneva: World Health Organization

World Health Organization (WHO), 2013b. *Report on the Global Tobacco Epidemic: Uganda Country Profile*. Geneva: World Health Organization

World Health Organization (WHO), 2014a. *Global progress report on the implementation of the WHO Framework Convention on Tobacco Control*. Geneva: World Health Organization

World Health Organization (WHO), 2014b. *Global status report on noncommunicable diseases*. Geneva: World Health Organization

World Health Organisation (WHO), 2014c. "Tobacco key facts". Available at http://www.who.int/mediacentre/factsheets/fs339/en/

World Health Organization (WHO), 2015. *Global report on trends in prevalence of tobacco smoking 2015*. Geneva: World Health Organization

Yach, D., 1982. "Economic aspects of smoking in South Africa". *South African Medical Journal*, 62: 167 – 170

Young, A., 2012. "The African growth miracle". *Journal of Political Economy*, 120(4): 696 – 739

Zambia Demographic and Health Survey, 2002. *Country Report*. Available at http://www.measuredhs.com/ (accessed July 2012)

Zambia Demographic and Health survey, 2007. *Country Report*. Available at http://www.measuredhs.com/ (accessed July 2012)

Zellner, A., 1962. "An efficient method for estimating Seemingly Unrelated Regressions and tests for Aggregation Bias". *Journal of the American Statistical Association,* 57(298): 348 – 368