ECONOMIC AND HEALTH IMPLICATIONS OF TOBACCO AND ALCOHOL USE IN SOUTH AFRICA: A HOUSEHOLD EXPENDITURE ANALYSIS

Zunda Chisha

Thesis Presented for the Degree of Doctor of Philosophy
in the School of Economics
University of Cape Town

January 2024

Supervisor: Professor Corné van Walbeek

Co-supervisor: Associate Professor Grieve Chelwa

[Word count: 70,423]

University of Cape Town

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

University of Cape Town

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

DECLARATION

I, Zunda Chisha, hereby declare that the work on which this thesis is based is my own original work (except where acknowledgements indicate otherwise), and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university. I authorise the University to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature: Signed by candidate Date: 31 January 2024

ABSTRACT

This study uses nationally representative data to analyse the economic and health implications of tobacco and alcohol use in South Africa. Issues covered include the crowding-out effect of tobacco expenditure on household expenses, and the impact of parental tobacco and alcohol consumption on child health outcomes. The thesis emphasises the interconnectedness of tobacco and alcohol use, and their place in South Africa's development and policy landscape.

Chapter 2 uses four income and expenditure surveys to investigate the categories of household expenditure displaced by tobacco expenditure. The analysis begins with a simple comparison of differences in expenditure shares between households that spent money on tobacco and those that did not. Next, the displacement of household expenditures is estimated using instrumental variables within a system-of-equations. Confidence intervals for the point estimates are then estimated using a method proposed by Nevo & Rosen (2012). Results indicate that expenditures on education, transport, housing, and household goods were displaced by tobacco expenditure, however, alcohol expenditure was sometimes crowded-in. Decreased household spending on education could lead to reduced human capital development.

Chapter 3 looks at how crowding-out patterns changed as tobacco taxes increased in the period 1995-2005/6. It uses a semi-parametric approach for the difference-in-difference technique developed by Abadie (2005), accounting for violations of the parallel-trends assumption. The results suggest a non-uniform pattern of adjustment of expenditures in tobacco-consuming households. While budget shares increased for some categories, such as food and alcohol, others, such as education, were crowded out. The displacement of expenditure on education is proportionately greater for low income households than for those with higher incomes.

Chapter 4 looks at whether parental tobacco and alcohol impact child health outcomes by crowding-out household food expenditure. Using instrumental variable estimation and mediation analysis, the study uses the National Income Dynamics Study (NIDS) panel datasets to investigate the lagged long-term effects of tobacco and alcohol expenditures on two child anthropometric measures (height-for-age and weight-for-age z-scores). Although a direct link between reduced food expenditure and observed effect is not established, tobacco and alcohol expenditure are shown to have a substantial negative impact on child anthropometric measures.

The study has three main implications. First, the dynamic analysis shows how various factors interplay with changes in fiscal policy, affecting household-level budgeting decisions. In doing

so it helps explain the effectiveness of cigarette tax policies over time, highlighting their continued relevance. Secondly, the study reveals possible unintended consequences among households that continue purchasing cigarettes after substantial price increases, such as deeper expenditure displacements, which policymakers must consider. Lastly, the evidence generated here suggests that reductions in tobacco use can be associated with a reduction in alcohol use, thus tobacco control could be used a tool for alcohol control.

DEDICATION

To my parents for giving me the encouragement to pursue my goals relentlessly in order to achieve them.

To my kids, Taliah, Caleb and Ella, all three of whom were born during this journey.

To my wife, Abby, for being the inspiration behind the scenes and the rock behind the success.

To my Father, King of the Universe, for all that I am is because of Him.

ACKNOWLEDGEMENTS

This thesis owes its existence to the unwavering patience and invaluable guidance of my supervisor, Prof. Corne van Walbeek. I embarked on this journey with uncertain footing, but thanks to Corne's steadfast support, I managed to reach my goal. Corne's resolute hand and enduring patience throughout my odyssey have been nothing short of extraordinary. I am profoundly indebted to him. Corne, I extend my heartfelt gratitude to you.

When it comes to technical matters within this thesis, Dr. Grieve Chelwa provided essential support and direction. I am sincerely grateful for his invaluable insights on the various topics explored before the final versions of these chapters took shape.

I count myself fortunate to have conducted my studies within the company of an exceptional team of researchers at REEP. The camaraderie and solidarity within this team are truly remarkable, and I am honoured to be part of it. I want to give special recognition to the post-doctoral colleagues who endured my endless questions and discussions over the years: Dr. Alfred Kechia Mukong, Dr. Ernest Tingum, Dr. Abel Nyagwachi, Dr. Michael Boachie, Dr. Zachary Gitonga, Dr. Joshua Magero and Dr. Chengetai Dare.

A heartfelt mention goes to Sharon, whose efforts ensured that my scholarship remained intact and lightened my administrative burdens in all aspects of this endeavour.

Within the PhD lab, Dr. Mashekwa Maboshe provided the encouragement that kept me going. I am grateful to the numerous PhD students who engaged in discussions with me over the years—there are too many to name individually, but each of them deserves my appreciation for the time they dedicated to enduring my musings.

I also want to express my gratitude to the ERSA PhD conference 2023 cohort for their insightful comments on my second chapter.

I am certain that many others lent their ears to my discussions on these chapters and contributed in various ways to shape this work. While I may not mention them by name here, I want to extend my thanks to all of them for taking the time to help me navigate the challenges that arose during this process.

To Elizabeth Baldwin for coming to my rescue and providing an excellent edit of this thesis.

Finally, my wife served as the inspiration and visionary behind our move to Cape Town many years ago, setting me on this trajectory. She has been a steadfast companion on this journey, and I salute her for standing by my side throughout it all.

TABLE OF CONTENTS

1. CHAPTER 1: INTRODUCTION 1

1.1 Background and motivation 1

1.1.1 Global trends in tobacco and alcohol use 1

1.1.2 Tobacco use in South Africa 5

1.1.3 Alcohol use in South Africa 7

1.1.4 The policy response to tobacco and alcohol use 8

1.2 The importance of understanding household expenditure patterns 12

1.3 Research objectives 14

1.4 Research Methods 14

1.5 Contribution 15

1.6 Themes and outline of the thesis 16

2. CHAPTER 2: THE OPPORTUNITY COST OF TOBACCO USE IN SOUTH AFRICA 18

2.1 Introduction 18

2.2. Relevant literature 22

2.3. Theoretical framework 25

2.4. Empirical strategy 26

2.5. Data and data quality issues 31

2.5.1. Changes in the measurement instrument 32

2.5.2. Changes in sampling procedures 33

2.5.3. Outliers and influential data points 35

2.6. Descriptive data 38

2.7. Empirical results 43

2.8 Summary and conclusion 60

2.9 Annex A 63

3. CHAPTER 3: TOBACCO TAX INCREASES AND THE EVOLUTION OF TOBACCO-INDUCED CROWDING-OUT 83
3.1 Introduction 83
3.2 Relevant literature 85
3.3 Difference-in-difference: Conceptual framework 89
3.4 Empirical strategy 92
3.5 Data 98
3.6 Empirical results 110
3.7 Discussion 122
3.8 Summary and conclusion 128
3.9 Annex B 130
4. CHAPTER 4: THE EFFECT OF TOBACCO AND ALCOHOL EXPENDITURE ON CHILD HEALTH OUTCOMES 137
4.1 Introduction 137
4.2 Literature review 139
4.3 A conceptual framework for child health and nutrition 146
4.4 Empirical strategy 148
4.5 Data 156
4.6 Descriptive statistics 157
4.7 Econometric results 159
4.8 Discussion 172
4.9 Conclusion 175
4.10 Annex C 176
5. CHAPTER 5: CONCLUSION 192
5.1 Main findings and contributions of this thesis 192
5.2 Main limitations of the thesis 194
5.3 Ideas for future research 195
6. REFERENCES 198

LIST OF FIGURES

Figure 1.1: Trends in current tobacco use among people aged 15 years and above.....2
Figure 1.2: Percentage of current drinkers, former drinkers, and lifetime abstainers (in %) among the total population (15+ years).....3
Figure 1.3: Trends in total alcohol per capita consumption (APC) among drinkers (15+ years) in litres of pure alcohol in WHO regions, 2000−2016.....4
Figure 1.4: Current Tobacco Smoking Prevalence by Age Group and Gender, 2021.....6
Figure 2.1: Tobacco excise taxes, cigarette prices and consumption in South Africa between 1995 and 2010.....21
Figure 2.2: Percent of single-person households across surveys.....34
Figure 2.3: Proportion of multi-person and single-person households among smokers and non-smokers.....35
Figure 2.4: Differences in expenditure shares between smoking and non-smoking households by settlement type.....48
Figure 2.5: Differences in expenditure shares between smoking and non-smoking households by tercile.....49
Figure 3.1: Difference in difference conceptual model.....91
Figure 3.2: Trends in mean budget shares disaggregated by tobacco use.....101
Figure 3.3: Trends in the mean budget share difference from 1995 to 2005/06.....113
Figure 3.4: Mean budget share differences for selected categories disaggregated by settlement type.....118
Figure 3.5: Mean budget share differences for selected categories disaggregated by income tercile.....121
Figure 3.6: Consumer Price Indices for selected categories in South Africa (1994-2006).....126
Figure 4.1: UNICEF Conceptual Framework on the Determinants of Maternal and Child Nutrition, 2020.....147
Figure 4.2: Mediated effect of cigarette or alcohol expenditure on child health outcomes.....156

LIST OF TABLES

Table 1.1: Total alcohol per capita consumption (in litres of pure alcohol) and prevalence of heavy episodic drinking (HED) (in %) among the total population aged 15+ years and among drinkers (15+ years) by WHO region, 2016 5
Table 1.2: Key indicators on levels of alcohol consumption among drinkers only for selected African countries, 2016 (both sexes; 15+ years) 7
Table 2.1: Identification of income outliers across IESs 37
Table 2.2: Impact of outliers on income 37
Table 2.3: Descriptive statistics from the income and expenditure surveys (IES) 1995, 2000, 2005-06 and 2010-11 41
Table 2.4: Mean budget shares from the income and expenditure surveys (IES) 1995, 2000, 2005-06 and 2010-11 disaggregated by tobacco use. 42
Table 2.5: Differences in expenditure shares between smoking and non-smoking households 47
Table 2.6: Chi-square statistics for test of heterogeneous preferences between tobacco users and non-users 51
Table 2.7: Three stage-least squares (3SLS) regression results – Full sample, urban and rural sub-samples.....53
Table 2.8: Three Stage-Least Squares (3SLS) Regression results – Tercile sub-samples.....54
Table 2.9: Chi-square statistics for endogeneity test for ptt lnM and (lnM)2 56
Table 2.10: Chi-Square statistic for Pagan-Hall test for heteroskedasticity 58
Table 2.11: Lower and upper bounds for the estimates of tobacco expenditure variable 60
Table 3.1: Descriptive statistics from the Income and Expenditure Surveys (IES) 1995, 2000 and 2005-06 99
Table 3.2: Test for violation of Parallel Trends Assumption.....103
Table 3.3: Difference in descriptive statistics between tobacco user and non-tobacco user households before and after reweighting procedure 105
Table 3.4: Descriptive statistics for comparison for pooled sample of 1995 and 2000 Income and Expenditure Surveys 107
Table 3.5: Descriptive statistics for comparison pooled sample of 2000 and 2005-06 Income and Expenditure Surveys 109

Table 3.6: Crowding-out transitions from 1995 to 2000 IES and from 2000 to 2005/06 IES using all samples 112
Table 3.7: Crowding-out transitions from 1995 to 2000 IES and from 2000 to 2005/06 IES using rural households 116
Table 3.8: Crowding-out transitions from 1995 to 2000 IES and from 2000 to 2005/06 IES using urban households 117
Table 4.1: Summary descriptive statistics 158
Table 4.2: The 2SLS Estimation of the Effect of Lagged Tobacco and Alcohol Expenditure on Child Nutrition Outcomes 161
Table 4.3: Chi-square statistic from endogeneity tests 163
Table 4.4: Auxiliary regression coefficients 164
Table 4.5: First-stage F-statistics for endogenous lagged cigarette expenditure, alcohol expenditure, and household income 165
Table 4.6: Results of regression of lagged food expenditure regressed on lagged cigarette and alcohol expenditure 168
Table 4.7: The mediated effect of cigarette and alcohol expenditure by food expenditure on child nutrition outcomes 169

Table A 1: Mean budget shares from the Income and Expenditure Surveys (IES) 1995, 2000, 2005-06 and 2010-11 disaggregated by tobacco use and settlement type 63
Table A 2:Mean budget shares from the income and expenditure surveys (IES) 1995, 2000, 2005-06 and 2010-11 disaggregated by tobacco use and tercile 64
Table A 3: Detailed budget share for food 65
Table A 4: Detailed budget share for housing 66
Table A 5: Detailed budget share for transport/ 67
Table A 6: Auxiliary OLS regression for validity of inclusivity criteria of adult-sex ratio in 1995, 2000, 2005/06 and 2010/11 samples 68
Table A 7: Auxiliary OLS regression for validity of inclusivity criteria of adult-sex ratio in 1995, 2000, 2005/06 and 2010/11 samples 69
Table A 8: Statistics for instrument validity – full sample, urban and rural samples 70
Table A 9: Statistics for instrument validity – Tercile sub-samples 70
Table A 10: Three Stage-Least square (3SLS) Regression 1995 IES Full Sample (Equation 5 specification) 71

Table A 11: Three Stage-Least square (3SLS) Regression 1995 IES Urban Sample (Equation 5 specification) 73
Table A 12: Three Stage-Least square (3SLS) Regression 1995 IES Rural Sample (Equation 5 specification) 75
Table A 13: Three Stage-Least square (3SLS) Regression 1995 IES Full Sample (Equation 7 specification) 77
Table A 14: Three Stage-Least square (3SLS) Regression 1995 IES Urban Sample (Equation 7 specification) 79
Table A 15: Three Stage-Least square (3SLS) Regression 1995 IES Rural Sample (Equation 7 specification) 81

Table B 1: Standard difference in difference estimation without covariates – Full sample .. 130
Table B 2: Standard difference in difference estimation with covariates – Full sample 131
Table B 3: Abadie's Difference in difference estimation without covariates – Full sample . 132
Table B 4: Crowding-out transitions from 1995 to 2000 IES and from 2000 to 2005/06 IES using Tercile 1 households 133
Table B 5: Crowding-out transitions from 1995 to 2000 IES and from 2000 to 2005/06 IES using Tercile 2 households 134
Table B 6: Crowding-out transitions from 1995 to 2000 IES and from 2000 to 2005/06 IES using Tercile 3 households 135
Table B 7: Crowding-out transitions from 1995 to 2000 IES and from 2000 to 2005/06 IES disaggregated by rural-urban and income category 136

Table C 1: Descriptive statistics for household assets used to construct asset index 178
Table C 2: Results for the Kaiser-Meyer-Olkin test of sampling adequacy 179
Table C 3: Coefficient estimates on asset variables 181
Table C 4: OLS regression results on the impact of lagged cigarette and alcohol expenditure on child health outcomes 182
Table C 5: OLS regression results on the impact of lagged alcohol expenditure on child health outcomes 184
Table C 6: 2SLS regression results on the impact of current cigarette expenditure on child health outcomes 185

Table C 7: 2SLS regression results on the impact of log of lagged cigarette expenditure on child health outcomes 186
Table C 8: 2SLS regression results on the impact of current alcohol expenditure on child health outcomes 187
Table C 9: 2SLS regression results on the impact of log of lagged alcohol expenditure on child health outcomes 188
Table C 10: The Mediated effect of Alcohol Expenditure by Food Expenditure on Child Nutrition Outcomes.....190

1. CHAPTER 1: INTRODUCTION

1.1 Background and motivation

Tobacco and alcohol consumption are significant preventable factors contributing to worldwide mortality and morbidity. In 2020, the World Health Organization (WHO) reported an annual death toll of approximately 8.7 million attributable to tobacco use (WHO, 2018). In comparison, an estimated 3.3 million annual deaths were linked to alcohol-related factors. The global economic burden associated with smoking was estimated at USD 1.4 trillion in 2016, representing nearly 2% of the annual global gross domestic product (GDP) (Goodchild, Nargis & d'Espaignet, 2018). Developing nations bore 40% of this economic cost. Likewise, economic costs associated with alcohol consumption are estimated to be around 2.6% of GDP (Manthey et al., 2021).

1.1.1 Global trends in tobacco and alcohol use

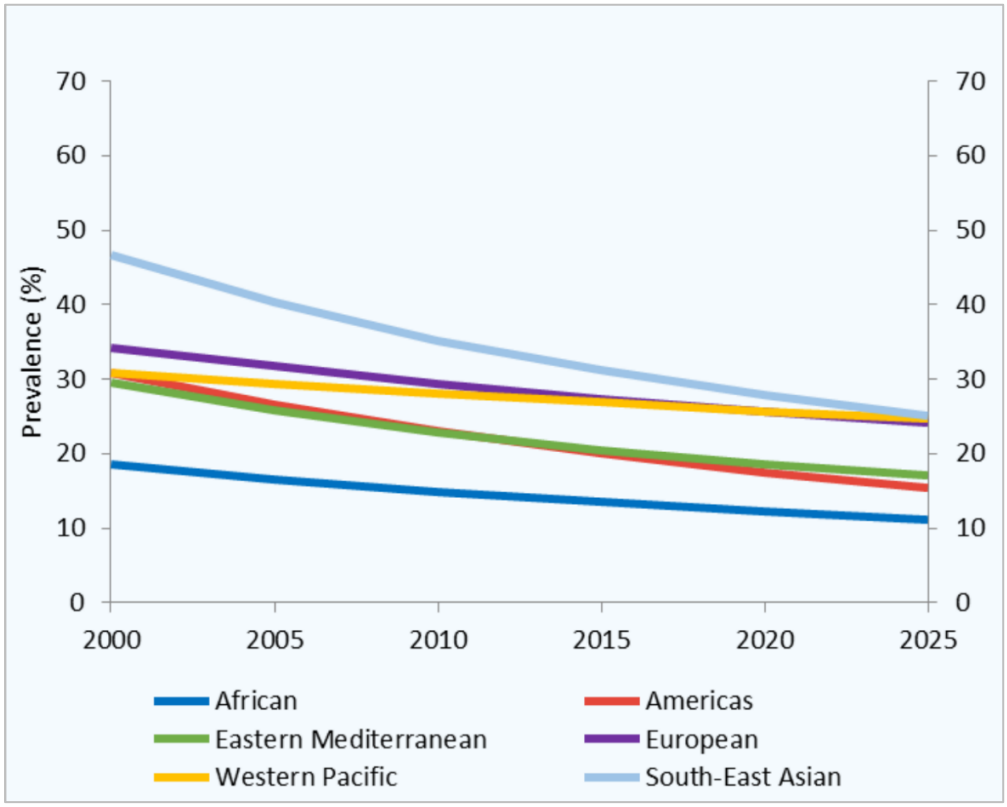
The prevalence of tobacco use has been on the decline in recent years across all WHO regions (refer to Figure 1.1) (WHO, 2018). For example, smoking rates in the historically highest-prevalence region, Southeast Asia, which had a prevalence rate of 47% in 2000, is projected to decrease to 25.1% by 2025. The African region, historically the lowest-prevalence region, is anticipated to decline from 18.5% in 2000 to 11.2% by 2025. Despite these declines in smoking prevalence, it is noteworthy that only the Americas region is on track to achieve the targeted 30% relative reduction in tobacco use prevalence by 2025, a target adopted 2014. Other regions, such as the Western Pacific and Europe, are expected to experience relatively slower rates of decline.

It is important to note that these reductions in prevalence mask a distinct issue arising from population growth. Specifically, although the global number of smokers had remained stable at approximately 1.1 billion for over two decades, the number of smokers in three regions was still projected to increase between 2015 and 2025 (WHO, 2018). These regions include Africa, where the number of smokers was expected to grow from 57 million to 67 million smokers, the Eastern Mediterranean, projected to increase from 77 million to 91 million smokers, and South-east Asia, where the number is expected to rise from 230 million to 252 million smokers.

In 2016, more than half the world's adult population (above 15 years) had not consumed alcohol in the preceding 12 months of the survey (See Figure 1.2) (World Health Organization,

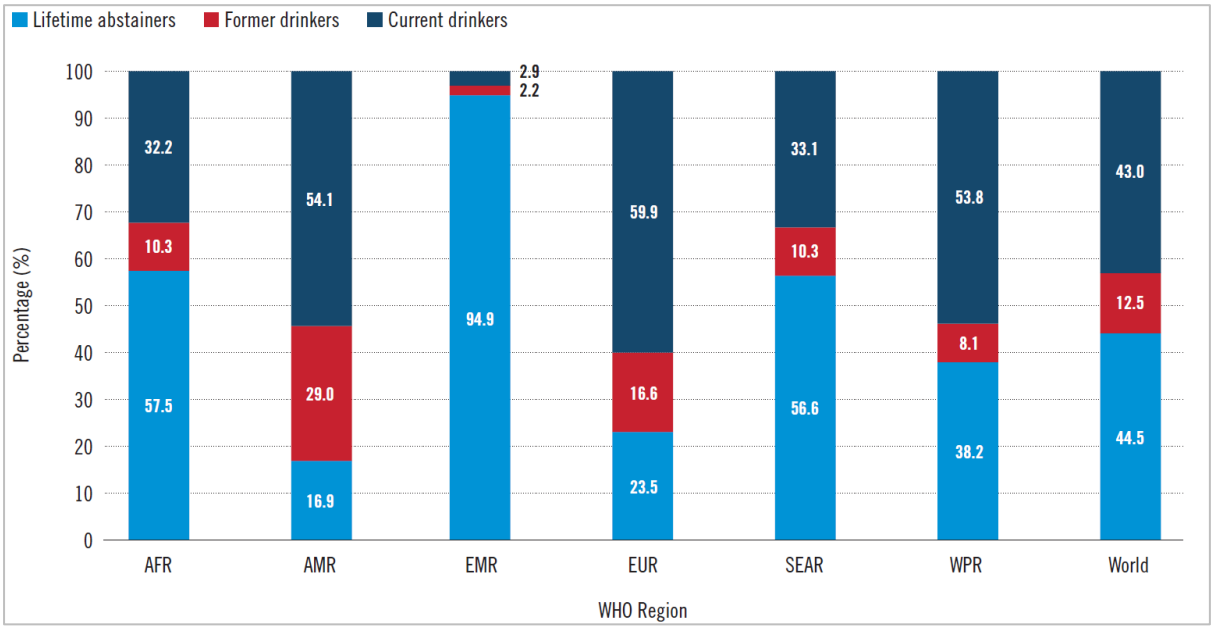
2019b). However, on a global level, the proportion of current alcohol drinkers is anticipated to increase from 47% in 2017 to 50% in 2030 (Manthey et al., 2019).

Figure 1.1: Trends in current tobacco use among people aged 15 years and above



Source: WHO global report on trends in prevalence of tobacco use 2000-2025

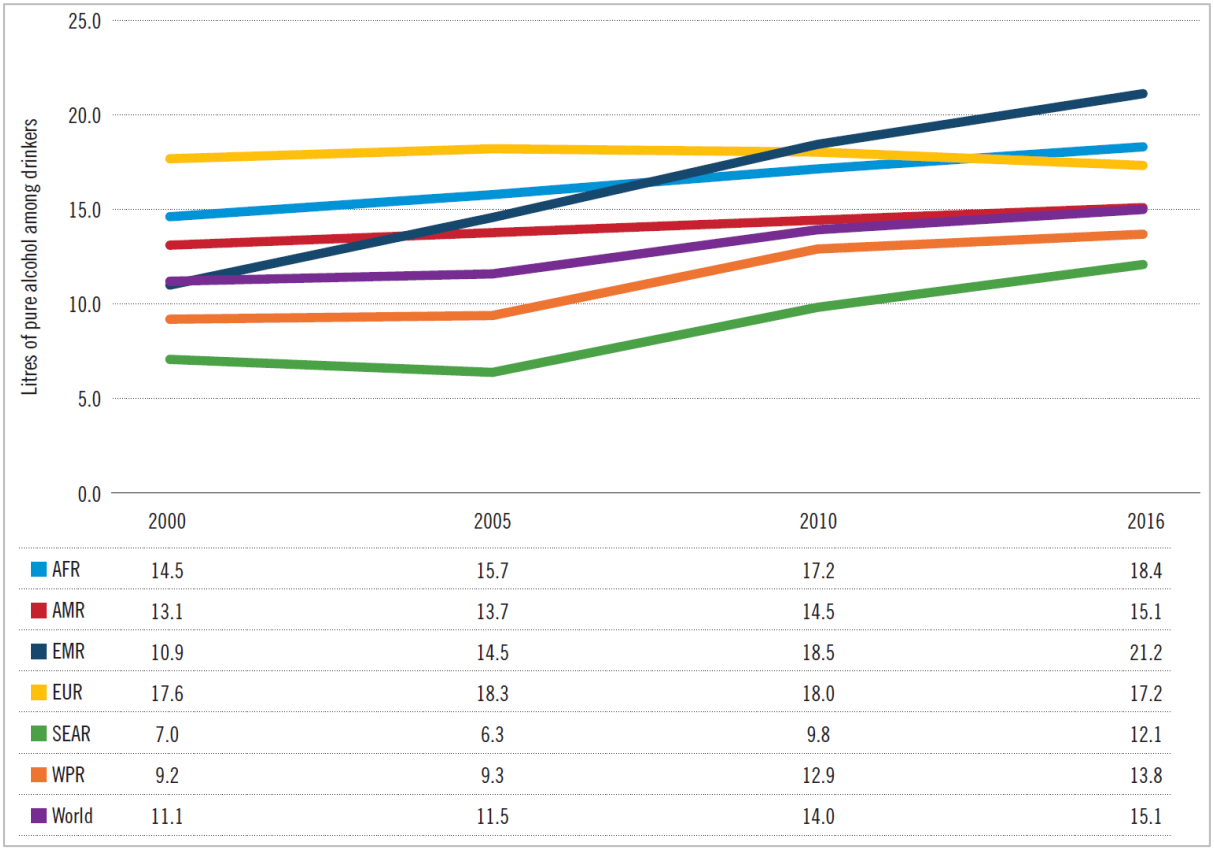
Figure 1.2: Percentage of current drinkers, former drinkers, and lifetime abstainers (in %) among the total population (15+ years)



Source: WHO Global Status Report on Alcohol 2018

A useful indicator to measure the level of alcohol consumption is the total alcohol per capita consumption (APC), defined as the total per capita pure alcohol consumption within a calendar year. This can be calculated for the total population, usually those above 15 years, or only for adult drinkers. Globally, total APC in the total population remained relatively stable between 2010 and 2016 at around 6.4 litres, although there were significant differences in trends across regions (World Health Organization, 2019b). In contrast, total APC among drinkers has increased across all regions since 2000, except the WHO European Region. This implies that, while the proportion of people that consume alcohol has decreased, the people that do drink alcohol have increased their per capita consumption.

Figure 1.3: Trends in total alcohol per capita consumption (APC) among drinkers (15+ years) in litres of pure alcohol in WHO regions, 2000−2016



	2000	2005	2010	2016
AFR	14.5	15.7	17.2	18.4
AMR	13.1	13.7	14.5	15.1
EMR	10.9	14.5	18.5	21.2
EUR	17.6	18.3	18.0	17.2
SEAR	7.0	6.3	9.8	12.1
WPR	9.2	9.3	12.9	13.8
World	11.1	11.5	14.0	15.1

Source: WHO Global Status Report on Alcohol 2018

Unlike tobacco use, the health and societal impacts of alcohol use are significantly shaped by patterns of use. The total APC among drinkers can vary based on on the primary type of alcoholic beverage consumed. For example, wine typically has a lower alcohol content than spirits. Additionally, the manner of consumption plays a crucial role, particularly whether alcohol is consumed regularly in moderation or in heavy episodic drinking patterns.

Heavy episodic drinking (HED) is defined as a pattern of drinking where 60 or more grams of pure alcohol is consumed on at least one occasion at least once a month. Higher total APC is positively associated with higher prevalence of HED. The global prevalence of HED among all adults (i.e., above 15 years) decreased from 22.6% in 2000 to 18.2% in 2016 (Manthey et al., 2019). However, about 50% of drinkers in the African region still consume alcohol in heavy doses, the highest prevalence of HED among drinkers across all regions (refer to Table 1.1).

Table 1.1: Total alcohol per capita consumption (in litres of pure alcohol) and prevalence of heavy episodic drinking (HED) (in %) among the total population aged 15+ years and among drinkers (15+ years) by WHO region, 2016

	Among all (15+ years)		Among drinkers (15+ years)		
WHO Region	**Total APC**	**HED prevalence (%)**	**Total APC**	**HED prevalence (%)**	**Number of HED drinkers (in millions)**
AFR	6.3	17.4	18.4	50.2	100.9
AMR	8	21.3	15.1	40.5	163.9
EMR	0.6	0.5	21.2	10.4	2.3
EUR	9.8	26.4	17.2	42.6	197.9
SEAR	4.5	13.9	12.1	40.7	195.7
WPR	7.3	21.9	13.8	40.6	332.4
World	**6.4**	**18.2**	**15.1**	**39.5**	**993.0**

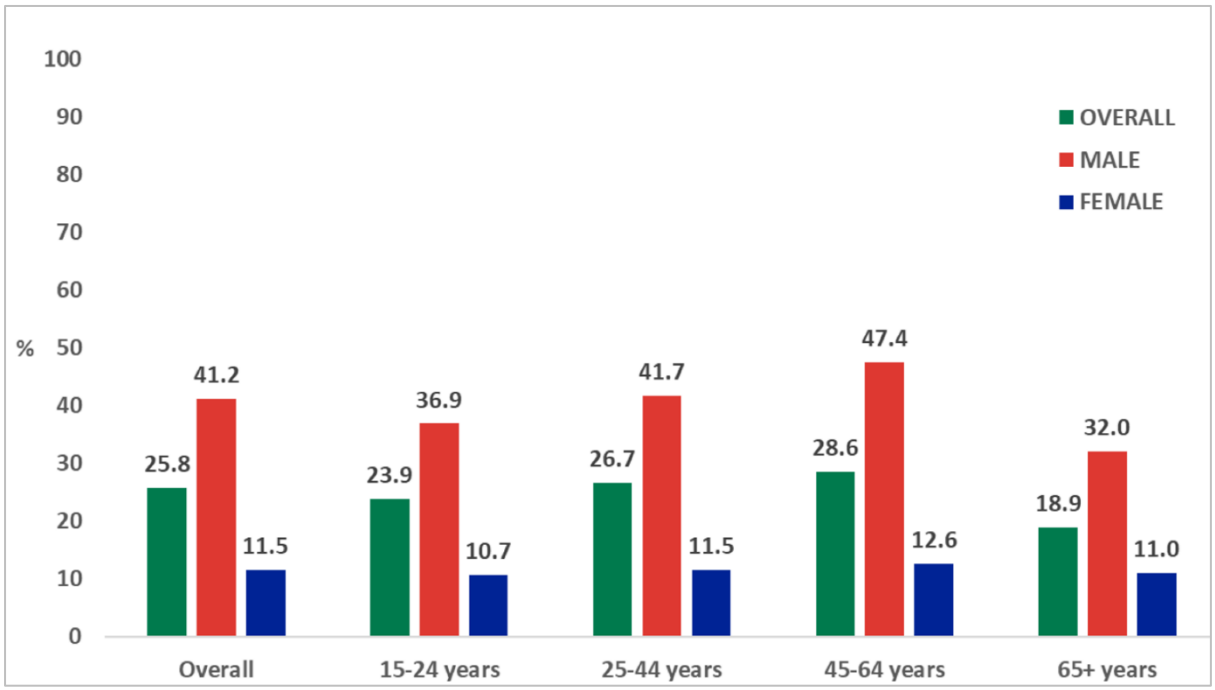
Source: WHO Global Status Report on Alcohol 2018

The trends discussed in this context provide a nuanced perspective. While certain indicators suggest an overall trend of progress in reducing the consumption of tobacco and alcohol products, there are significant regional disparities. In specific cases, such as the total APC among drinkers (see Figure 1.3), there are reasons for concern. What further complicates this situation is the epidemiological nature of the problem, characterised by a sizable time lag between tobacco or alcohol use and the onset of related non-communicable diseases (NCD), morbidity, and mortality. Consequently, even where a decline in usage is observed, there remains the potential for NCD-related outcomes to emerge for many years following the observed reductions in usage.

1.1.2 Tobacco use in South Africa

The most recent Global Adult Tobacco Survey (GATS) conducted in South Africa estimates that the prevalence of tobacco use is about 25.8% among adults 15 years and older (South African Medical Research Council, 2022). Men have a higher prevalence (41.2%) compared to women (11.4%). This overall tobacco use prevalence represents a substantial increase from previous estimates of 17.6% in 2012 (Reddy et al., 2015), 22.5% in 2016 (National Department of Health & ICF, 2019), and about 19.9% in 2017 (Vellios, van Walbeek & Ross, 2020). While there was some variation in smoking prevalence between 2010 and 2020, recent estimates suggest a substantial increase in tobacco use in the country.

Figure 1.4: Current Tobacco Smoking Prevalence by Age Group and Gender, 2021



Source: Global Adult Tobacco Survey South Africa Fact sheet, 2021

As illustrated in Figure 1.4, the highest prevalence is found among those aged 45 to 64, followed by the 25-to-44 and 15-to24 age groups. Adults older than 64 have the lowest prevalence, at 18.9% overall. This distribution of prevalence is similar irrespective of the gender divide.

A major driver of the observed increase in tobacco use in the country is likely to be the proliferation of the illicit cigarette trade. Prior to 2010, South Africa's illicit cigarette market share was less than 10% (Vellios, van Walbeek & Ross, 2020). However, after 2010, and in particular after 2015, the share of illicit cigarettes in the market rose dramatically to over 30% in 2017 (Van Der Zee, Van Walbeek & Magadla, 2020), and 54%in 2021 (Vellios, 2022).

The impact of tobacco use on South African society is profound. Boachie, Rossouw & Ross (2021) estimated that the economic cost of smoking in 2016 was as much as R42 billion, or the equivalent of 0.97% of the GDP in that year. Of this, about R14.5 billion was for healthcare costs alone, which include hospitalizations and outpatient department visits. The healthcare cost of smoking-related diseases was 4.1% of total South African health expenditure.

1.1.3 Alcohol use in South Africa

In 2016, approximately 69% of South African adults aged 15 years and above indicated that they had not consumed alcohol in the preceding 12 months (World Health Organization, 2019b). The majority were lifetime abstainers (53.5%), with a smaller portion classified as former drinkers (15.5%). Consequently, the prevalence of alcohol use during the previous year stood at approximately 31%. However, despite this relatively low prevalence, the volume of alcohol consumed per drinker ranked among the highest on the African continent.

Specifically, South African drinkers averaged nearly 30 litres of pure alcohol consumption per capita, an amount exceeded only by Namibian drinkers, at 32.4 litres, and Eswatini drinkers at 34.4 litres of pure alcohol each (refer to Table 1.2).

Table 1.2: Key indicators on levels of alcohol consumption among drinkers only for selected African countries, 2016 (both sexes; 15+ years)

Country	Alcohol, heavy episodic drinking (15+), drinkers only, past 30 days (%)*	Total consumption (drinkers only)§	Average daily intake±
Eswatini	60.0	34.4	74.3
Namibia	59.7	32.4	70.1
South Africa	59.0	29.9	64.6
Botswana	59.4	26.2	56.6
Uganda	56.9	26.0	56.3
United Republic of Tanzania	55.8	25.8	55.7
Rwanda	54.6	25.6	55.4
Nigeria	55.0	25.5	55.0
Cameroon	47.4	22.6	48.8
Burkina Faso	46.3	22.2	47.9

Source: WHO Global Status Report on Alcohol 2018. * Consumed at least 60 grams or more of pure alcohol on at least one occasion in the past 30 days. § Total (recorded + unrecorded) alcohol per capita consumption, drinkers only, 2016 (in litres of pure alcohol; 15+ years). ± Average daily intake of alcohol per capita, drinkers only, 2016 (in grams of pure alcohol; 15+ years).

The public health implications of this situation are substantial. For instance, Probst et al. (2018) estimated that about 62,300 adults died from alcohol-related causes in 2015, of which 60% of deaths occurred in the lowest third of the socioeconomic distribution.

1.1.4 The policy response to tobacco and alcohol use

In 2003, the World Health Assembly unanimously adopted the WHO Framework Convention on Tobacco Control (WHO FCTC) as a response to the worldwide tobacco epidemic (World Health Organization, 2003). As the first global health treaty negotiated under the auspices of the WHO, the treaty aimed to mandate cost-effective tobacco-control measures in participating countries. These interventions included increasing excise taxes; legislating for smoke-free areas; prohibiting tobacco advertising, promotion, and sponsorship; supporting alternative economic activities for tobacco growers and workers; implementing health warnings on tobacco packaging; and safeguarding public health policies from tobacco-industry interference. The treaty came into effect in 2005 and currently has 183 signatories (United Nations, 2003).

In addition, the *Global Strategy to Accelerate Tobacco Control* was adopted at the seventh Conference of the Parties to the WHO FCTC in 2017 (World Health Organization, 2019a). Spanning from 2019 to 2025, this strategy guides ongoing tobacco-control endeavours at the global level, including initiatives such as WHO's MPOWER[1] measures, designed to assist countries in implementing the provisions of the WHO FCTC to reduce demand for tobacco products.

Global alcohol-control efforts lack a dedicated treaty, but the World Health Assembly adopted the *Global Strategy to Reduce the Harmful Use of Alcohol* at its sixty-third sitting in 2010 (World Health Organization, 2010). The Global Strategy represents the only global policy framework on alcohol. It is the culmination of several regional initiatives aimed at promoting and supporting local, regional, and global actions to prevent and reduce the harmful use of alcohol. Related to this is the Global Alcohol Action Plan 2022-2030 that aims to strengthen the implementation of the Global Strategy (World Health Organization, 2022). It relies on the implementation of an initiative, known as SAFER. Like MPOWER in tobacco control, this initiative by the WHO represents five highly cost-effective interventions to mitigate alcohol-related harm. These interventions comprise: (1) **S**trengthening restrictions on alcohol availability; (2) **A**dvancing and enforcing drink-driving countermeasures; (3) **F**acilitating

[1] MPOWER is an acronym for six evidence-based components, namely: (1) **M**onitor tobacco-use and prevention policies; (2) **P**rotect people from tobacco smoke; (3) **O**ffer help to quit tobacco use; (4) **W**arn about the dangers of tobacco; (5) **E**nforce bans on tobacco advertising, promotion, and sponsorship; and (6) **R**aise taxes on tobacco.

access to screening, brief interventions, and treatment; (4) **E**nforcing bans or comprehensive restrictions on alcohol advertising, sponsorship, and promotion; and (5) **R**ising alcohol prices through excise taxes and pricing policies.

The WHO's attention has been focused strongly on tobacco, especially since the ratification of the WHO FCTC in 2005. For many years, alcohol use and misuse have not received the same degree of attention. However, in recent years this has been changing, with increased recognition of the similarities these two products. In fact, the shift in focus has turned towards addressing the four modifiable risk-factors for NCDs, namely: tobacco, alcohol, unhealthy diets, and physical inactivity.

Tobacco control in South Africa

South Africa became a Party to the WHO FCTC on July 18, 2005. However, the foundation of the country's tobacco-control efforts predates this commitment, and is primarily rooted in the Tobacco Control Act 83 of 1993. Subsequently, this Act underwent four amendments between 1993 and 2008, through the General Law Fifth Amendment Act 157 of 1993, the Tobacco Products Control Amendment Act 12 of 1999, the Tobacco Products Control Amendment Act 23 of 2007, and the Tobacco Products Control Amendment Act 63 of 2008. The Act encompasses various aspects of tobacco control, including smoking restrictions, tobacco advertising, promotion, and sponsorship regulations, and packaging and labelling standards.

In 1995, regulations pertaining to the labelling, advertising, and sale of tobacco products were introduced. These regulations required cigarette packaging to display rotating text-only health warnings, covering 15% of the front and 25% of the back of the package. While these regulations did not ban tobacco advertising outright, they mandated warnings for written, pictorial, video, and radio advertisements. Subsequently, in 1999, these regulations were reinforced with a complete ban on the advertising of tobacco products.

In 2000, additional regulations regarding smoke-free areas were introduced, mandating that all public places, including bars, pubs, restaurants, casinos, and airports, become smoke-free. However, as a concession largely influenced by the tobacco industry under the guise of representing the hospitality sector, the legislation permitted up to 25% of the floor space to be allocated to designated smoking areas.

Further amendments to the regulations were made in 2007 and 2008. These amendments focused on ensuring that exported tobacco products conformed to the standards of the

destination country, and further reinforced the ban on advertising or promoting tobacco products in South Africa.

However, many of these regulations fall short of what is considered best practice today, and are not entirely aligned with WHO FCTC recommendations. For instance, while many countries have adopted graphic health warnings, South Africa has retained text-only warnings on cigarette packaging. Further, although this has not been tested in the South African case, the partial smoking ban has proven largely ineffective in reducing second-hand smoke exposure amongst other nations (Catalano & Gilleskie, 2021).

Recognising these limitations, the National Department of Health introduced the Tobacco Products and Electronic Delivery Systems Control Bill in 2018. The bill is currently (2024) being debated, and aims to align South Africa's tobacco-control measures with WHO FCTC guidelines and global best practices. Proposed measures include a comprehensive ban on indoor smoking in public places and specific outdoor areas, the prohibition of cigarette sales via vending machines, the removal of point-of-sale displays, the introduction of plain packaging featuring graphic health warnings, and the regulation of electronic nicotine delivery systems (ENDS) and electronic non-nicotine delivery systems (ENNDS) within the country.

Regarding tobacco taxation, the government implemented substantial excise tax increases on cigarettes as early as 1994. These increases led to the real excise tax per pack rising by over 250% between 1994 and 2004, accompanied by a corresponding 127% increase in the real price over the same period (Linegar & van Walbeek, 2018, van Walbeek, 2004). Nevertheless, since 2010, the real excise tax has remained relatively stable. Post-2010, the government implemented annual adjustments to the excise tax, in line with inflation, to the present day.

Alcohol control in South Africa

As for tobacco control, many of the initiatives to enhance alcohol control in South Africa began in the late 1990s (Parry, 2010). However, they were not passed into law until well into the early 2000s. The South African National Department of Health initiated policy development on the restriction of alcohol advertising and at-counter advertising in 1997, while regulations mandating health warnings were only promulgated in 2007. Similarly, the National Department of Trade and Industry initiated reforms to the liquor trade that were only promulgated in 2003. These regulations included restrictions related to the employment of people who are aged less than 18 years in the liquor trade, prohibitions on the sale of liquor to individuals aged less than

18 years of age, and a provision that the Minister of Trade and Industry, in consultation with the Minister of Health, may prescribe public health notices to be displayed at points of sale.

Excise taxes on alcoholic beverages in South Africa have a long history. Specifically, spirits and beer have been subject to these taxes since the early twentieth century, while sorghum beer was added to the taxation system in 1992 (Van Walbeek & Blecher, 2014). These taxes are applied in the form of specific excise taxes, meaning that they are based on a fixed amount of volume of alcohol in the beverage or the volume of the beverage, rather than being *ad valorem* (i.e. levied as a percentage of value).

As of the fiscal year 2023/24 (South African Revenue Services, 2023), the excise tax rates for alcoholic beverages stand as follows:

- beer: R127.40 per litre of absolute alcohol,
- spirits: R257.23 per litre of absolute alcohol,
- wine: R5.20, R8.77 and R16.64 per litre of unfortified, fortified, and sparkling wine, regardless of alcohol content; and
- sorghum beer: 7.82 cents per litre of sorghum beer, regardless of its alcohol content.

Alcohol spending has been found to exhibit regressive patterns (Fontes Marx et al., 2019). However, this approach is often indiscriminate, as it does not consistently target the affordability of the cheapest alcoholic beverages, which are frequently consumed by those engaging in harmful drinking behaviours. As a result, there is increasing interest in the adoption of minimum unit pricing (MUP) as a complementary policy measure. This interest is particularly driven by evidence suggesting that harmful alcohol consumption is disproportionately prevalent among low-income populations, where affordability plays a critical role in consumption patterns (Callinan, Room & Dietze, 2015).

A study by Van Walbeek & Chelwa (2021) sheds light on this issue, revealing that 84% of all alcohol consumption in South Africa originates from households characterised by occasional to regular heavy-drinking habits. Strikingly, these households constitute only a quarter of all households in the country. The study's analysis strongly indicates that the introduction of an MUP could be influential in curtailing alcohol consumption within the segment of regular heavy-drinking households, with discernible, albeit somewhat lesser, effects expected among occasional heavy-drinking and intermediate-drinking households.

Furthermore, an additional study underscores the potential of this policy in not only reducing the harm associated with alcohol consumption, but also in diminishing health inequalities (Gibbs et al., 2022). Although the current thesis does not touch on MUP policies, it may have important implications for overall alcohol policy in South Africa.

1.2 The importance of understanding household expenditure patterns

Over the past decade, tobacco and alcohol policy issues have attracted significant global attention as substantial obstacles to global development efforts. They are acknowledged to be a major threat to the three dimensions of development, namely health, economies, and the environment, thereby impeding progress towards sustainable development (Dünnbier, 2020, World Health Organization, 2019a). Consequently, these issues are explicitly incorporated into the Sustainable Development Goal (SDG) targets and indicators. Goal Three, focusing on Health, naturally includes these concerns. Specifically, Target 3.5 aims to *strengthen the prevention and treatment of substance abuse, including narcotic drug abuse and harmful alcohol use*. For tobacco, Target 3a aims to *enhance the implementation of the WHO FCTC in all countries as appropriate* (WHO Framework Convention on Tobacco Control Secretariat, 2017).

The broader socio-economic impacts of tobacco and alcohol extend beyond health. These substances have the potential to indirectly influence other SDG targets due to their multifaceted effects. For instance, alcohol-related issues intersect with various other SDGs, including those related to gender equality and the reduction of violence. Similarly, tobacco is linked to goals pertaining to poverty alleviation, food security, economic growth, and education (World Health Organization, 2019a).

Understanding household expenditure patterns is crucial within this context as it reveals how consumption of tobacco and alcohol can lead to significant economic trade-offs for households. This thesis explores three key themes in the economics of tobacco and alcohol control in South Africa with a focus on household expenditure patterns and their broader socio-economic impacts.

Economic trade-offs and consumption choices

Household expenditure patterns provide valuable insights into the financial decisions made by families and individuals. When a significant portion of household income is allocated to tobacco and alcohol, it inevitably affects other areas of expenditure (Chelwa & Koch, 2019, Gallego, Juan Miguel, Paraje & Rodríguez-Lesmes, 2024, Husain et al., 2018, Jin & Cho, 2021,

Nguyen & Nguyen, 2020, Nyagwachi, Chelwa & van Walbeek, 2020, Paraje & Araya, 2018, Sánchez & Gómez, 2024, Vladisavljevic et al., 2024). This trade-off can lead to reduced spending on essential goods and services such as healthcare, education, nutrition, and housing, which are crucial for the well-being and development of household members.

For instance, higher spending on tobacco and alcohol might result in less investment in children's education, adversely affecting their future earning potential and perpetuating cycles of poverty. Similarly, reduced spending on nutritious food due to high expenditure on these substances can lead to health issues, further straining household finances through increased medical costs.

Socio-economic impacts of tobacco and alcohol consumption

The socio-economic impacts of tobacco and alcohol consumption extend beyond individual households to the broader community and economy. High levels of consumption can lead to increased healthcare costs at the national level due to the treatment of diseases related to smoking and alcohol abuse, such as cancer, liver disease, and cardiovascular conditions (Manthey et al., 2021, Pichon-Riviere et al., 2020). Additionally, productivity losses due to illness or premature death of individuals addicted to these substances can negatively impact the economy.

Analysing household expenditure on tobacco and alcohol also helps in understanding the distribution of these impacts across different socio-economic groups. Lower-income households tend to spend a higher proportion of their income on these products, exacerbating their economic vulnerability. This unequal burden underscores the need for targeted public health interventions and policies that address the specific needs of these groups.

Policy implications and control measures

Exploring household expenditure patterns in the context of tobacco and alcohol consumption also informs the development and implementation of effective control measures and policies. By understanding where and how households allocate their resources, policymakers can design targeted interventions such as taxation, subsidies, and public awareness campaigns that can effectively reduce consumption and mitigate its negative impacts.

For example, higher taxes on tobacco and alcohol can deter consumption and generate additional revenue that can be invested in healthcare and education. Subsidies for healthier alternatives or smoking cessation programs can support individuals in making better

consumption choices. Public awareness campaigns can educate the population on the harmful effects of tobacco and alcohol, encouraging more informed and healthier decision-making.

1.3 Research objectives

This thesis concerns itself with addressing three primary research objectives:

- To analyse the economic impact of tobacco expenditure on household expenses in South Africa.
- To investigate the crowding-out effect of tobacco and alcohol expenditure on other critical household spending categories.
- To assess the impact of parental tobacco and alcohol consumption on child health outcomes.

The specific research questions to be addressed are the following:

- What categories of household expenditure are displaced by tobacco expenditure?
- How do these displacement patterns change with varying levels of tobacco taxes?
- How does parental tobacco and alcohol use affect child health outcomes, particularly in terms of nutrition?

1.4 Research Methods

To investigate the displacement of household expenditures by tobacco use, this study employs a comprehensive cross-sectional analysis of crowding-out, drawing on methodologies commonly found in the literature (Husain et al., 2018, John, 2008, John et al., 2019, Pu et al., 2008, San & Chaloupka, 2016). The analysis begins with a simple comparison of budget shares and progresses to the application of a demand system model, specifically the Quadratic Almost Ideal Demand System (QuAIDS) developed by Banks, Blundell & Lewbel (1997). To address potential endogeneity in tobacco expenditure, an instrumental variable (IV) approach is adopted, utilising the adult-sex ratio as an instrument (Chelwa & Van Walbeek, 2014, John, 2008). Recognising the limitations of the instrument, the study employs the method proposed by Nevo & Rosen (2012) to relax the exclusion restriction, thereby using imperfect instrumental variables (IIV) to identify a range of parameter values within which the true point estimate is likely to fall.

To examine changes in crowding-out over the period from 1995 to 2000, the study employs a difference-in-differences (DiD) approach. The study incorporates a DiD method that integrates

propensity scores into a direct weighting scheme proposed by Abadie (2005). The study also performs various tests to ascertain the plausibility of the parallel trends assumption.

Lastly, to investigate the effect of parental tobacco and alcohol expenditure on child nutrition outcomes, the study uses an IV method augmented with mediation analysis. The mediation analysis aims to reveal the mechanisms through which the observed effects occur. This type of analysis decomposes the total treatment effect into two components: the indirect effect and the direct effect (Celli, 2021, Heckman, James J & Pinto, 2015).

1.5 Contribution

This study contributes to the literature in several significant ways.

Chapter 2 enhances the existing body of knowledge regarding the categories of expenditure displaced by cigarette expenditure. Previous studies in sub-Saharan Africa, including those conducted in Zambia (Chelwa & Van Walbeek, 2014), South Africa (Chelwa & Koch, 2019, Koch & Tshiswaka-Kashalala, 2008), Ghana (Masa-Ud, Chelwa & Van Walbeek, 2020), and Kenya (Nyagwachi, Chelwa & van Walbeek, 2020) have explored this issue. This study adds to the limited number of studies in the region by reinforcing the findings of earlier research on the impact of cigarette expenditure on household spending patterns.

Second, the study extends our understanding of crowding-out by analysing how this displacement phenomenon has evolved over time, especially in the presence of substantial cigarette-price increases. Employing the DiD method integrated with propensity scores in a direct weighting scheme (Abadie, 2005), the study sidesteps the potential pitfalls associated with propensity-score matching DiD methods and approaches used in previous research (Nyagwachi, Chelwa & van Walbeek, 2020). This methodological improvement allows for a more robust analysis of the dynamic impacts of tobacco price changes on household expenditure.

Beyond its scholarly contributions, the dynamic analysis of tobacco tax impacts holds significant policy implications. The literature indicates that tobacco consumption declines due to price increases through reductions on both the extensive margin (fewer smokers) and the intensive margin (fewer cigarettes smoked per smoker). This study's dynamic analysis elucidates how these factors interact with fiscal policy changes, influencing household budgeting decisions. It highlights the effectiveness of cigarette tax policies over time, underscoring their continued relevance. Additionally, it uncovers potential unintended

consequences for households that continue purchasing cigarettes despite substantial price hikes, such as deeper expenditure displacements, which are crucial for policymakers to consider.

The study also contributes to our understanding of child health by examining the impact of parental alcohol and tobacco consumption on children aged 6 months to 5 years, a critical developmental phase often overlooked in existing literature. It adopts a holistic approach by considering both parents' health behaviours, as indicated by household expenditure on cigarettes and alcohol, rather than focusing solely on maternal behaviour, which is more commonly studied. Furthermore, it introduces two methodological enhancements. First, it employs various instruments to address endogeneity in cigarette and alcohol expenditure alongside household income. Second, it utilises the inverse hyperbolic sine (IHS) transformation for expenditure variables, addressing potential distribution issues. These methodological refinements enhance the credibility of the findings, providing a more accurate understanding of the relationships involved.

1.6 Themes and outline of the thesis

This thesis is composed of five main chapters, structured around three key themes in the economics of tobacco and alcohol control in South Africa, with an introductory chapter and a concluding chapter. These themes are briefly discussed below:

Chapter 1: Introduction

This chapter introduces the thesis, outlining the background and overall motivation for the study. It details the research objectives, the global and local context and the overall contribution of the research.

Chapter 2: The opportunity cost of tobacco use in South Africa

The second chapter introduces the first major theme of the thesis: the opportunity cost of tobacco consumption at the household level. This chapter investigates the categories of expenditure displaced in South African households due to tobacco expenditure. Using data from four Income and Expenditure Surveys (IESs) spanning from 1995 to 2010, it provides insights into the opportunity costs associated with cigarette consumption.

Chapter 3: Tobacco tax increases and the evolution of tobacco-induced crowding-out

The third chapter addresses the second key theme by examining changes in crowding-out patterns with increasing tobacco taxes. This analysis is set against the backdrop of South Africa's robust tobacco excise-tax policy implemented between 1995 and 2000, a period considered the “golden years of tobacco control” in the country. As cigarette prices rise, households that continue to consume cigarettes adjust their spending patterns, impacting their crowding-out behaviour.

Chapter 4: The effect of tobacco and alcohol expenditure on child health outcomes

The fourth chapter considers the third theme, focusing on the impact of parental tobacco and alcohol consumption on child health outcomes. Building on the preceding chapters’ crowding-out framework, this chapter explores whether parental tobacco and alcohol use displaces food expenditure in the home, thereby affecting child health outcomes. Using the National Income Dynamics Surveys (NIDS) panel datasets, it examines the medium- to long-term impact of alcohol and tobacco expenditure on child health.

Chapter 5: Conclusion

The final chapter concludes the thesis by summarising the research findings and contributions. It offers policy recommendations and outlines potential directions for future research.

2. CHAPTER 2: THE OPPORTUNITY COST OF TOBACCO USE IN SOUTH AFRICA

2.1 Introduction

Tobacco use, particularly cigarette smoking, significantly impacts individual health, contributing to a range of severe illnesses including cancer, cardiovascular disease, stroke, and pulmonary disease. The U.S. Surgeon General's report underscores the pervasive harm of smoking, stating that "cigarette smoking has been causally linked to diseases of nearly all organs of the body, to diminished health status, and to harm to the foetus" (U.S. Department of Health and Human Services et al., 2014). On a global scale, tobacco use was responsible for an estimated 8.7 million deaths in 2019 (Reitsma et al., 2021). Moreover, historical analyses estimate that approximately 100 million people died prematurely due to smoking throughout the 20th century (Jha, 2009).

The economic rationale for the taxation and regulation of tobacco is founded on three core principles: correcting externalities, addressing internalities, and generating revenue. Externalities due to smoking arise from two fronts. First, direct externalities accrue due to the impact of second hand smoke (SHS). For instance, of the 8.7 million deaths due to tobacco use in 2019, 1.3 million were non-smokers exposed to SHS (Reitsma et al., 2021). Second, fiscal externalities are brought on by deteriorating health status that leads to increased medical needs and reduced productivity. One study estimated that the amount of healthcare expenditure due to smoking-attributable diseases amounted to USD422 billion in 2012 or about 5.7% of global health expenditure, while productivity losses accounted for just over USD 1 trillion globally during the same period (Goodchild, Nargis & d'Espaignet, 2018).

Smokers also incur negative internalities through decisions that fail to maximise their lifetime utility. Specifically, smokers may be influenced by hyperbolic discounting, uncertainty, or asymmetric information regarding the future dangers of smoking. These factors contribute to negative spillover effects on their future selves, resulting in suboptimal demand for a longer, healthier life.

Finally, while revenue generation is one of the oldest arguments for taxation, it has not always been a frequently cited reason for taxing these products. However, this trend is shifting due to renewed interest in domestic resource mobilization and health care financing. The underlying economic principle is that, given the relatively price-inelastic demand for cigarettes, an

increase in taxes will raise prices and reduce demand. However, the reduction in consumption will not fully offset the revenue gain, as the tax rate increase is more than proportional to the decrease in demand.

The regressivity debate and opportunity cost of tobacco expenditure

Taxation is often cited as one of the most effective policy tools available to curb tobacco consumption (Chaloupka, Yurekli & Fong, 2012, US National Cancer Institute & World Health Organization, 2016). Based on established economic principles, it is argued that taxing tobacco products will lead to behavioural changes that mitigate negative externalities and internalities while also raising government revenue.

However, a common criticism of this approach is that excise taxes on tobacco, like other consumption taxes, are regressive. Additionally, since the targeted consumption is disproportionately higher among low-income populations, these taxes are seen as unfairly targeting poor or vulnerable minorities (Franks et al., 2007, Remler, 2004). Indeed, Van Walbeek (2002) and Koch (2018) both find cigarette taxes to be regressive in South Africa.

Ross and Chaloupka (2006) argue that the distributional impacts of tobacco taxes must be considered alongside their welfare effects. Even if excise taxes on tobacco are regressive, tax increases may not be, because the poor are more price-sensitive and thus reduce their consumption by a greater proportion in response to a tax-induced price increase compared to the rich. Several studies support this argument, finding that price elasticities are highest among the poorest South African households (Koch & Tshiswaka-Kashalala, 2008, van Walbeek, 2005).

Bosch and Koch (2014) examined tax regressivity before and after a tax increase using the 2005-06 Income and Expenditure Survey (IES), noting that while tobacco taxes are regressive, the tax burden on lower-income groups is reduced after a tax increase. More recently, Koch (2018) uses both the 2005-06 and 2010-11 IESs to investigate this phenomenon. He found a 30% increase in taxes over the period, noting that, although taxes were regressive, there was a reduction in the level of regressivity over the period.

One way of thinking about the effect of tobacco expenditure on households is in the context of opportunity cost. In economics, opportunity cost is pivotal in guiding decision-making across

domains by unravelling consumption patterns and resource allocation. Economists grapple with trade-offs, where choosing one option means giving up others. This concept is particularly relevant in consumer behaviour, where choices entail distinct opportunity costs.

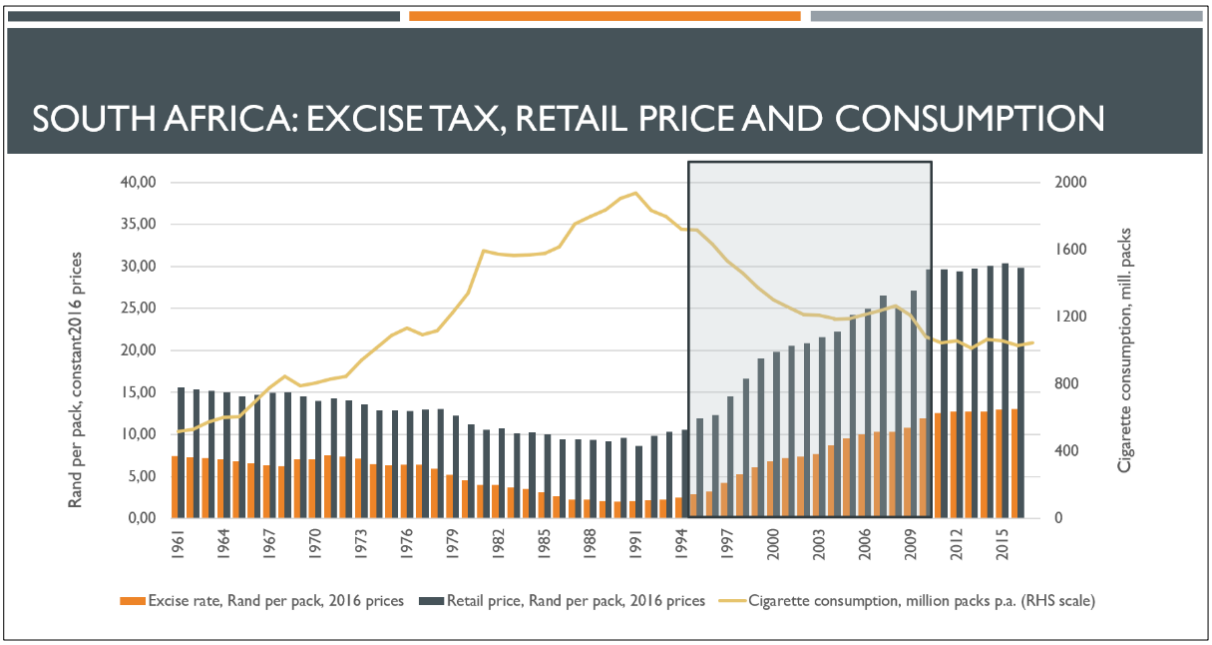
The notion of the opportunity cost associated with tobacco consumption at household level was made popular by the seminal work of Efroymson et al. (2001) on Bangladeshi households. Understanding these trade-offs is particularly relevant in the context of policy interventions, such as tobacco taxation, which influence household spending patterns. Studies indicate that despite tax hikes, some economically disadvantaged smokers persist in their habits (Adda & Cornaglia, 2010, Evans & Farrelly, 1998, Farrelly et al., 2005), exacerbating financial strain within households.

A brief history of tobacco excise taxation in South Africa

From the mid-1990s onwards, South Africa's tobacco-control strategy was widely regarded as exceptional among low- and middle-income countries. In particular, the use of excise taxes to curb consumption proved highly successful. In 1994, the government announced a target total tax burden (excise tax and VAT) on cigarettes of 50% of the retail sales price. At that time, the combined tax burden was only 33% of the retail price. The target was achieved in 1997, and subsequent tax increases were aligned with the inflation rate to maintain the 50% threshold. In 2006, this tax threshold was adjusted to 52% of the retail sales price of the most popular brand and has been maintained since then (Fuchs, Del Carmen & Mukong, 2018; Linegar & Van Walbeek, 2018).

This rapid increase in cigarette taxes, and prices, reduced smoking prevalence from about a third in 1994 to less than a fifth in 2012 (Mukong & Tingum, 2018). Chelwa, Van Walbeek and Blecher (2016), using a synthetic control method, estimate that by 2004 per capita cigarette consumption was 36% less than it would have been without the aggressive tax policy implemented in the period. Figure 2.1 shows the evolution of the excise tax changes, changes in the real price of cigarettes, and the impact these had on the consumption of cigarettes over the period before the 1994 tax changes began, as well as in the period afterwards.

Figure 2.1: Tobacco excise taxes, cigarette prices and consumption in South Africa between 1995 and 2010



Building on the preceding discussion, this chapter seeks to address two key questions. First, ”Which categories of household expenditure are displaced by tobacco expenditure in South Africa?”, and second, “How have these patterns of expenditure changed over time?”

To answer these questions, I turn to four income and expenditure surveys conducted between 1995 and 2010. Analysing this issue across these four surveys allows for a robust and comprehensive understanding of the displacement effects of tobacco expenditure on household spending. Each survey provides a snapshot of household behaviour at different points in time, enabling the examination of trends and changes over an extended period. This long-term perspective is essential for identifying persistent patterns and the impact of policy interventions, such as tobacco taxation, over time.

However, it is important to recognise that the analysis may not fully capture all contextual factors influencing household behaviour, such as changes in economic conditions, social norms, or other concurrent policy measures. Additionally, while the surveys provide valuable quantitative data, they may not offer deep qualitative insights into the motivations and decision-making processes of individual households.

Importantly, as these are cross-sectional surveys, making them comparable can be challenging. Fortunately, earlier attempts by Yu, D. (2008) and Yu, D. (2016) provide a useful starting point for making them more comparable and I discuss these extensively in subsequent sections.

The next section will introduce the relevant literature, highlighting why this remains an area of interest for further research. Section 3 will focus on the theoretical framework, while Section 4 discusses the empirical strategy. Section 5 presents a discussion of the data sets used in this chapter, highlighting the known issues with the data and how they were addressed to make the surveys comparable. In Section 6, the descriptive statistics are presented, followed by the main results in Section 7.

2.2. Relevant literature

The literature on household-level tobacco expenditure and its impacts provides a foundational understanding of the economic consequences of smoking on resource allocation. Early studies, such as Efroymson et al. (2001)'s work on Bangladeshi households and Thomson et al. (2002) work in New Zealand framed this issue within the 'crowding-out' paradigm. They posited that tobacco spending reduces household allocation to essential goods, especially in low-income settings. These studies, though largely descriptive, introduced a compelling narrative that linked tobacco spending to reductions in welfare and inspired subsequent work that increasingly applied econometric rigor to quantify these effects. However, the absence of price effects in these early models limits the generalisability and economic interpretability of findings, particularly as economic intuition around demand allocation generally relies on price-based substitution and income effects.

The assumption that tobacco expenditure alone captures cross-price elasticity effects is particularly tenuous. Without explicit price data, interpreting tobacco's share of household expenditure as capturing substitution patterns is problematic. Cross-price elasticities inherently require comparative price information to reveal substitution or complementarity across goods, a relationship overlooked when only expenditure shares are analysed. Thus, interpreting crowding-out as a quasi-cross-price effect oversimplifies the underlying substitution dynamics, as no cross-elasticity information allows us to confirm tobacco expenditure's unique role relative to other discretionary expenses.

Later studies, such as those by Busch et al. (2004) and Wang, Sindelar & Busch (2006), addressed this shortcoming by incorporating cigarette prices and estimating demand systems, thereby offering a more nuanced view of household budget constraints and potential crowding-

out effects. Busch et al. (2004) used regional and state cigarette prices in the United States between 1995 and 2001 to estimate a demand system, analysing how higher cigarette prices impact household expenditure patterns (Deaton & Muellbauer, 1980). Their study linked higher cigarette prices to adjustments in housing expenditure, suggesting that price changes influence household allocations more broadly. However, these analyses are predicated on functional forms such as the Almost Ideal Demand System (AIDS), where the chosen form heavily dictates results. If the functional form is incorrect, estimated expenditure shares and elasticities may misrepresent underlying preferences. Furthermore, interpreting results within a fixed-budget framework suggests implicit symmetry in cross-category adjustments; however, economic theory does not guarantee that reductions in tobacco expenditure would uniformly result in increased spending in other areas, especially as one equation must be omitted in estimation. This omission complicates the interpretation of expenditure shifts, as we cannot ascertain if positive and negative reallocations counterbalance across categories.

John (2008) and subsequent studies addressed endogeneity in tobacco expenditure using instrumental variables (IVs). John's use of a three-stage least squares (3SLS) model with Quadratic Almost Ideal Demand System or QuAIDS (Banks, Blundell & Lewbel, 1997) provides a technical advancement by accounting for contemporaneous correlation of errors across expenditure categories and addressing endogeneity in tobacco spending through instruments like the adult-sex ratio. This approach gained traction in later studies seen in Taiwan (Pu et al., 2008); South Africa (Koch & Tshiswaka-Kashalala, 2008); Zambia (Chelwa & Van Walbeek, 2014); Turkey (San & Chaloupka, 2016); India (Jumrani & Birthal, 2017); Bangladesh (Husain et al., 2018); Ghana (Masa-Ud, Chelwa & Van Walbeek, 2020); Vietnam (Nguyen & Nguyen, 2020); and Greece (Kostakis, 2020). However, the validity of IVs relies on exclusion criteria, which often remain conceptually, rather than empirically, justified. Furthermore, instruments like gender ratios may only weakly correlate with tobacco expenditure, risking bias if they do not fully satisfy the exclusion restriction.

As in the case of John (2008), studies by Pu et al. (2008), Chelwa & Van Walbeek (2014), Masa-Ud, Chelwa & Van Walbeek (2020) and Nguyen & Nguyen (2020) all use the adult-sex ratio as an instrument for tobacco expenditure. The selection of this instrument draws on the fact that, in most countries, smoking prevalence is much higher among men than women. Other instruments used in the literature include a composite of the smoking prevalence rate (Koch & Tshiswaka-Kashalala, 2008); the ratio of women (San & Chaloupka, 2016); a peer effect measure (Jumrani & Birthal, 2017); the proportion of males among adults (Husain et al., 2018);

and the adult ratio (Kostakis, 2020). Some studies, such as John, Ross & Blecher (2012) in Cambodia and Paraje & Araya (2018) in Chile could not find useful instruments, and estimated models without them. However, they maintained similar functional forms and estimation techniques.

Alternative approaches, including the reduced-form equations proposed by Block & Webb (2009b) and genetic matching methods introduced by Chelwa & Koch (2019), represent further attempts to circumvent endogeneity issues. By comparing outcome variables related to tobacco and non-tobacco expenditures under a unified set of covariates, Block and Webb provide an indirect test of crowding-out effects on welfare. Genetic matching offers another methodological innovation, avoiding reliance on exclusion criteria through non-parametric controls. Despite their strengths, both methods introduce interpretational limitations: reduced-form analyses risk conflating direct effects with omitted variables, while matching approaches assume that matched characteristics fully account for unobserved confounders, an assumption often hard to verify empirically.

Most recently, Nyagwachi, Chelwa & van Walbeek (2020) utilise a natural experiment framework, applying matched difference-in-difference methods to estimate crowding-out effects under rising tobacco taxation in Kenya. By exploiting exogenous price changes due to tax policy, they align the analysis closer to an ideal experimental design, controlling for endogeneity driven by time-invariant unobservables. However, even with refined methods, the interpretive limitations of fixed budgets and functional forms remain. The estimated parameters continue to depend on underlying assumptions about consumer substitution effects, which, without explicit cross-price elasticities, remain speculative.

Nevertheless, this study builds on previous econometric methodologies introduced by John (2008), and expounded in John et al. (2019) despite the constraints of not having direct price elasticity data. The absence of explicit price data means that tobacco expenditure is treated as the primary variable influencing shifts in other household spending categories, implicitly assuming rigidity in budget allocation. This approach may capture general crowding-out trends but cannot address the nuanced substitution effects that cross-price elasticities would reveal.

It applies these techniques within the South African context to explore how sustained tobacco tax increases may impact household spending over time. By using data from four Income and Expenditure Surveys (IESs) collected between 1994 and 2010, this study aligns with previous work in regions like the US and Kenya (Busch et al., 2004, Chelwa & Koch, 2019, Nyagwachi,

Chelwa & van Walbeek, 2020). It provides valuable insights into the effects of tobacco-related spending patterns on households. In this sense, while it shares the limitations found in previous research, this study contributes additional empirical evidence that enhances our understanding of crowding-out effects, especially in the context of long-term tobacco tax policy in South Africa.

2.3. Theoretical framework

In line with other studies in this literature (Chelwa & Van Walbeek, 2014, John, Ross & Blecher, 2012, Paraje & Araya, 2018, San & Chaloupka, 2016), I assume that a household seeks to maximise a single utility function subject to a budget constraint, as depicted below:

$$Max\ U = U(x_1, \dots, x_n; \boldsymbol{a}) \quad \text{s.t.} \sum_{i=1}^{n} p_i x_i = Y \tag{1}$$

Where x_i denotes consumption of the ith good, p_i denotes the price of the good and Y denotes total household expenditures, given household characteristics, $\boldsymbol{a}$. Solving this problem yields the normal unconditional Marshallian demand function of the form:

$$x_i = h_i(p_i, \dots, p_n; Y; \boldsymbol{a}) \tag{2}$$

If a household predetermines the consumption of one good, say tobacco, the function is modified to allow for this (John, 2008, Pollak, 1969). Taking tobacco as the *n*th good means that the first *n-1* goods are available in the market for the prices $(p_1, \dots, p_{n-1})$ and the total expenditure is now given by *M* (income available after tobacco expenditure). This conditional demand framework, was made precise by Pollak (1969), and results in a utility function of the form:

$$Max\ U = U(x_1, \dots, x_n; \boldsymbol{a}) \quad s.t. \sum_{i=1}^{n-1} p_i x_i = M \tag{3}$$

where M is total income net of tobacco expenditure ($M = Y - p_t t$, where $p_t t$ is tobacco expenditure).

The corresponding set of demand functions, conditional on the consumption of tobacco (in this case), are solved as follows:

$$x_i = g^{i,n}(p_i, \dots, p_{n-1}; M; \boldsymbol{a}; d) \quad (\forall\, i \neq n) \tag{4}$$

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.

Marshallian demand functions, based on equation 5 for non-smoking households are compared with those for households with at least one smoker for a common set of commodities. The variable d in this function is the coefficient of interest. In addition, a later analysis will estimate how this difference changed over time, following a period of tobacco excise tax increases.

2.4. Empirical strategy

Empirically, I first perform the standard cross-sectional analysis of crowding-out as found in most of the literature (Husain et al., 2018, John, 2008, John et al., 2019, Pu et al., 2008, San & Chaloupka, 2016). To do this, I define twelve expenditure categories commensurate with the main Classification of Individual Consumption by Purpose (COICOP) expenditure groups as follows: (1) food; (2) tobacco; (3) alcohol; (4) clothing; (5) education; (6) utilities; (7) transport; (8) communication; (9) healthcare; (10) household goods; (11) recreation; and (12) a miscellaneous category (United Nations, 2018). I then use a standard two-sample Student's t-test for the equality of means to determine whether there is a statistically significant difference in the mean budget shares between households that report expenditure on tobacco and those that do not. As this is only a rudimentary comparison of means, it could well be confounded by differences in household characteristics.

In the next step, I compare the Marshallian demand functions for the defined categories. This type of analysis is typically done within the context of a demand system that uses price and other data to estimate elasticities. In lieu of this, as is typically the case in many low- and middle-income countries, several alternative methods that estimate (conditional) Engel curves are proposed. I opted to use the QuAIDS developed by Banks, Blundell & Lewbel (1997) for this study. The Engel curves from QuAIDS introduce a quadratic income term to the well-known Almost Ideal Demand (AID) system proposed by Deaton & Muellbauer (1980). In doing this, QuAIDS allows greater flexibility of the Engel curves estimated at different income levels, while remaining consistent with utility theory. The conditional Engel curve is specified as follows:

$$w_{ij} = \alpha_{1i} + \alpha_{2i}\,(p_t t)_j + \alpha_{3i} lnM_j + \alpha_{4i}(lnM_j)^2 + \gamma_i \boldsymbol{x_j} + u_{ij} \tag{5}$$

Where w_{ij} represents the monthly adjusted expenditure share of category i for household j in the remaining budget excluding expenditure on tobacco; $(p_t t)_j$ represents the expenditure on tobacco; lnM_j is the log of total monthly expenditure minus the expenditure on tobacco for household j; $(lnM_j)^2$ represents the squared log of total monthly expenditure minus the expenditure on tobacco for household j, and $\boldsymbol{x_j}$ represents a vector of household characteristics. These will include the usual characteristics of gender, race, and age of household head, household size, ratio of adults to children and settlement-type dummies indicating urban and rural areas; u_{ij} is the error term.

According to this specification, focus is placed on how the amount spent on tobacco affects the expenditure shares of other household goods. Thus, if the coefficient α_{2i} is negative and statistically significant, then the amount spent on tobacco crowds out expenditure on commodity *i*, *ceteris paribus.* An alternative specification examines the difference in expenditure shares between households that use tobacco and those that do not, but replaces the expenditure amount, $(p_t t)_j$, with a tobacco use dummy, (d), as follows:

$$w_{ij} = \alpha_{1i} + \alpha_{2i}(d)_j + \alpha_{3i} lnM_j + \alpha_{4i}(lnM_j)^2 + \gamma_i \boldsymbol{x_j} + u_{ij} \qquad (6)$$

In this case, the focus is not on the amount of money that is spent on tobacco, but rather whether the household spends money on tobacco or not. A common challenge in large household surveys is how to deal with zero expenditures for a particular commodity. Generally, zero expenditures arise, due to either corner solutions or abstention. A corner solution occurs if a consumer is unable to afford tobacco at its current price and at his or her income levels. If, however, the price fell or the consumer's income were to increase, the non-consumer may decide to purchase the commodity. Abstention, in contrast, implies that consumers and non-consumers of tobacco have different preferences. Thus, for the non-consumer, tobacco is not an argument in the direct utility function; and that person or household would therefore not use tobacco, even if it were available for free. Vermeulen (2003) developed a simple test to determine whether zero expenditures can be found to be due to corner solutions or to abstention. To do this, the QuAIDS model in Equation 5 above is augmented by introducing a binary variable, d, representing smoking status as follows:

$$w_{ij} = (\alpha_{1i} + \alpha_{2i} d_j + \alpha_{3i} (p_t t)_j + \gamma_i \boldsymbol{x_j}) + (\beta_{1i} + \beta_{2i} d_j) lnM_j + (\zeta_{1i} + \zeta_{2i} d_j)(lnM_j)^2 + u_{ij} \quad (7)$$

From this equation, a test of the joint significance of the coefficients associated with the binary variable indicates the existence of heterogeneous behaviours between tobacco users and non-users. The test implies a null hypothesis that the coefficients associated with d_j are jointly zero, i.e. $H_0: \alpha_{2i} = \beta_{2i} = \zeta_{2i} = 0$. A rejection of this null hypothesis means that one may conclude that households that record zero expenditures on tobacco only do so because it is unaffordable. The test is also akin to a weak separability test of consumption of tobacco from other commodities. Separability implies that consumption of tobacco only generates an income effect, with no substitution effect.

Following the recommendations of John et al. (2019), the specification in equation 7 is run first. If the parameters associated withd_j are jointly significant, this indicates that the specification is appropriate. Conversely, if they are not jointly significant, equation 5 is preferred. For diagnostic purposes, robust standard errors are applied throughout the analysis to account for heteroskedasticity, which is commonly present in cross-sectional studies (Deaton, 2018).

In addition, all three specifications (equations 5, 6, and 7) are likely subject to endogeneity issues, specifically concerning the regressors (d), $(p_t t)_j$, lnM_j and $(lnM_j)^2$ due to potential simultaneity(John, 2008). As a result, estimation of the model via OLS would likely yield biased estimates. To address this, the present study tests the exogeneity of these variables using the endog test option within Stata's ivreg2 implementation. Given that robust standard errors are already used, the endog test also adjusts for any heteroskedasticity, making additional tests for heteroskedasticity unnecessary. This streamlined approach ensures consistency in addressing both endogeneity and heteroskedasticity within the analysis.

To address the endogeneity problem, an instrumental variable (IV) approach is taken. I follow others (Chelwa & Van Walbeek, 2014, John, 2008) in using the adult-sex ratio as an instrument for tobacco expenditure. This is motivated by the gendered nature of smoking prevalence in South Africa, which is largely male-dominated, with current tobacco-smoking prevalence at 37% for men compared to 6.8% for women (South African Medical Research Council, 2022). I also instrument for lnM_j and $(lnM_j)^2$ using the log of household income and the squared log of household income, respectively. This is commensurate with other studies that have used this

instrument (Husain et al., 2018, Pu et al., 2008), although it may not be ideal. For instance, income variables may suffer greater measurement errors than consumption does (Lee, 2009). However, given the limitations of the current dataset, there are no more suitable alternatives available.

In this study, we estimate the IV approach using the Three-Stage Least Squares (3SLS) method, specifically implemented through Stata's reg3 command. The 3SLS method is appropriate due to the presence of contemporaneous correlation of the errors across equations, which it effectively handles by combining the Seemingly Unrelated Regression (SUR) approach with IV techniques.[2]

However, due to expected heteroskedasticity in the data, a two-step Generalised Method of Moments (GMM) estimator generally offers superior performance over traditional 3SLS. Wooldridge (2010) highlights that the efficiency of "traditional 3SLS" is contingent upon homoskedastic conditions. While the GMM estimator may exhibit suboptimal finite sample properties, asymptotically its utilization of a heteroskedasticity-consistent weighting matrix ensures at least equivalent performance compared to traditional 3SLS and surpasses it in certain contexts. GMM exhibits resilience to heteroskedasticity and enhances efficiency by minimising a weighted sum of sample moments. Recent studies suggest that iterative GMM is preferable over two-step GMM due to several advantages: it avoids arbitrary initial weighting matrix selection, simplifies robust inference against misspecification, and tends to reduce variance, thereby improving efficiency (Hansen & Lee, 2020).

Despite these advantages, practical experience during this analysis revealed significant computational challenges with GMM. While reg3 completes estimation within minutes, GMM often fails to converge. This discrepancy stems from reg3's utilization of a closed-form matrix solution for direct estimator computation, while GMM relies on a nonlinear numerical search process. In essence, reg3 exploits known algebraic properties and matrix operations, resulting in swift performance even with multiple equations. Conversely, GMM iteratively searches for optimal estimators by iteratively refining values to minimise errors, a process inherently slower and computationally demanding, particularly with an increasing number of equations.

To diagnose the convergence issues with GMM, various potential causes were examined, supplemented by the application of three different GMM estimation strategies. Firstly, model

specification was scrutinised to assess whether any assumptions regarding functional form or variable inclusion might be contributing to the convergence issues. The relevance and validity of instrumental variables were re-evaluated to ensure adherence to exclusion restrictions, using first-stage F-statistics.

Secondly, the impact of initial weighting matrix selection on convergence was investigated. The initial weighting matrix significantly influences the starting point for the optimization process in GMM. Attempts were made to enhance convergence by employing the output from reg3 as initial values for GMM, but the issue persisted.

In an effort to address heteroskedasticity while maintaining computational efficiency, two additional GMM estimation strategies were explored: direct two-step GMM and restricted iteration GMM. The former entails estimation of the model using GMM with an initial guess for parameter values, followed by refined estimation based on initial results to improve parameter estimate accuracy. This approach applies the GMM framework directly without additional constraints. In the latter attempt, GMM estimation was conducted with a restriction of 100 iterations during the optimization process, aimed at expediting convergence.

Despite these efforts, none of these strategies converged satisfactorily. Therefore, while GMM has theoretical advantages under heteroskedasticity, practical constraints led us to prefer *reg3* with bootstrapped standard errors. This method, although computationally intensive, is a compromise that delivers consistent parameter estimates and correct variances under heteroskedasticity without the excessive computational burden associated with GMM. Thus, this methodology ensures both efficiency and accuracy, providing a practical solution to the challenges posed by heteroskedasticity in simultaneous equation models.

Relaxing the exclusion criteria

The validity of the adult-sex ratio as an instrument used in this model constitutes a potential limitation, as noted from the diagnostic tests. In fact, the results from the disaggregated population groups suggest that it may not be valid at all in some settings. The shortcoming is also acknowledged by others in the literature (Chelwa & Koch, 2019). To address this, I adopt the partial identification strategy developed by Nevo & Rosen (2012). This section describes the method and aims to assess the robustness of the 3SLS results presented in the preceding section.

In contrast to the standard IV approach which focuses on point estimates, Nevo & Rosen (2012) use imperfect instrumental variables (IIV) to identify a set of parameter values. An IIV is an instrumental variable that may be correlated with the error term. The key identifying assumptions are two-fold:

1. the sign of the correlation between the IIV and the error term is the same as the correlation between the endogenous regressor and the error term; and
2. the correlation of the IIV with the error term is assumed to be less than the correlation between the endogenous variable and the error term.

These are weaker exogeneity assumptions than those in the standard IV and OLS methods. Formally, let the imperfect instrument be denoted Z, variable Y as a function of the endogenous regressor X, other covariates W and the error term, ε. Nevo & Rosen (2012) show that if the relationship between the instrumental variable and the endogenous variable is positive, only one bound can be estimated, as follows: $\beta \leq min\{\beta_{v(1)}^{IV}, \beta_z^{IV}\}$ if $\sigma_{x\varepsilon}, \sigma_{z\varepsilon} \geq 0$ or $\beta \geq max\{\beta_{v(1)}^{IV}, \beta_z^{IV}\}$ if $\sigma_{x\varepsilon}, \sigma_{z\varepsilon} \leq 0$, where $\beta_{v(1)}^{IV}$ is the probability limit of the traditional IV (2SLS) estimator for β when $v(1)$ is used as an instrument for X, β_z^{IV} is the probability limit of the standard IV estimator for β, while $\sigma_{x\varepsilon}$ and $\sigma_{z\varepsilon}$ denote the covariance of variable X and the error term, and of instrument Z and the error term, respectively. Furthermore, $v(\lambda)$ is a synthetic IV proposed by Nevo & Rosen (2012) with $v(\lambda) = \sigma_x Z - \lambda \sigma_z X$, where Z is the IIV and X is the endogenous regressor; σ_x and σ_z are respectively the standard deviations of X and Z; and λ is some parameter defined as the ratio of the correlation between Z and ε to the correlation between X and ε. On the other hand, a negative relationship between the instrumental variable and the endogenous variable would allow for the estimation of two-sided bounds. In the latter case, the bounds are given by $\beta_z^{IV} \leq \beta \leq \beta_{v(1)}^{IV}$ if $\sigma_{x\varepsilon}, \sigma_{z\varepsilon} \geq 0$ and $\beta_z^{IV} \geq \beta \geq \beta_{v(1)}^{IV}$ if $\sigma_{x\varepsilon}, \sigma_{z\varepsilon} \leq 0$, where the parameters are as defined before[3].

2.5. Data and data quality issues

Data for this study were taken from IESs conducted in 1995, 2000, 2005-2006 and 2010-2011. The IES is a quinquennial nationally representative survey conducted by Statistics South Africa (Stats SA). The primary purpose of the survey is to obtain the consumption basket used to construct the consumer price index. A two-stage stratified sampling method was used in each case, with the 1991 population census serving as sampling frame for the 1995/2000 IESs and

[3] The exposition presented here draws extensively from the doctoral thesis by Chelwa (2015).

the 2001 population census serving as the sampling frame for the 2005-2006 and 2010-2011 IESs.

Although the IES datasets present a rich panoply of geographic, demographic, and socioeconomic data in addition to the expenditure data, several data problems have been cited (Ardington et al., 2006, Kerr & Wittenberg, 2015, Simkins, 2004, Wittenberg, 2008, Wittenberg, 2017a, Wittenberg, 2017b). Some of the pertinent data quality issues are highlighted here, with their possible implications for the analysis and interpretation of results.

2.5.1. Changes in the measurement instrument

In order to improve the quality of data collected in the IESs, Stats SA made several changes to the data collection instrument between 1995 and 2010-2011. Broadly summarised, these changes to the instrument encompass both changes to the survey questionnaire and changes in fieldwork implementation. Although only a summary is presented here, a detailed account of many of these changes and their implications for comparisons across surveys can be found elsewhere (Yu, D., 2008, Yu, D., 2016).

One of the more significant changes in terms of the questionnaire was in the taxonomy of expenditure categories used. This changed from the Standard International Trade Classification (SITC) to the Classification of Individual Consumption by Purpose (COICOP) expenditure groups.[4] COICOP is the internationally accepted reference classification of household expenditures developed by the United Nations Statistics Division, while the SITC, also developed by the same institution, is primarily used for the economic analysis of traded commodities (United Nations, 2018). These changes affect the type of expenditure items that are included in various categories, as well as how they are captured. In order to make reasonable comparisons across the surveys, Yu, D. (2008) suggests one of two methods, viz., either re-categorise the income and expenditure items prior to 2005/06 using the 2005/06 COICOP structure or, conversely, re-categorise the income and expenditure items post-2000 using the

[4] Note that this refers to the COICOP 1999 version. A more recent revision took place in 2018 that set out to modernise the classification, considering new consumption patterns and removing obsolete classifications. (See United Nations. 2018. *Classification of Individual Consumption According to Purpose (COICOP) 2018*. New York.)

SITC.[5] The COICOP is more fit for purpose in this study, thus the first approach (changing the SITC to COICOP) is adopted, resulting in twelve expenditure categories as defined earlier.[6]

Turning to fieldwork implementation, a big change here concerns the difference in data collection periods across surveys. For the 1995 and 2000 IESs, data were collected from households over a period of about a month, coinciding with the October Household Survey (OHS) of the same year, while data for the 2005-2006 and 2010-2011 IESs were collected over the period of a year. Furthermore, the method of data collection changed from a recall-only approach in 1995 and 2000, to both diary and recall methods from 2005 onwards.

For the latter two surveys, expenditure or income values were obtained from both diary entries and recall interviews. When both instruments were utilised, the total amount was recorded. Conversely, if the information was sourced from either the main questionnaire or the diary alone, the respective values from the recall variable or the diary were transferred to the final value column.

The two earlier surveys may have measurement error due to potential recall biases related to of the data collection method applied. The use of the diary method in the latter surveys may have reduced potential recall biases present in the earlier surveys (Yu, D., 2008, Yu, D., 2016). However, it is still possible that the survey may suffer from measurement error associated with the failure to record small purchases (Neter & Waksberg, 1964).

In the latter two surveys, i.e. 2005/06 and 2010/11, the income/expenditure on an item was collected through both the diary and main questionnaire instruments wherever possible. These amounts were simply added together to reflect the total amount spent. Where data was collected from only one instrument, the values were transferred to the value column to represent the total amount spent.

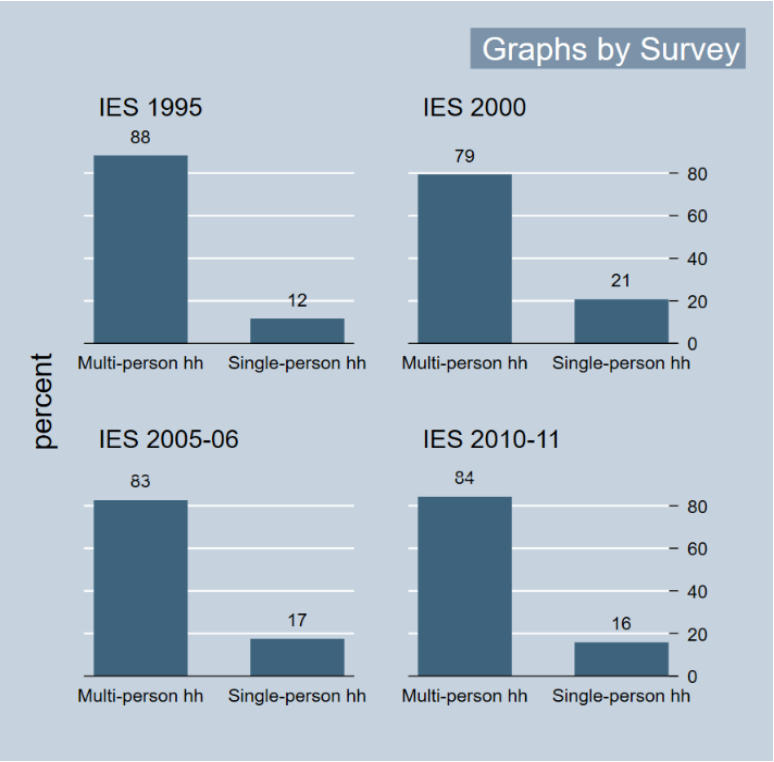
2.5.2. Changes in sampling procedures

The changes described in the preceding section are broadly measurement related. However, a more complicated modification occurred between 1995 and 2000, with respect to the sampling procedures used. Kerr & Wittenberg (2015) note that the pre-1999 OHSs used a sampling method that significantly under-sampled households at multiple dwelling units, especially

[5] Yu (2016) suggests excluding some sub-categories (*imputed rent (04211000), imputed rent 7% per year of dwelling (04211001)* and *cost of other dwelling)* from the aggregations on housing expenditures across surveys to ensure uniformity of both 2005-2006 and 2010-2011 IESs.
[6] Recall that these were food, tobacco, alcohol, clothing, education, utilities, transport, communication, healthcare, household goods, recreation, and a miscellaneous category.

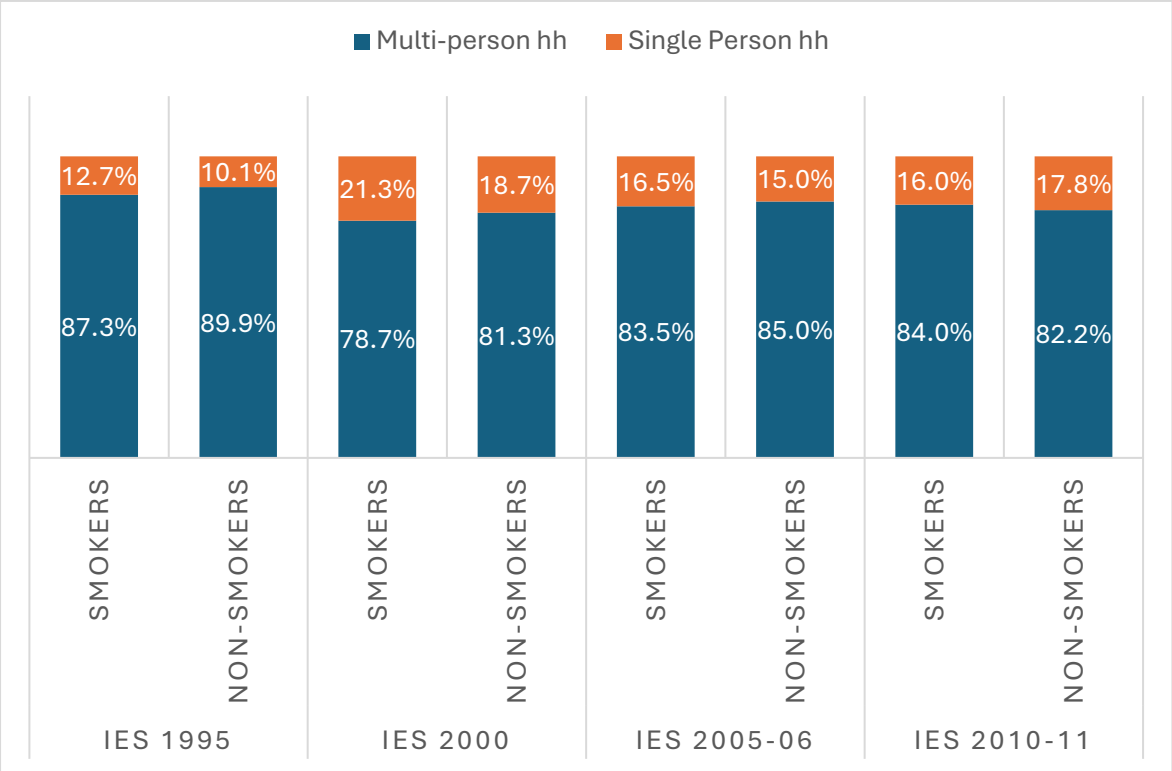
small households, without any commensurate increase in weights for such households in the final dataset. Since the 1995 OHS and IES were conducted simultaneously, it is impossible to separate most of the issues in the IES from those in the OHS. By way of contrast, subsequent surveys reportedly interviewed all households at multiple dwelling units. This means that the distribution of households in the 1995 IES is fundamentally different from those of subsequent IESs. The choice of which datasets to use is thus a non-trivial endeavour, with the risk of possible distortions in the trends proposed for analysis in subsequent chapters.

Figure 2.2: Percent of single-person households across surveys



A look at the weighted proportion of single- versus multi-person households, shown in Figure 2.2, reveals a substantial difference across surveys. The 1995 IES reports only 12% of households as single-person households, compared to 21%, 17%, and 16% in subsequent surveys. This probably confirms the under-sampling of single-person households in 1995, as mentioned earlier. Further disaggregation of these proportions by tobacco-consuming and non-consuming households reveals a similar pattern (see Figure 2.3). Machemedze, Kerr & Wittenberg (2014) suggest a recalibration of 1995 OHS weights, using the 1996 census, as one possible solution, but admit that this exercise has mixed results. In this chapter, it is difficult to determine the overall effect this difference may have on the models to be estimated, and it is thus acknowledged to be a likely limitation.

Figure 2.3: Proportion of multi-person and single-person households among smokers and non-smokers



Related to this problem is the issue of non-response. Stats SA admits that the refusal rate for the IES is often higher than the refusal rate in other household surveys. High income households are the least responsive. However, across IESs, the response rates are relatively similar. For instance, the 2000 IES reported an overall response rate of 87.6%, while the 2005-2006 and 2010-2011 IESs reported 89.8% and 91.6%, respectively. Differences in response rates across regions were fairly stark in some cases. Notably, Gauteng Province was most affected, with a reported response rate of 82.9% in 2010-2011, compared to 95.3% in the 2005-2006 IES (Statistics South Africa, 2017). The implication is that, since Gauteng province alone accounts for over a third of all household expenditures in the country, some underreporting is expected on various areas of expenditure, such as food, beverages, and tobacco.

2.5.3. Outliers and influential data points

A final issue concerns the presence of outliers and influential data points, particularly for income and expenditure variables in the IES datasets. This is a well-documented issue of some importance (see for instance Greene (2000)), not least because such observations may arise from a different data-generating process than the rest of the data. These outliers may reflect

fundamentally different underlying dynamics, potentially stemming from unique economic behaviours or atypical income sources that are not representative of the broader population.

The inclusion of such extreme values in the analysis can bias the estimation of relationships between income and expenditure, as well as distort the observed distribution of spending across household budget categories. This distortion has implications not only for total expenditure patterns but also for the allocation of expenditures across categories (e.g., the proportion of income allocated to luxury versus essential goods). Such biases undermine the generalizability of the findings, complicating efforts to draw conclusions applicable to the wider population.

To mitigate these risks, this study identifies and appropriately addresses outliers to produce a dataset that more accurately represents typical household expenditure relationships. This approach enhances the reliability and external validity of the results, ensuring that the conclusions drawn are robust and applicable to the general population. The obvious challenge is two-fold, namely: (1) how to detect outliers, and (2) what to do about them.

As expenditure is traditionally regarded as a function of income, I have chosen to focus the outlier analysis exclusively on the income variables within each survey. First, all "millionaires" are identified in the data by flagging households that earn more than R83,334 a month. This follows Burger & Yu (2007), who highlighted the importance of a small cohort of "millionaires" in their analysis of OHS and Labour Force Survey (LFS) data between 1995 and 2005. However, this only flags extreme values at the high end of the distribution, and not at the low end. Two additional strategies are employed to address this. First, a mean-variance test is used to flag observations as outliers if they are more than three times the standard deviation from the mean. Assuming a normal distribution, a reasonable expectation would be that about 99.87% of observations would be non-outliers. Conversely, outliers would constitute about 0.13% of observations. The natural log of income closely approximates a normal distribution and is used in this case. The method identifies observations on either end of the distribution and is popular in the literature, but has some weaknesses (Leys et al., 2013). For instance, the calculation of the mean and the standard deviation is not impervious to outliers, where the calculation of the mean and the standard deviation is itself influenced by the presence of outliers. Another limitation arises if the underlying distribution is non-normal or heavily skewed.

An increasingly common alternative is the median absolute deviation (MAD) from the median made popular by Hampel (1974) (see Huber (2004) for a textbook treatment). This method is

akin to the previous one, but replaces the mean with the median and the standard deviation with the MAD. The median is insensitive to the presence of outliers while the MAD is immune to sample size variations and outliers, making this method more robust. Further, it uses a similar decision criterion, where observations are flagged as outliers if they are more than three times the MAD from the median.

Table 2.1: Identification of income outliers across IESs

	1995		2000		2005-2006		2010-2011	
	N	**%**	**N**	**%**	**N**	**%**	**N**	**%**
Original sample	29,595	100	26,265	100	21,114	100	25,328	100
Millionaires (earn >R83,334 p.m.)	171	0.58	43	0.16	114	0.54	428	1.69
Flag 1 (>3SD from mean)	100	0.34	133	0.51	117	0.55	495	1.95
Flag 2 (>3MAD from median)	68	0.23	186	0.71	318	1.51	534	2.11
Flag 1 & Flag 2	**68**	**0.23**	**133**	**0.51**	**117**	**0.55**	**495**	**1.95**
Final sample	**29,514**	**99.80**	**26,132**	**99.50**	**21,027**	**99.50**	**24,833**	**98.00**

Using these methods, and triangulating the observations flagged as outliers, gives the final set as shown in Table 2.1 above. The IES 1995 only flagged 68 observations (0.23% of the overall sample), while nearly 500 observations (1.95% of the overall sample) were flagged as outliers in the 2010-2011 IES. Table 2.1 also shows the final sample sizes used in each of the surveys. In each case, fewer than 2% of the original sample were removed after being flagged as outliers. The removal of these outliers results in a reduced dispersion from the mean in the form of smaller standard deviations and overall smoother distributions (see Table 2.2 below).

Table 2.2: Impact of outliers on income

Variable	**Mean**	**Median**	**Standard Deviation**	**Minimum**	**Maximum**
1995 IES (outliers incl.)	8,551	3,874	15,756	84	1,039,725
1995 IES (outliers excl.)	8,110	3,859	11,340	126	128,614
2000 IES (outliers incl.)	5,948	2,418	12,111	2	743,851
2000 IES (outliers excl.)	5,677	2,429	8,796	76	78,879
2005 - 06 IES (outliers incl.)	9,017	3,131	19,399	0	613,897
2005 - 06 IES (outliers excl.)	8,171	3,128	13,321	90	109,273
2010 - 11 IES (outliers incl.)	10,259	3,812	16,919	0	341,999
2010 - 11 IES (outliers excl.)	10,319	3,877	16,748	77	178,487

2.6. Descriptive data

Having addressed the data issues, attention is now given to the descriptive statistics of the final datasets used. Table 2.3 gives the relevant descriptive statistics across all four IESs. In terms of the prevalence of tobacco consumption, the proportion of tobacco-consuming households was highest in the 1995 IES, 46.7%, followed by 37.4% in 2005-06, 34% in 2000 and 24% in 2010-11. This shows an overall decline in tobacco consumption over the period between 1995 and 2010, closely mimicking the individual-level smoking prevalence reported in the country for the same period (Reddy et al., 2015, van Walbeek, 2004, Vellios & Van Walbeek, 2016). Most tobacco consumption was reported in urban areas in all surveys. The lowest proportion of urban tobacco consumers was in 1995 (57.8%), while the highest was reported in 2010-11 (74.7%).

The general sociodemographic trends are not contentious, and are generally consistent with what others observe in the literature. For instance, Wittenberg, Collinson & Harris (2017) report similar reductions in household sizes between 1992 and 2012 using other datasets. Reductions observed in the average number of adults and children likely reflects a link with the overall household size. The average age of children was between 6.9 years and 7.5 years of age, while the average age among adults in a household ranged from 37.3 years to 38.6 years. Turning to the racial profile of the household heads, this again is broadly consistent with the country's demographic profile and is uncontroversial. Significant variations in the proportion of male-headed versus female-headed households were also noted and mimic the largely male-dominated smoking prevalence patterns in South Africa.

Expenditure data were collected using only the recall method in 1995 and 2000, while the 2005-06 and 2010-11 surveys used both recall and diary methods. The reference periods for these expenditures differed, as some were monthly, while others were annual. In the latter case, expenditures were divided by 12 to obtain monthly figures. As noted earlier, the focus of this chapter is on twelve categories, namely food, tobacco, alcohol, clothes and shoes, education, housing and household maintenance, transport, communication, health, household goods and services, recreation, and a miscellaneous category. Budget shares for these categories were calculated as a proportion of the total expenditures in the household.[7]

[7] Total household expenditures were obtained by simply summing up the expenditure categories. All missing values were converted to zero to limit the number of observations dropped from the analysis.

Table 2.4 highlights the budget shares for each category across all four surveys and disaggregated by tobacco-consumption status. On average, the largest share of the household budget was dedicated to food and housing. Food shares accounted for no less than a quarter of the household budget and in some cases almost half the household budget, regardless of tobacco-consumption status. Rural households expended more of their income on food than urban households. Conversely, wealthier households spent a smaller proportion of their income on food than poorer ones. These results were consistent across surveys.

Among households that reported tobacco consumption, tobacco represented a significant portion of their total household expenditure over the years. In 1995, it accounted for approximately 3.2% of total household expenditure, which increased to 5.5% in 2000, then decreased to 3.6% in the 2005-06 period, and finally reached 4% in 2010-11. It is worth noting that, in general, urban households consistently reported slightly higher proportions of their expenditures that were allocated to tobacco, compared to rural households. However, there was an exception in the 2005-06 survey, where the opposite trend was observed. Additionally, it is important to highlight that low-income households consistently allocated significantly higher budget shares to tobacco consumption when compared to their high-income counterparts. In some cases, the budget shares reported by low-income households were more than double those reported by high-income households.

Some surprising fluctuations or increases in budget shares were also noted in a few categories, specifically food, housing, and transport. Examining the comparability of various estimates from the 1995, 2000 and 2005-06 IESs, Yu, D. (2008) finds similar sudden changes. He notes that the precipitous rise in transport costs in the 2005-06 IES were driven by increased purchases of new and used motor cars, station wagons, minibuses, bakkies, four-wheel-drive vehicles, and fuel. Furthermore, the decrease in food expenditures between 2000 and 2005-06 was underpinned by decreased purchases in food items, such as white bread, brown bread, whole wheat bread, mealie meal and maize flour, fresh full cream milk, fresh low-fat milk, onions, and white sugar.

A breakdown of the expenditure shares of the three categories (food, housing, and transport) is provided in Tables A3, A4 and A5. Decreases in food expenditure shares appear to be driven by reductions in the shares of sub-categories, such as coffee and tea, fish, fruits, oils and fats, sugar, vegetables, and unclassified food expenses. On the other hand, in the transport category, while there was an initial drop in the expenditure shares spent on motor vehicles between 1995

and 2000, there was a consistent increase in budget shares for this sub-category for the subsequent surveys. Finally, housing expenditure sub-categories present some noisy fluctuations. A case in point is the *actual rent* budget share sub-category, where a general increase is noted between 1995 and 2005-06, but the share drops off quickly in the 2010-11 survey. Similarly, the *imputed rent* share was 35.4% in 1995, after which it suddenly drops to 2.1% in 2000, before shooting up to 54.8% and 66.4% in 2005-06 and 2010-11. These and other drastic differences can only be partially explained by some of the data quality issues raised earlier, and reinforce the need for a cautious interpretation of the results.

Table 2.3: Descriptive statistics from the income and expenditure surveys (IES) 1995, 2000, 2005-06 and 2010-11

	IES 1995			IES 2000			IES 2005-06			IES 2010-11		
Variable	**Tob**	**No Tob**	**Full 1995**	**Tob**	**No Tob**	**Full 2000**	**Tob**	**No Tob**	**Full 2005-06**	**Tob**	**No Tob**	**Full 2010-11**
Number of households	8,751	9,546	18,297	9,181	16,906	26,087	8,157	12,852	21,009	6,030	18,803	24,833
Tobacco consuming households (%)	100.0	0.0	46.7	100.0	0.0	34.0	100.0	0.0	37.4	100.0	0.0	23.6
Non-tobacco consuming households (%)	0.0	100.0	53.3	0.0	100.0	66.0	0.0	100.0	62.6	0.0	100.0	76.4
Households in rural areas (%)	42.2	45.7	44.1	30.3	35.6	33.8	31.7	37.1	35.1	25.3	35.1	32.8
Households in urban areas (%)	57.8	54.3	55.9	69.7	64.4	66.2	68.3	62.9	64.9	74.7	64.9	67.2
Mean monthly tobacco expenditure*	150.1	0.0	70.2	173.6	0.0	59.1	116.8	0.0	43.6	178.0	0.0	42.0
Mean monthly household expenditure*	7,112.5	7,396.2	7,263.6	4,860.9	4,193.1	4,420.4	5,230.6	6,117.0	5,785.8	7,604.7	7,762.7	7,725.5
Mean household size	4.5	4.3	4.4	3.9	3.9	3.9	3.9	3.8	3.8	3.7	3.9	3.9
Mean number of kids	1.4	1.5	1.4	1.1	1.3	1.3	1.2	1.3	1.2	1.0	1.3	1.2
Mean number of adults	3.1	2.8	3.0	2.8	2.5	2.6	2.9	2.6	2.7	2.9	2.8	2.8
Mean age of kids#	7.5	7.6	7.5	7.3	7.4	7.4	0.0	0.0	0.0	6.9	7.0	7.0
Mean age of adults#	37.5	38.6	38.1	37.7	37.1	37.3	0.0	0.0	0.0	38.3	37.5	37.7
African-headed households (%)	66.2	73.4	70.0	70.2	82.4	78.3	72.5	80.0	77.2	64.1	80.4	76.6
Coloured-headed households (%)	12.3	4.7	8.3	14.0	5.0	8.1	12.8	4.8	7.8	18.6	5.4	8.5
Indian-headed households (%)	2.9	2.4	2.6	2.8	2.2	2.4	2.9	2.3	2.5	2.5	2.6	2.6
White-headed households (%)	18.7	19.4	19.1	12.9	10.3	11.2	11.9	12.9	12.5	14.8	11.6	12.3
Female-headed households (%)	20.0	41.1	31.3	41.9	52.8	49.1	41.7	50.5	47.2	27.2	43.0	39.3
Male-headed households (%)	80.0	58.9	68.7	58.1	47.2	50.9	58.3	49.5	52.8	72.8	57.0	60.7

*Expenditures prior to the IES 2010/2011 were all inflated to March 2011. # Age in the 2005-06 IES was captured as a categorical variable, hence no mean was calculated.

Table 2.4: Mean budget shares from the income and expenditure surveys (IES) 1995, 2000, 2005-06 and 2010-11 disaggregated by tobacco use.

	Full sample				Tobacco user				Non-tobacco users			
Category	**IES 1995**	**IES 2000**	**IES 2005-06**	**IES 2010-11**	**IES 1995**	**IES 2000**	**IES 2005-06**	**IES 2010-11**	**IES 1995**	**IES 2000**	**IES 2005-06**	**IES 2010-11**
Food	29.5%	45.4%	27.6%	25.2%	29.1%	42.2%	28.2%	24.7%	29.8%	47.0%	27.2%	25.4%
Tobacco	1.5%	1.9%	1.3%	0.9%	3.2%	5.5%	3.6%	4.0%	0.0%	0.0%	0.0%	0.0%
Alcohol	1.1%	2.0%	0.8%	1.1%	1.7%	4.1%	1.6%	3.1%	0.6%	0.9%	0.3%	0.4%
Clothes & shoes	6.2%	6.7%	8.2%	7.3%	6.2%	6.3%	8.2%	6.7%	6.1%	6.9%	8.2%	7.5%
Education	0.9%	2.1%	1.8%	1.7%	0.8%	1.5%	1.3%	1.3%	1.0%	2.4%	2.1%	1.8%
Housing	37.6%	17.3%	21.4%	27.3%	36.4%	17.2%	20.5%	25.9%	38.7%	17.4%	21.9%	27.8%
Transport	3.4%	7.4%	11.7%	12.8%	3.3%	6.9%	11.0%	11.6%	3.5%	7.7%	12.2%	13.2%
Communication	0.2%	1.9%	3.8%	3.4%	0.1%	1.7%	3.5%	3.2%	0.2%	2.0%	3.9%	3.5%
Health	2.0%	1.1%	1.7%	1.5%	2.0%	1.0%	1.7%	1.5%	2.0%	1.2%	1.7%	1.5%
Household goods	4.3%	2.9%	7.5%	5.5%	4.1%	2.9%	7.1%	5.2%	4.4%	2.9%	7.8%	5.6%
Recreation	1.5%	1.9%	2.1%	1.4%	1.6%	1.8%	2.1%	1.5%	1.5%	1.9%	2.1%	1.4%
Miscellaneous	11.9%	9.4%	12.1%	11.9%	11.6%	9.0%	11.1%	11.2%	12.2%	9.6%	12.6%	12.1%

2.7. Empirical results

The empirical results are presented in two parts. The results from the analysis of differences in expenditure shares are presented first, followed by the results of the 3SLS regression procedures conducted on each cross-sectional survey dataset.

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

The outcomes of the analysis of differences in budget shares across all the surveys are presented in Table 2.5 below. In this analysis, I investigate variations in budget allocations between households with smokers and those without. A positive and statistically significant result in the table suggests the presence of crowding-out, while a negative and statistically significant result indicates crowding-in. These results shed light on the overall, though somewhat simplified, effects of tobacco spending at the household level. Nonetheless, it is crucial to recognise that these effects are meaningful primarily in the context of how they impact individuals within these households. Consequently, the results are framed in terms of their potential influence on nutrition, which is reflected in reduced expenditure on food; on human capital, as indicated by diminished shares allocated to education and healthcare; and on intra-family well-being, as evidenced by expenditures related to clothing, housing, household goods, transportation, communication, and recreation.

In terms of the impact on food expenditure, most households that reported tobacco expenditure spent less on food than did non-tobacco-spending households. The differences in the budget share for food were calculated to be 0.7, 4.9, and 0.7 percentage points in 1995, 2000 and 2010-11. Access to nutritious food, particularly during childhood, are both directly and indirectly linked to good health (Block & Webb, 2009b, Efroymson et al., 2001, Kennedy & Peters, 1992). Access to nutritious food in turn has indirect long-term and intergenerational impacts on education and earnings (Bütikofer, Mølland & Salvanes, 2018). The notable exception is the 2005-06 survey, where the food expenditure share is higher among households that report expenditure on tobacco. Similar inconsistent patterns of food expenditure have been observed in the literature. For example, in their examination of 40 low- and middle-income countries, Do & Bautista (2015) find that households that report daily tobacco use saw increased food expenditure. Another study in Bangladesh recently found that food expenditures were higher among households that purchased tobacco (Husain et al., 2018).

The human capital effect of tobacco expenditure can be gleaned from its influence on education and health care expenditures. Consistent with findings from the literature, tobacco-spending

households spent less on education than their non-tobacco spending counterparts. The difference was as low as 0.2 percentage points for the 1995 IES, and as high as 0.9 percentage points for the 2000 survey. This suggests reduced investment in education among households that report tobacco expenditure. As noted earlier, this has the potential to negatively impact earnings and increase poverty in the long term (Currie & Almond, 2011).

Expenditure on healthcare forms the other part of the human capital effect. The table shows that only the 2000 survey finds a statistically significant difference in budget shares on health. The difference is only 0.1 percentage points. This may not be a surprising result, for at least two reasons. First, the health impact of tobacco use is only seen over several years, and may not be easily identified using household-level data. Secondly, most South Africans are uninsured, and rely on the public healthcare system for their health needs (Mayosi & Benatar, 2014), which is almost entirely financed by taxes. Households are therefore unlikely to spend that much on healthcare. This is also reflected in the low proportion of healthcare spending in the household budget, as shown earlier in Table 2.4.

Household welfare is often proxied by consumption of goods and services within a household. If an individual within the household can be made better off without making others within the same household worse off, that is said to be a Pareto efficient improvement (Deaton, 2018). Tobacco expenditure can be thought of as making smokers within a household better off, in the sense of greater pleasure or satisfaction. The non-smokers within the household, as well as being subjected to second-hand smoke, are forced to bear the additional opportunity cost of lower expenditure on other items that may be of greater interest to them. The ensuing discussion focuses on these types of trade-offs and their potential implication for overall household welfare.

The results in Table 2.5 suggest that transport and communication budget shares were consistently lower for households that reported tobacco expenditure across all surveys. For transport, the highest difference in budget shares was in 2010-11 (about 1.6 percentage points), while the lowest was in 1995 (about 0.2 percentage points). The differences in budget shares for communication-related expenditures were much lower by comparison, that is, less than 0.01 percentage points in 1995, 0.03 percentage points in both the 2000 and 2010-11 surveys, and 0.4 percentage points in 2005-06.

Expenditure on housing and household goods expenditures were also possibly crowded-out in the 1995, 2005-06 and 2010-11 surveys. The difference in budget share spent on housing was

highest in 1995, with a 2.3 percentage point difference, followed by differences of 1.9 and 1.4 percentage points in 2010-11 and 2005-06, respectively. In comparison, the difference in household goods expenditure was highest in 2005-06, with 0.7 percentage points, while the differences in the 1995 and 2010-11 surveys were both 0.4 percentage points.

The preceding results suggest that expenditures on transport, communication, housing, and household goods are lower in tobacco-consuming households than in non-tobacco-consuming households. In addition, the 2000 and 2010-11 surveys also noted similar reductions in the expenditures for clothes and shoes, while the miscellaneous expenditures were consistently lower for tobacco-spending households across all surveys. The miscellaneous category included any expenses[8] that did not fit into any of the other eleven categories.

The magnitude of the difference in budget shares of rural households compared to urban households was somewhat mixed (see Figure 2.4).[9] In 1995, rural areas experienced higher levels of crowding-out. For instance, there was an average 0.2 percentage point difference in urban areas, compared to a 0.3 percentage point difference in rural areas, while that difference was 0.2 percentage points and 0.7 percentage points respectively for household goods. On the other hand, there was a 3.1 percentage point difference in housing budget shares in urban areas, compared to only 0.9 percentage points in rural areas that year. A similar comparison for the 2000 survey shows that rural households experienced the biggest difference in food budget shares (7.0 percentage points) compared to urban ones. Conversely, most other categories had higher differences in urban households than in rural households. For example, for clothing and shoes urban households averaged a 0.7 percentage point difference, compared to 0.4 percentage points in rural areas, while transport and communication recorded 1.5 and 0.4 percentage point differences in urban areas, compared to 0.3 and 0.2 in rural areas.

Similarly, when disaggregated by income tercile (i.e., thirds), some results, such as the food expenditure share differences, were reversed (see Figure 2.5). For example, among the lowest income group (Tercile 1) of the population, the average share of household budget spent on food was lower among households reporting tobacco expenditures than those reporting no tobacco expenditures. On the other hand, households in the top third of the income distribution (Tercile 3) that reported tobacco expenditures also reported higher food expenditures compared

[8] This included, for instance, personal care expenses, savings, donations, jewellery, funeral expenses, insurance, and other financial expenses.

[9] Note that Figures 2.4 and 2.5 only report the magnitudes of the difference in budget shares for the indicated categories. It does not necessarily reflect their statistical significance.

to those that did not report tobacco expenditures. Thus, crowding-out was detected in the lowest third of the income distribution, while crowding-in was found in the top third.

An interesting result in this part of the analysis is the percentage point difference for alcohol. Across all sub-populations and surveys, alcohol expenditure is consistently higher among households that report tobacco use than those that do not. Furthermore, the percentage point difference was higher among rural households than urban ones. This suggests that households in rural areas that reported tobacco expenditure tended on average to spend larger shares of their household budget on alcohol than did urban households. For instance, in 2010-11 the percentage point difference of the alcohol budget share was an average of 2.6 percentage points in urban households compared to 3.0 percentage points in rural areas.

Disaggregation along the income gradient presents even more troubling results. Poorer households in the first and second terciles (Terciles 1 and 2) had much higher budget share differences than the top tercile (Tercile 3). The difference in budget shares between the lower two terciles were similar, however; for instance, in 1995 the difference in Tercile 1 was 1.3 percentage points, compared to 1.4 percentage points in Tercile 2 of the same year. Similarly, the differences in Tercile 1 of 2000 and 2005-06 were 3.7 and 1.6 percentage points, compared to 3.8 and 1.7 percentage points in Tercile 2 for the respective survey years. On the other hand, the 2010-11 survey showed a marked difference between the two terciles, with a much higher difference in the lowest tercile. Tercile 1 had a difference of 4.1 percentage points, while Tercile 2 had one of 3.0 percentage points. This last point hints at a possibly worsening situation, in which the poorest households not only spend more on tobacco, but on alcohol as well.

Table 2.5: Differences in expenditure shares between smoking and non-smoking households

	Full sample				Urban sample				Rural sample			
Category	**1995**	**2000**	**2005-06**	**2010-11**	**1995**	**2000**	**2005-06**	**2010-11**	**1995**	**2000**	**2005-06**	**2010-11**
Food	0.7***	4.9***	-1.0**	0.7**	-0.9***	2.5***	-2.3***	-0.8**	2.5***	7.0***	0.02	0.3
Alcohol	-1.1***	-3.1***	-1.3***	-2.7***	-0.9***	-2.9***	-1.2***	-2.6***	-1.4***	-3.6***	-1.5***	-3.0***
Clothes & shoes	-0.1	0.6***	0.0	0.8***	-0.3**	0.7***	-0.3*	0.4**	0.3**	0.4**	0.3*	1.3***
Education	0.2***	0.9***	0.8***	0.5***	0.2**	0.9***	0.8***	0.6***	0.3***	0.9***	0.7***	0.5***
Housing	2.3***	0.1	1.4***	1.9***	3.1***	0.9**	1.8***	2.6***	0.9**	-0.3	1.3***	2.3***
Transport	0.2**	0.9***	1.1***	1.6***	0.3	1.5***	1.3***	2.0***	0.4***	0.3*	1.2***	1.4***
Communication	0.0***	0.3***	0.4***	0.3**	0.1***	0.4***	0.5***	0.5***	0.0**	0.2***	0.4***	0.0
Health	0.0	0.1**	0.0	0.0	0.1	0.2**	0.1	0.0	0.0	0.0	-0.1	-0.1
Household goods	0.4***	0.0	0.7***	0.4**	0.2*	0.0	0.8***	0.1	0.7***	0.1	0.2	0.3
Recreation	0.0	0.1	0.0	-0.1**	0.1	0.2*	0.1	-0.1	-0.1	0.2*	0.0	0.0
Miscellaneous	0.5**	0.7**	1.5***	0.9***	1.4***	1.1***	2.1***	1.3***	-0.4	0.2	0.8**	1.1***
	Tercile 1 – Poorest				**Tercile 2 – Middle**				**Tercile 3 – Wealthiest**			
Category	**1995**	**2000**	**2005-06**	**2010-11**	**1995**	**2000**	**2005-06**	**2010-11**	**1995**	**2000**	**2005-06**	**2010-11**
Food	3.5***	6.1***	1.7***	1.9***	0.3	3.8***	-1.1**	-0.8*	-1.6***	-0.4	-2.3***	-0.6*
Alcohol	-1.3***	-3.7***	-1.6***	-4.1***	-1.4***	-3.8***	-1.7***	-3.0***	-0.6***	-1.9***	-0.6***	-1.2***
Clothes & shoes	-0.1	0.8***	0.5*	1.5***	0.0	0.9***	0.0	0.6**	0.1	0.3**	-0.3	0.1
Education	0.2***	0.8***	0.8***	0.6***	0.2**	0.7***	0.6***	0.5***	0.3**	1.3***	0.8**	0.4**
Housing	1.9***	0.2	0.5	1.0**	2.4***	1.3***	1.9***	4.2***	1.9***	-0.5	1.4**	1.2*Yes
Transport	0.1**	0.8***	0.8*	2.0***	0.3**	1.2***	1.8***	1.6***	0.3	2.1***	0.4	1.6**
Communication	0.0***	0.3***	0.6***	0.5**	0.0***	0.5***	0.6***	0.2*	0.0	0.5***	0.0	0.1
Health	0.2**	0.1*	0.1	0.0	0.0	0.0	-0.1	-0.1	-0.2	0.3**	0.1	0.0
Household goods	-0.1	0.1	0.7**	0.7***	1.0***	0.1	0.5*	0.0	0.4**	0.4**	0.9**	0.3*
Recreation	0.0	0.4**	-0.2**	0.0	0.0	0.1	0.1	-0.1*	0.0	0.3*	-0.1	-0.2**
Miscellaneous	-0.6*	0.7**	0.6**	1.3***	0.9**	0.7**	1.0**	1.1***	1.7***	2.0***	2.3***	1.0**

The coefficients are the differences in weighted mean expenditure shares expressed as percentage points. A positive coefficient indicates that tobacco consuming households had a higher expenditure share than households that did not consume tobacco. *p<0.1, **p<0.05, ***p<0.01.

Figure 2.4: Differences in expenditure shares between smoking and non-smoking households by settlement type

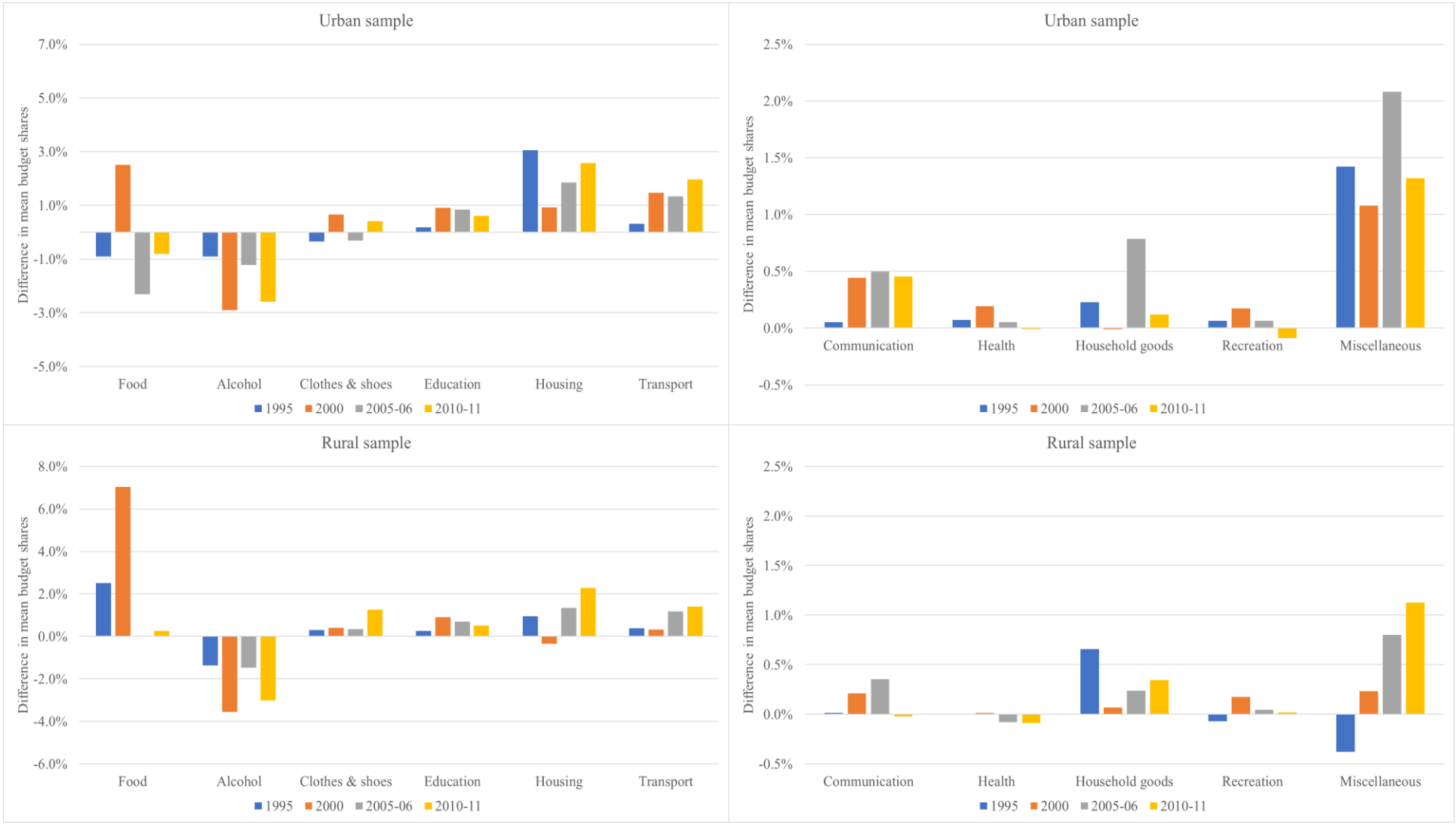


Figure 2.5: Differences in expenditure shares between smoking and non-smoking households by tercile

2.7.2. Consumer separability

As noted earlier, an important consideration for the specification of an appropriate econometric model is how the model deals with zero expenditure on tobacco. The consumer separability test by Vermeulen (2003) builds on the work of Browning & Meghir (1991) to determine whether smokers and non-smokers have the same preferences.

To recapitulate, Equation 5 is specified as follows:

$$w_{ij} = \alpha_{1i} + \alpha_{2i}(p_t t)_j + \alpha_{3i} lnM_j + \alpha_{4i}(lnM_j)^2 + \gamma_i \boldsymbol{x_j} + u_{ij} \quad (5)$$

This equation is concerned with determining the effect of tobacco expenditure $(p_t t)$ on the budget share (w_{ij}). Implicit in the specification is the assumption that smokers and non-smokers have similar utility functions. Thus, the only reason that non-smokers report zero expenditure on tobacco is that it is unaffordable for them (corner solution).

In comparison, Equation 7 is specified as follows:

$$w_{ij} = (\alpha_{1i} + \alpha_{2i} d_j + \alpha_{3i}(p_t t)_j + \gamma_i \boldsymbol{x_j}) + (\beta_{1i} + \beta_{2i} d_j) lnM_j + (\zeta_{1i} + \zeta_{2i} d_j)(lnM_j)^2 + u_{ij} \quad (7)$$

This equation also aims to determine the effect of tobacco expenditure on the budget shares of other household goods and services. However, the specification now implicitly assumes a fundamental difference in the utility functions of smokers versus non-smokers by including the binary variable, d_j. It suggests that non-smokers record zero on tobacco expenditure not because it is unaffordable for them, but because it is not an argument in their utility function, and they would still record zero purchases even if tobacco were free (abstention). Using Equation 7, (Vermeulen, 2003) test simply determines if the coefficients on d_j are jointly zero, i.e. $H_0: \alpha_{2i} = \beta_{2i} = \zeta_{2i} = 0$. If this is indeed the case, then Equation 7 effectively reduces to Equation 5, and we can reject any differences in utility function between smokers and non-smokers. The results of this test are shown in Table 2.6 below.

In summary, the null hypothesis of zeros arising from a corner solution was rejected for most household categories in the 1995 and 2005-06 IESs, while the null hypothesis was rejected for only some categories in both the 2000 and 2010-11 IESs. In fact, for the 2010-11 IES only the food category rejected the null hypothesis that smokers and non-smokers had heterogenous preferences.

Table 2.6: Chi-square statistics for test of heterogeneous preferences between tobacco users and non-users

Category	IES 1995	IES 2000	IES 2005-06	IES 2010-11
Food	**298.1*****	2.5	**23.6*****	**5.73****
Tobacco	**490.9*****	**712.8*****	**179.9*****	0.7
Alcohol	**33.2*****	1.7	**11.5*****	1.9
Clothes & shoes	**209.5*****	2.2	**5.5****	0.2
Education	0.6	0.6	1.0	1.9
Housing	**163.1*****	**13.2*****	**2.8***	1.0
Transport	**13.1*****	0.8	1.1	1.6
Communication	**5.1****	0.7	**7.9****	1.8
Health	**10.5****	1.8	**23.93****	2.6
Household goods	**45.8*****	1.5	**16.4*****	2.0
Recreation	0.4	**3.1***	**10.3****	1.9
Miscellaneous	0.0	**4.4****	**18.3*****	2.3

A Wald test is performed to test a composite linear hypothesis that all three coefficients associated with the tobacco use dummy variable in Equation 7 are jointly zero. A rejection (i.e. significant test statistic) suggests that Equation 7 is a more appropriate specification, whereas no rejection implies Equation 5 is the better specification. The results in most categories of the 1995 and 2005-06 IESs, as well some categories in the 2000 IES, suggest the Equation 7 specification to be appropriate, i.e., that the heterogeneity of tobacco-users specification is correct and is used going forward.

For consistency, the Equation 7 specification is adopted throughout and presented in the subsequent sections of these results. However, examples of the full regression result for both Equation 5 and Equation 7 are presented in the annexes (see Tables A10 to A15). By and large, the conclusions remain the same regardless of the specification, however, the magnitude of the effects appears to be larger for the coefficients in Equation 7 compared to Equation 5.

2.7.3. Three stage least squares (3SLS)

The results for the coefficients on $p_t t$ (tobacco expenditure) are given in Tables 2.7 and 2.8. Recall that, in this section, I examine the effect of tobacco expenditure on the budget share of various categories of goods and services at the household level. To report the results conveniently, the coefficients were multiplied by 100. The interpretation given is the effect of a R100 change in monthly expenditure on tobacco on the budget share of other household expenses, expressed as a percentage.

Crowding-out effects were observed in various expenditure categories across different surveys. In the 1995 survey, it was evident that tobacco expenditure had a crowding-out impact on expenditures related to housing and household goods. Specifically, a R100 increase in tobacco

spending was associated with a reduction in the budget shares allocated to housing and household goods by 0.17 and 0.06 percentage points, respectively. These findings provide only partial validation of the outcomes presented in the previous analysis of differences in budget shares. To be more precise, while the earlier analysis indicated potential crowding-out effects in seven categories, this subsequent analysis confirmed crowding-out in only two of the seven categories.[10]

The 2000 survey results confirm crowding-out effects in four of the seven categories highlighted in the earlier analysis. On average, the largest reductions in budget shares resulting from a R100 increase in tobacco spending were on transport (0.35 percentage points), household goods (0.29 percentage points), and education (0.28 percentage points).

Both the 2005-06 and 2010-11 surveys found crowding-out on education and household goods. However, crowding-out was also noticed on housing in the 2005-06 survey, whereas in the 2010-11 survey, crowding-out was also detected on transport. A R100 rise in tobacco spending in 2005-06 led to 0.03 percentage point decrease in the budget share for education, while the same increase led to 0.10 percentage point decrease for education in the 2010-11 survey.

[10] Strictly speaking, this is out of six, not seven, categories. This is because the 3SLS procedure requires that one category be dropped from the analysis – *miscellaneous expenditures* was dropped. It would therefore not appear in the later analysis, but is one of the crowded-out categories in the difference in budget share analysis.

Table 2.7: Three stage-least squares (3SLS) regression results – Full sample, urban and rural sub-samples

	Full sample				Urban sample				Rural sample			
	IES 1995	**IES 2000**	**IES 2005-06**	**IES 2010-11**	**IES 1995**	**IES 2000**	**IES 2005-06**	**IES 2010-11**	**IES 1995**	**IES 2000**	**IES 2005-06**	**IES 2010-11**
Food	-0.201	-1.9	-0.0688	8.69**	-1.68	-0.0418	0.186	10.1**	-0.198	-2.39	-0.0166	0.696
	(0.379)	(1.73)	(0.393)	(4.4)	(1.97)	(0.247)	(0.974)	(4.43)	(0.63)	(2.24)	(0.43)	(1.87)
Alcohol	0.601	6.22*	0.76***	5.66**	0.489	0.132	0.963***	3.05***	0.855*	6.95*	0.716**	1.14
	(0.17)	(3.24)	(0.205)	(2.82)	(0.436)	(0.362)	(0.372)	(1.12)	(0.459)	(4.05)	(0.28)	(2.24)
Cloths	-0.187	-0.397	0.637***	-1.27	-0.457	0.0389	0.545	-1.18	-0.136	0.5	0.714**	-0.0532
	(0.162)	(0.621)	(0.247)	(0.838)	(0.617)	(0.119)	(0.441)	(0.81)	(0.271)	(1.12)	(0.345)	(0.377)
Educ	0.106	-2.81*	-0.275*	-1.01*	0.058	0.00222	-0.182	-0.845***	0.156	-4.99*	-0.405	-0.121
	(0.0711)	(1.55)	(0.147)	(0.603)	(0.155)	(0.0421)	(0.242)	(0.387)	(0.151)	(2.98)	(0.259)	(0.284)
Housing	-1.65**	-0.335	-1.14**	4.88	0.52	-0.147	0.0341	-2.26	-2.45	-2.27	-1.74**	1.41
	(0.738)	(1.29)	(0.455)	(3)	(1.9)	(0.192)	(0.804)	(2.25)	(1.53)	(2.21)	(0.846)	(2.76)
Trans	0.187	-3.47*	2.1***	-1.36	0.319	0.0588	2.54**	-0.306	-0.0041	-6.66*	1.93**	-0.17
	(0.159)	(2.05)	(0.573)	(1.16)	(0.404)	(0.113)	(1)	(1.15)	(0.289)	(3.97)	(0.863)	(1.18)
Comm	-0.0098	-1.8*	0.479***	1	-0.0059	0.00205	0.544*	-0.302	-0.0173	-2.86*	0.435**	0.309
	(0.0097)	(0.98)	(0.167)	(0.636)	(0.0158)	(0.0454)	(0.318)	(0.414)	(0.0228)	(1.73)	(0.216)	(0.664)
Health	-0.176	-0.681*	-0.135	-1.36*	0.126	-0.0096	-0.319	-0.372	-0.558	-1.06	-0.0506	-0.337
	(0.112)	(0.395)	(0.0957)	(0.706)	(0.227)	(0.0266)	(0.203)	(0.257)	(0.355)	(0.689)	(0.107)	(0.753)
Hhgds	-0.589**	-2.9*	-1.4***	-5.5*	-1.1	-0.0451	-1.98**	-2.62**	-0.366	-3.03	-1.22**	-1.09
	(0.25)	(1.6)	(0.455)	(2.85)	(1.03)	(0.163)	(0.82)	(1.16)	(0.391)	(1.93)	(0.512)	(2.11)
Recre	0.0867	0.521	0.359***	0.749*	0.152	-0.0202	0.438**	-0.139	0.0617	2.24	0.321*	0.254
	(0.0721)	(0.649)	(0.126)	(0.419)	(0.199)	(0.0519)	(0.211)	(0.216)	(0.138)	(1.67)	(0.168)	(0.597)

The results are the 3SLS estimation of equation 7. Coefficients on $\boldsymbol{p_i t}$ is the change in expenditure shares associated with a R100 increase in monthly expenditure on tobacco. Negative coefficient on $\boldsymbol{p_i t}$ indicates that tobacco expenditure crowds out that commodity. For brevity household characteristics included in the regression are not reported. *p<0.1, **p<0.05, ***p<0.01. Bootstrapped standard errors are in brackets. (Tob – Tobacco, Educ – Education, Trans – Transport, Comm – Communication, Hhgds – Household goods, Recre – Recreation)

Table 2.8: Three Stage-Least Squares (3SLS) Regression results – Tercile sub-samples

	Tercile 1				Tercile 2				Tercile 3			
	IES 1995	**IES 2000**	**IES 2005-06**	**IES 2010-11**	**IES 1995**	**IES 2000**	**IES 2005-06**	**IES 2010-11**	**IES 1995**	**IES 2000**	**IES 2005-06**	**IES 2010-11**
Food	-4.13	5.1*	0.77	7.37	-0.104	-1.2	-0.509	3.14	0.276	-2.19	-0.852	-0.323
	(5.11)	(2.96)	(1.02)	(5.3)	(0.396)	(2)	(0.537)	(2.83)	(0.896)	(1.49)	(0.744)	(1.03)
Alcohol	0.335	3.6**	1.11***	2.62*	0.591***	0.979	0.324***	1.07	0.37	1.54	0.717*	-1.15
	(0.296)	(1.72)	(0.408)	(1.44)	(0.153)	(1.1)	(0.117)	(1.2)	(0.439)	(0.985)	(0.37)	(0.963)
Cloths	0.495	-1.38	1.55**	0.011	-0.263	0.368	0.455*	-0.958	-0.165	-0.393	-0.231	-0.46
	(1.07)	(0.945)	(0.601)	(1.01)	(0.203)	(0.526)	(0.241)	(0.875)	(0.498)	(0.409)	(0.384)	(0.519)
Educ	0.0347	-0.336	-0.188	0.143	0.11	-0.7	-0.372*	0.192	0.192	-1.05	-0.303	1.28
	(0.126)	(0.404)	(0.259)	(0.534)	(0.0836)	(0.771)	(0.208)	(0.585)	(0.391)	(0.695)	(0.344)	(1.1)
Housing	-3.7	-1.03	-1.39	-4.83	-2.72***	-1.18	-0.725*	-1.88	-1.97	0.661	-3.75*	-2.84
	(9.07)	(1.7)	(1.07)	(4.08)	(0.79)	(0.82)	(0.422)	(2.9)	(2.61)	(1.49)	(1.98)	(2.35)
Trans	0.356*	-0.744	0.322	-0.955	0.332**	-0.158	1.95***	0.548	-0.374	-1.14	7.87**	-1.41
	(0.215)	(0.699)	(0.47)	(1.51)	(0.153)	(0.338)	(0.488)	(1.13)	(0.951)	(0.994)	(3.63)	(1.6)
Comm	-0.00772	-0.419	0.478	-0.262	0.00719	-0.359	0.502***	0.0766	-0.036	-0.885	0.784*	-0.898
	(0.0367)	(0.283)	(0.251)	(0.56)	(0.0123)	(0.363)	(0.155)	(0.748)	(0.0505)	(0.635)	(0.43)	(0.755)
Health	0.253	-0.154	-0.448*	-0.379	0.24**	-0.0965	-0.0632	-0.121	-1.22	-0.459	0.47	0.393
	(0.281)	(0.273)	(0.235)	(0.448)	(0.118)	(0.169)	(0.112)	(0.208)	(1.5)	(0.34)	(0.323)	(0.365)
Hhgds	0.557	-0.864	-1.14**	-1.91	-0.0914	-0.667	-1.16***	-1.35	-0.504	-1.56	-3.69**	1.74
	(1.28)	(0.5530)	(0.552)	(1.32)	(0.234)	(0.673)	(0.329)	(1.37)	(0.835)	(1)	(1.72)	(1.43)
Recre	0.254	0.084	0.552**	0.497*	0.0504	-0.0641	0.344***	0.0831	0.177	-0.295	-0.192	-0.177
	(0.202)	(0.492)	(0.236)	(0.266)	(0.0918)	(0.198)	(0.108)	(0.167)	(0.374)	(0.32)	(0.259)	(0.242)

The results are the 3SLS estimation of equation 7. Coefficients on $\boldsymbol{p_t t}$ is the change in expenditure shares associated with a R100 increase in monthly expenditure on tobacco. Negative coefficient on $\boldsymbol{p_t t}$ indicates that tobacco expenditure crowds out that commodity. For brevity household characteristics included in the regression are not reported. *p<0.1, **p<0.05, ***p<0.01. Bootstrapped standard errors are in brackets.

Furthermore, the same increase in tobacco expenditure led to decreases in the budget shares for household goods expenses by 0.14 and 0.56 percentage points. Thus, the effect of a R100 increase in tobacco spending had larger impacts on the same categories in the 2010-11 survey compared to the 2005-06 survey. The decrease in the budget share on transport was estimated to be 0.14 percentage points.

Qualitatively, the patterns of crowding-out are broadly similar when disaggregated geographically (rural vs. urban). However, a few categories retain their statistical significance. Among urban households, the magnitude of the coefficients that are statistically significant were all found to be smaller than for the full sample, while the opposite is true for rural households. This suggests that the opportunity cost of tobacco expenditure is likely to have a greater effect among rural households. This confirms a similar result in the earlier analysis.

2.7.4. Diagnostics

Three diagnostic tests were conducted to ascertain the robustness of the results presented. Earlier studies emphasise the potential endogeneity of $p_t t$ and M in equation 7 (John, 2008, Jumrani & Birthal, 2017, Nguyen & Nguyen, 2020, Pu et al., 2008, San & Chaloupka, 2016). Table 2.9 presents the chi-square statistics from the Durbin-Wu-Hausman test for endogeneity. The null hypothesis that $p_t t$, lnM and $(lnM)^2$ are exogenous is rejected for most of the categories across all surveys.

Table 2.9: Chi-square statistics for endogeneity test for $p_t t$ lnM and (lnM)2

	1995						2000					
	Full	**Urban**	**Rural**	**Terc 1**	**Terc 2**	**Terc 3**	**Full**	**Urban**	**Rural**	**Terc 1**	**Terc 2**	**Terc 3**
Food	**1378.63*****	**526.50*****	**848.16*****	**693.54*****	**79.52*****	**75.06*****	**39.96*****	**28.07*****	**18.44*****	**21.79*****	**35.90*****	**22.48*****
Tob	**246.54*****	**149.71*****	**54.60*****	**21.19*****	**43.27*****	**17.78*****	1.55	1.62	13.02**	3.59	4.78	1.50
Alc	**216.11*****	**129.15*****	**91.44*****	**96.62*****	**45.76*****	**30.16*****	**183.41*****	**113.08*****	**67.69*****	**82.76*****	**52.29*****	**61.04*****
Clth & shs	**862.97*****	**335.89*****	**512.21*****	**427.34*****	**44.71*****	**25.65*****	**238.76*****	**122.43*****	**108.33*****	**59.95*****	**23.91*****	**9.90****
Educ	**129.86*****	**60.31*****	**67.62*****	**9.37****	**13.13****	**31.57*****	**95.93*****	**63.86*****	**30.02*****	**9.05****	**20.42*****	**46.40*****
Hous	**3693.57*****	**1694.32*****	**1872.94*****	**1838.32*****	**864.38*****	**307.81*****	**624.79*****	**447.61*****	**172.29*****	**93.34*****	**127.6*****	**334.86*****
Trans	**417.43*****	**246.68*****	**186.53*****	**83.45*****	**137.67*****	**86.65*****	**232.20*****	**150.29*****	**70.53*****	5.69	**14.28****	**114.72*****
Comm	**326.96*****	**209.37*****	**107.24*****	**44.58*****	**46.50*****	**75.63*****	**224.54*****	**163.89*****	**63.75*****	**15.73****	**24.51*****	**101.49*****
Health	**626.99*****	**400.54*****	**217.30*****	**82.33*****	**130.66*****	**114.40*****	**11.19****	**9.36****	**8.51****	**13.53****	**3.05**	**10.97****
HH gds	**562.37*****	**195.99*****	**371.20*****	**247.58*****	**54.37*****	**6.48*****	**185.77*****	**109.92*****	**64.40*****	**19.81*****	**34.95*****	**27.75*****
Recr	**371.98*****	**211.17*****	**150.52*****	**65.16*****	**50.55*****	**64.45*****	**88.97*****	**46.83*****	**130.50*****	**30.65*****	5.25	**38.81*****
Misc	**689.70*****	**343.46*****	**298.82*****	**319.85*****	**112.57*****	**74.68*****	**410.16*****	**309.39*****	**106.35*****	**26.29*****	**179.09*****	**168.62*****
	2005-06						**2010-11**					
	Full	**Urban**	**Rural**	**Terc 1**	**Terc 2**	**Terc 3**	**Full**	**Urban**	**Rural**	**Terc 1**	**Terc 2**	**Terc 3**
Food	**307.05*****	**128.29*****	**188.33*****	**21.07*****	**51.39*****	**42.18*****	**823.32*****	**404.50*****	**428.94*****	**127.95*****	**86.86*****	**77.22*****
Tob	**83.88*****	**67.49*****	**24.55*****	**11.92****	**12.02****	3.46	**36.48*****	**17.97*****	6.12	**10.92****	3.85	1.49
Alc	**110.37*****	**80.28*****	**33.49*****	**62.04*****	**21.35*****	**29.06*****	**158.84*****	**106.82*****	**55.67*****	**68.26*****	**50.84*****	**29.06*****
Clth & shs	**67.04*****	**38.05*****	**27.19*****	**22.11*****	**13.86****	**16.09****	**61.97*****	**27.24*****	**24.25*****	**9.12****	**18.50*****	**24.19*****
Educ	**37.18*****	**41.01*****	2.36	**10.42****	**6.51***	4.69	**12.21****	**13.31****	**13.25****	1.61	**7.94****	**26.36*****
Hous	**26.60*****	**26.55*****	5.05	**18.65*****	3.12	**14.36****	**111.60*****	**34.25*****	**130.10*****	**96.62*****	**15.69****	5.53
Trans	**135.18*****	**80.09*****	**62.31*****	4.04	**44.57*****	**78.47*****	**95.36*****	**92.08*****	4.73	**6.35***	2.07	**100.51*****
Comm	**213.93*****	**122.78*****	**86.77*****	**7.28***	**37.19*****	**53.73*****	**199.34*****	**116.91*****	**88.61*****	**9.71****	**31.78*****	**78.02*****
Health	**13.10****	**14.49****	**12.03****	**22.75*****	**6.40***	5.91	**24.26*****	**25.18*****	**14.33****	**11.76****	4.11	4.77
HH gds	**53.50*****	**28.99*****	**29.63*****	**13.09****	**16.47*****	**54.03*****	**83.21*****	**47.13*****	**33.32*****	**23.64*****	**36.70*****	**36.88*****
Recr	**103.54*****	**75.75*****	**30.89*****	**28.30*****	**14.11****	**49.92*****	**86.03*****	**57.88*****	**27.59*****	**20.23*****	**9.64****	**14.86****
Misc	**229.70*****	**185.22*****	**71.67*****	**24.04*****	**14.50****	**60.00*****	**406.13*****	**266.05*****	**117.65*****	**22.31*****	**43.16*****	**143.52*****

The Chi-square statistic presented for the different sub-samples is a variant of the C-statistic, defined as the difference of two Sargan-Hansen statistics. In Stata, it is generated by the *endogtest ()* option of *ivreg2* and unlike the Durbin-Wu-Hausman tests reported by *ivendog*, can report test statistics that are robust to various violations of conditional homoskedasticity. The null hypothesis that the specified endogenous regressors can be treated as exogenous is rejected if it is statistically significant, indicated as *** p<0.01, ** p<0.05, * p<0.1. Most categories reject this null hypothesis.

Table 2.10 presents results for the Pagan-Hall test for heteroskedasticity. In the presence of heteroskedasticity, IV regression produces inconsistent standard errors, while 3SLS estimates are less efficient and produce inconsistent standard errors (Cameron & Trivedi, 2005). As in most of the literature (Deaton, 2018), the test results indicate that the null hypothesis of homoskedasticity is rejected for most of the expenditure categories in the 1995 survey. The other surveys, however, reflect some irregularity with homoskedasticity, as noted in many categories of the 2000, 2005-06 and 2010-11 surveys.

The third set of tests concerns the validity of the instruments used for the 3SLS procedure. According to John et al. (2019), the R^2 and the *F-statistic* from the first stage regressions are used to determine whether these instruments are weak or strong. The rule of thumb is that an *F-statistic* greater than 10 indicates a relatively strong instrument (Stock, James H., Wright & Yogo, 2002). In the case of more than one endogenous regressor, however, a different set of statistics can be computed such as *Shea's Partial* R^2 (Shea, 1997). Although there is no clear rule of thumb to distinguish weak instruments from strong ones, higher values would indicate stronger instruments, and would be likely to exhibit less bias. Tables A8 and A9 in the annex provide the results of various tests of instrument relevance. The *F-statistic* for the full and urban samples in all the surveys are greater than 10, indicating strong instruments. However, the results for the variable $p_t t$ are inconsistent in the samples disaggregated by rural-urban area or by income terciles, with some greater and others less than 10. The results for *Shea's Partial* R^2 are also potentially concerning, as the values are particularly low on the variable $p_t t$. In comparison, all the income variables *lnM* and *lnM*^{2} have *F-statistics* greater than 10 and relatively high *Shea's Partial* R^2 values.

Regarding the interpretation of these relevance tests, Shea (1997) cautions against using the results as a screening devise. He instead suggests using them to highlight possible areas of concern, for instance, whether there are high standard errors estimated. He further maintains that justification for the use of a particular instrumental variable ought to rely more on prior reasoning to determine its relevance. The results in Table 2.10 are indicative of possible imperfect instrumental variables used in the model and serve as the basis for relaxing the exclusion criteria in the model.

Table 2.10: Chi-Square statistic for Pagan-Hall test for heteroskedasticity

	1995						2000					
	Full	**Urban**	**Rural**	**Terc 1**	**Terc 2**	**Terc 3**	**Full**	**Urban**	**Rural**	**Terc 1**	**Terc 2**	**Terc 3**
Food	**412.968*****	**395.591*****	6.218	1.788	**71.015*****	**272.275*****	11.402	5.412	0.247	7.486	0.563	0.729
Tobacco	**126.529*****	**37.960****	**108.653*****	**33.999****	2.825	**49.319*****	**138.272*****	17.728	4.870	173.312	19.984	**228.259*****
Alcohol	18.846	5.315	9.671	**50.752*****	28.906	1.925	0.474	0.326	0.266	6.953	0.244	0.880
Clothes & shoes	**533.180*****	**477.984*****	26.967	3.654	**120.252*****	**337.701*****	**58.691*****	6.574	0.712	**53.688*****	1.513	**88.889*****
Education	**275.710*****	**158.839*****	**124.980*****	**35.973****	**78.322*****	**61.542*****	0.968	0.355	**78.821*****	**82.172*****	0.320	1.503
Housing	**207.258*****	**43.690****	**551.227*****	2.865	16.359	10.409	**522.227*****	17.782	3.651	**62.648*****	6.740	**34.395****
Transport	**1164.892*****	**679.700*****	**347.347*****	**40.450****	30.806	**142.437*****	2.720	0.631	12.400	**46.439****	52.506	8.028
Communication	**158.099*****	**98.074*****	**40.342****	**62.246*****	**37.943****	16.280	0.445	0.310	0.347	**43.164****	0.629	0.929
Health	**529.585*****	16.941	**475.305*****	**35.342****	**171.890*****	1.011	2.254	0.857	0.257	7.962	7.883	2.318
Household goods	**105.363*****	**301.662*****	4.554	1.459	**188.060*****	**51.549*****	0.777	0.779	0.302	15.950	1.064	1.112
Recreation	**160.056*****	**77.738*****	**70.885*****	**46.135****	22.625	**32.841***	**38.595****	2.931	**29.520***	**39.083****	29.496	28.539
Miscellaneous	**51.270*****	9.842	**35.160****	2.614	**40.098****	1.892	0.545	0.340	366.897	6.571	0.986	0.747
	2005-06						**2010-11**					
	Full	**Urban**	**Rural**	**Terc 1**	**Terc 2**	**Terc 3**	**Full**	**Urban**	**Rural**	**Terc 1**	**Terc 2**	**Terc 3**
Food	**408.750*****	**285.975*****	**130.180*****	10.988	1.914	20.233	2.213	0.257	5.929	2.090	4.540	66.969***
Tobacco	**1060.495*****	10.386	**567.592*****	**36.046****	2.204	**58.054*****	4.726	1.737	**240.319*****	**164.388*****	**57.224*****	**150.572*****
Alcohol	16.546	8.074	19.796	6.832	7.837	1.929	0.481	0.054	6.792	3.211	4.466	0.490
Clothes & shoes	28.228	9.207	**73.524*****	3.783	**45.603****	**127.998*****	5.031	21.464	**35.618****	1.139	5.490	9.978
Education	**206.867*****	**160.560*****	**148.757*****	**201.337*****	11.501	**63.899*****	5.065	4.124	**32.529***	**73.199*****	8.828	0.631
Housing	**50.312*****	7.161	**148.150*****	4.989	**137.244*****	0.630	1.006	0.075	27.240	3.297	0.875	0.440
Transport	13.536	7.674	19.080	**89.988*****	16.756	0.595	**41.242****	11.089	**371.141*****	4.313	**64.681*****	11.655
Communication	14.846	8.628	**39.792****	4.393	6.426	1.830	1.840	0.081	**49.780*****	3.691	3.546	0.382
Health	**159.520*****	**71.156*****	**98.304*****	11.897	3.211	20.050	0.684	0.068	**47.630*****	3.517	**39.054****	2.174
Household goods	9.593	5.232	9.655	3.643	**39.297****	0.522	0.470	0.062	4.832	1.703	2.605	0.471
Recreation	18.441	8.276	**43.071****	4.589	19.478	**34.489****	1.029	0.059	**51.574*****	11.512	**42.851****	**30.830***
Miscellaneous	16.897	**52.393*****	8.125	3.370	**41.121****	**29.839***	0.481	0.060	7.295	1.558	1.812	1.110

The Pagan-Hall general statistic is presented for the different sub-samples. The null hypothesis of homoskedasticity is rejected if it is statistically significant indicated as *** p<0.01, ** p<0.05, * p<0.1. Most categories reject this null hypothesis.

2.7.5. Relaxing the exclusion criteria

Chelwa & Van Walbeek (2014) use this method to estimate bounds for the impact of tobacco use on household expenditure in Zambia. Motivated by experimental evidence on the relationship between smoking and the discount rate, they assume a positive relationship between tobacco use (the endogenous regressor) and the error term. That is, they assume that the discount rate is one of the variables subsumed in the error term. Evidence suggests that the discount rates of smokers and non-smokers are quite different, with smokers having higher discount rates than non-smokers (Bishai, 2001). A similar argument can be made for the gender differences in risk preference. Studies show that women are more risk-averse than men (Croson & Gneezy, 2009). One can infer from this that it is likely that the adult-sex ratio (calculated as number adult males divided by number of adults in the household) would be more significantly correlated with the error term than the female ratio (calculated as number of adult females divided by number of adults in the household). If we further consider the higher smoking rate among men, it is quite plausible that the female ratio (the imperfect IV) is less correlated with the error term than the endogenous regressor (in this case tobacco expenditure). Thus, the key identifying assumptions for Nevo & Rosen (2012) method are satisfied.

In this section of the analysis, the female ratio is used as the IIV. This differs from most of the literature, which uses the adult-sex ratio (Chelwa & Van Walbeek, 2014, Husain et al., 2018, John, 2008, Nguyen & Nguyen, 2020, Pu et al., 2008). However, the female ratio takes precedence in the literature (see, for example, San & Chaloupka (2016)). Importantly, the advantage of using it is that it imposes a negative relationship between the IIV and the endogenous regressor. As Nyagwachi (2019) points out, this provides the opportunity to generate both upper and lower bounds of the point estimates using the Nevo & Rosen (2012) method. A further advantage is that, whether the adult-sex ratio or the female ratio is used as an instrument, the point estimates from the 3SLS regression are numerically identical.

Robustness of the 3SLS results can be assessed by comparing those point estimates to the bounds estimated using the Nevo & Rosen (2012) method. When the range of the upper and lower bounds encompasses zero, this signals a lack of mutual consistency in the IIV assumptions, thereby leading to their rejection. Conversely, if the upper and lower bounds are either both positive or both negative, this suggests that the assumptions remain unchallenged. This occurrence of bounds crossing provides a valuable specification test for the underlying assumptions.

Table 2.11: Lower and upper bounds for the estimates of tobacco expenditure variable

	IES 1995	IES 2000	IES 2005-06	IES 2010-11
Food	[-0.592, 0.197]	[-12.940, 0.166]	[-2.644, 0.153]	[-4.434, 0.122]
Alcohol	**[0.045, 0.350]**	[-28.954, 0.072]	**[0.050, 0.421]**	[-18.486, 0.053]
Clothes and shoes	[-0.334, 0.019]	[-2.396, 0.022]	**[0.056, 0.238]**	[-2.494, 0.033]
Education	[-0.004, 0.007]	**[-3.634, -0.018]**	**[-0.823, -0.010]**	**[-7.306, -0.014]**
Housing	**[-4.713, -0.087]**	**[-14.159, -0.111]**	**[-3.128, -0.017]**	[-26.460, 0.010]
Transport	[-0.025, 0.014]	**[-3.367, -0.012]**	**[0.032, 1.130]**	**[-4.100, -0.009]**
Communication	[-0.022, 0.001]	[-1.935, 0.0002]	**[0.044, 0.279]**	[-5.357, 0.029]
Health	[-0.271, 0.007]	**[-1.267, -0.002]**	[-0.422, 0.002]	[-7.506, 0.006]
HH goods	**[-0.880, -0.018]**	**[-3.644, -0.020]**	**[-3.106, -0.066]**	**[-37.324, -0.027]**
Recreation	[-0.006, 0.008]	**[-2.129, -0.001]**	**[0.028, 0.204]**	[-3.213, 0.012]
Miscellaneous	[0.953, -0.067]	[-51.136, 0.051]	**[-1.756, -0.075]**	**[-55.783, -0.079]**
Sample Size	**29,512**	**26,069**	**20,995**	**24,826**

The table shows estimates of bounds for the causal effect of tobacco expenditure ($p_t t$) using the method in Nevo and Rosen (2012). Bounds in bold imply that the causal effect of $p_t t$ is statistically significant since such bounds do not overlap zero. Positive values indicate tobacco consuming households have a higher budget share than households that do not consume tobacco.

The results from the relaxation of the exclusion criteria using Nevo & Rosen (2012) are shown in Table 2.11 above. The bounds in bold are statistically significant at the 95% level. Negative values indicate crowding-out, while positive ones indicate crowding-in. Qualitatively, the results largely validate the results of the 3SLS and the difference analysis. For instance, in the 1995 survey, both housing and household goods expenses were associated with lower expenditure among households that reported tobacco use. Similarly, crowding-out was noted on education, housing, transport, health, and household goods expenditures in 2000. In 2005-06 education, housing, housing, and household goods expenditures were crowded-out, while in 2010-11 it was education, transport, household goods and miscellaneous expenditures. The 3SLS point estimates were all within the range of the bounds estimated for those categories where crowding-out was detected. Interestingly, however, where crowding-in was detected, for product categories like alcohol, 3SLS point estimates were typically higher than the upper bounds of estimates found in Table 2.11. Despite this, the overall conclusions remain unchanged across all four surveys.

2.8 Summary and conclusion

The aim of this chapter was to determine the opportunity cost of tobacco expenditure among South African households in an environment of increasing tobacco taxes. To do this, the paper

analysed four cross-sectional surveys conducted in 1995, 2000, 2005-06 and 2010-11. The analysis follows the steps provided by John et al. (2019). The initial steps involve a simple comparison of differences in expenditure shares between households reporting tobacco expenditure and those that did not. The second phase uses an instrumental variables technique to account for endogeneity within a system of equations framework that considers contemporaneous correlation of errors using a 3SLS approach. A final approach accounts for an imperfect instrument using the method proposed by Nevo & Rosen (2012) to estimate intervals for the point estimates.

The analysis of the difference in expenditure shares concludes that households that reported tobacco expenditure spent less of their budget on food, education, housing, clothes and shoes, transport, communication health, household goods and miscellaneous expenses. This provided the first indication of crowding-out due to tobacco expenditure. The econometric results from the 3SLS regressions confirmed that some of the above categories were indeed crowded-out. Specifically, education, transport, housing, and household goods expenditures were confirmed to have been crowded-out in various years. The potential impact on development may be seen in various ways, such as decreased human capital development because of reduced expenditure on education at the household level.

These results are presented with certain limitations. First, the cross-sectional analysis using 3SLS did not include regional fixed effects, as would typically be the case (Deaton, 2018). Fixed effects would have accounted for the clustering of relative prices within geographic regions. However, although the IES 1995 included a suitable cluster variable in the dataset, the remaining IES datasets did not. Thus, for consistency, I opted to omit it in my analysis.

As well, the use of the IV method potentially has the usual limitation of not meeting the exclusion criteria. The adult-sex ratio used in this case is possibly correlated with other covariates in the model, and is thus an imperfect instrument. This may have biased the point estimates. The dataset was, however, limited on other possible instruments to use in the model. The chapter attempts to address this by estimating bounds for the point estimates.

Finally, this chapter analyses the data cross-sectionally, and fails to account for any dynamics over the period. Several factors were changing over the 15-year period between the first and last surveys. For instance, the prevalence of smoking decreased from around 33% in 1995 to around 20% in 2010. This means the composition of households reporting tobacco expenditure and those not reporting such expenditure varied over time. Despite this, the cross-sectional

analysis presents a first look at the state of crowding-out in South Africa during a period of large tax increases.

2.9 Annex A

Table A 1: Mean budget shares from the Income and Expenditure Surveys (IES) 1995, 2000, 2005-06 and 2010-11 disaggregated by tobacco use and settlement type

	Full sample								Tobacco consuming								Tobacco non-consuming							
	IES 1995		IES 2000		IES 2005-06		IES 2010-11		IES 1995		IES 2000		IES 2005-06		IES 2010-11		IES 1995		IES 2000		IES 2005-06		IES 2010-11	
Category	Urb	Rur	Urb	Rur	Urb	Rur	Urb	Rur	Urb	Rur	Urb	Rur	Urb	Rur	Urb	Rur	Urb	Rur	Urb	Rur	Urb	Rur	Urb	Rur
Food	27.7	31.7	39.5	56.8	24.1	34.0	21.2	33.5	28.2	30.3	37.9	51.9	25.5	34.0	21.8	33.3	27.2	32.8	40.4	59.0	23.2	34.0	21.0	33.5
Tob	1.6	1.5	2.0	1.7	1.4	1.2	1.0	0.7	3.2	3.3	5.4	5.5	3.7	3.4	4.0	4.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Alc	1.1	1.1	2.0	2.0	0.8	0.7	1.1	0.9	1.6	1.8	3.9	4.5	1.5	1.7	3.0	3.4	0.7	0.4	1.0	0.9	0.3	0.3	0.4	0.4
Clth & shs	6.2	6.1	6.6	6.8	7.7	9.1	6.9	8.1	6.4	5.9	6.2	6.5	7.9	8.8	6.6	7.0	6.1	6.2	6.8	6.9	7.6	9.2	7.0	8.3
Educ	1.0	0.7	2.2	1.8	2.0	1.6	2.1	0.8	0.9	0.6	1.6	1.2	1.4	1.1	1.6	0.4	1.1	0.8	2.5	2.1	2.3	1.8	2.2	0.9
Hous	34.8	41.2	19.8	12.4	22.9	18.6	29.2	23.4	33.2	40.7	19.2	12.6	21.8	17.7	27.3	21.6	36.3	41.6	20.2	12.3	23.6	19.1	29.9	23.8
Trans	4.9	1.5	8.8	4.8	12.8	9.7	13.7	11.1	4.7	1.3	7.8	4.6	12.0	8.9	12.2	9.9	5.1	1.7	9.3	4.9	13.4	10.1	14.2	11.4
Comm	0.2	0.1	2.4	1.0	4.1	3.2	3.7	2.9	0.2	0.0	2.1	0.8	3.8	3.0	3.3	2.9	0.3	0.1	2.6	1.0	4.3	3.3	3.8	2.9
Health	2.8	1.0	1.2	1.1	1.6	1.8	1.6	1.4	2.8	1.0	1.0	1.1	1.6	1.9	1.6	1.5	2.8	1.0	1.2	1.1	1.6	1.8	1.6	1.4
HH gds	4.6	3.8	3.0	2.7	7.0	8.5	4.9	6.7	4.5	3.4	3.0	2.7	6.5	8.3	4.8	6.4	4.7	4.1	3.0	2.8	7.3	8.6	5.0	6.8
Recr	2.1	0.8	2.2	1.2	2.5	1.4	1.7	0.9	2.1	0.9	2.1	1.1	2.5	1.3	1.7	0.9	2.1	0.8	2.3	1.3	2.6	1.4	1.6	0.9
Misc	12.9	10.6	10.3	7.7	13.1	10.2	13.0	9.5	12.2	10.8	9.6	7.5	11.8	9.6	12.0	8.6	13.6	10.4	10.7	7.7	13.9	10.4	13.4	

Table A 2:Mean budget shares from the income and expenditure surveys (IES) 1995, 2000, 2005-06 and 2010-11 disaggregated by tobacco use and tercile

	Full sample											
	IES 1995			IES 2000			IES 2005-06			IES 2010-11		
Category	T1 (%)	T2 (%)	T3 (%)	T1 (%)	T2 (%)	T3 (%)	T1 (%)	T2 (%)	T3 (%)	T1 (%)	T2 (%)	T3 (%)
Food	33.9	32.4	22.0	59.3	47.1	29.8	36.1	30.6	16.1	35.2	27.5	13.1
Tob	1.7	1.8	1.1	1.9	2.1	1.6	1.7	1.4	0.9	1.2	1.0	0.7
Alc	1.0	1.4	0.9	1.8	2.4	1.8	0.9	0.9	0.5	1.4	1.1	0.6
Clth & shs	5.0	7.5	5.9	5.8	7.7	6.4	9.3	9.0	6.2	8.2	8.3	5.4
Educ	0.5	0.8	1.5	1.5	1.6	3.1	1.6	1.5	2.4	0.9	1.4	2.7
Hous	48.8	34.6	29.3	15.8	15.8	20.3	20.4	19.1	24.7	23.8	26.1	32.1
Trans	0.6	1.9	7.8	3.9	6.3	12.1	8.9	11.3	15.1	9.7	12.7	16.0
Comm	0.1	0.2	0.3	0.9	1.7	3.2	3.2	3.8	4.3	3.4	3.6	3.3
Health	0.6	1.7	3.7	0.9	1.0	1.4	1.7	1.7	1.7	1.5	1.5	1.5
HH gds	2.4	5.2	5.2	1.4	3.2	4.1	7.0	8.0	7.5	5.9	5.6	5.0
Recr	0.5	1.4	2.7	1.1	1.5	3.0	1.3	1.9	3.1	0.9	1.2	2.2
Misc	5.1	11.1	19.5	5.6	9.6	13.0	8.0	10.7	17.5	8.0	10.0	17.5
Tobacco user												
Food	32.0	32.3	22.8	55.0	44.6	30.1	35.0	31.3	17.6	33.7	28.1	13.6
Tob	3.7	3.7	2.3	6.6	5.7	4.4	4.5	3.6	2.6	5.4	4.1	2.6
Alc	1.7	2.1	1.3	4.5	4.8	3.0	1.9	1.9	0.9	4.6	3.4	1.6
Clth & shs	5.1	7.5	5.9	5.2	7.1	6.2	9.1	9.0	6.4	7.0	7.8	5.3
Educ	0.4	0.7	1.3	0.9	1.1	2.3	1.0	1.2	1.9	0.4	1.0	2.4
Hous	47.8	33.4	28.3	15.7	15.0	20.6	20.1	17.9	23.9	23.0	22.9	31.2
Trans	0.5	1.7	7.7	3.3	5.5	10.8	8.3	10.2	14.9	8.2	11.5	14.8
Comm	0.0	0.1	0.3	0.7	1.4	2.9	2.9	3.5	4.2	3.0	3.4	3.3
Health	0.5	1.7	3.8	0.8	1.0	1.2	1.7	1.8	1.6	1.5	1.6	1.4
HH gds	2.4	4.7	5.0	1.3	3.1	3.9	6.6	7.7	6.9	5.3	5.7	4.7
Recr	0.5	1.4	2.7	0.8	1.4	2.9	1.4	1.9	3.2	0.8	1.3	2.3
Misc	5.5	10.6	18.6	5.1	9.2	11.8	7.6	10.1	16.0	7.0	9.2	16.7
Non-tobacco users												
Food	35.5	32.6	21.2	61.0	48.4	29.7	36.7	30.2	15.3	35.6	27.3	13.0
Tob	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Alc	0.4	0.7	0.6	0.8	1.0	1.1	0.3	0.3	0.3	0.5	0.4	0.3
Clth & shs	5.0	7.5	6.0	6.0	8.1	6.6	9.5	9.0	6.1	8.5	8.4	5.5
Educ	0.5	0.9	1.6	1.8	1.9	3.6	1.9	1.7	2.7	1.0	1.5	2.8
Hous	49.7	35.8	30.2	15.9	16.3	20.1	20.5	19.9	25.2	24.0	27.0	32.3
Trans	0.6	2.0	8.0	4.1	6.7	12.9	9.2	12.0	15.2	10.2	13.1	16.4
Comm	0.1	0.2	0.3	1.0	1.9	3.4	3.4	4.0	4.3	3.5	3.6	3.3
Health	0.6	1.7	3.6	1.0	1.1	1.5	1.7	1.7	1.7	1.5	1.5	1.5
HH gds	2.3	5.7	5.4	1.4	3.3	4.3	7.3	8.2	7.8	6.1	5.6	5.1
Recr	0.5	1.4	2.7	1.2	1.5	3.1	1.3	2.0	3.1	0.9	1.2	2.1
Misc	4.8	11.5	20.3	5.8	9.9	13.8	8.3	11.1	18.3	8.3	10.3	17.8

Table A 3: Detailed budget share for food

	Full sample				Tobacco user				Non-tobacco users			
	IES 1995	IES 2000	IES 2005-06	IES 2010-11	IES 1995	IES 2000	IES 2005-06	IES 2010-11	IES 1995	IES 2000	IES 2005-06	IES 2010-11
Bread	20.9%	23.2%	21.0%	24.2%	20.2%	21.4%	20.7%	22.2%	21.5%	24.3%	21.2%	24.9%
Coffee & Tea	3.8%	3.0%	1.8%	1.9%	3.9%	3.0%	1.9%	2.2%	3.6%	2.9%	1.8%	1.8%
Fish	3.3%	3.3%	3.5%	2.3%	3.4%	3.5%	3.7%	2.5%	3.3%	3.1%	3.4%	2.2%
Food_n.e.c.	5.4%	5.1%	4.8%	5.5%	5.2%	5.0%	5.0%	6.0%	5.5%	5.2%	4.6%	5.4%
Fruits	3.8%	3.8%	2.5%	2.5%	3.6%	3.5%	1.9%	2.2%	4.1%	3.9%	2.9%	2.6%
Meat	25.0%	22.5%	25.3%	25.3%	25.9%	24.0%	27.0%	27.9%	24.2%	21.6%	24.3%	24.5%
Milk, cheese & eggs	8.1%	8.3%	9.6%	9.6%	8.1%	8.3%	9.3%	9.6%	8.0%	8.3%	9.8%	9.6%
Water & soft drinks	4.6%	5.5%	5.5%	5.4%	4.6%	5.7%	5.4%	5.7%	4.6%	5.3%	5.6%	5.4%
Oils & Fats	4.6%	3.8%	3.3%	4.0%	4.6%	3.7%	3.2%	3.6%	4.5%	3.9%	3.3%	4.2%
Sugar	7.4%	5.7%	5.0%	4.9%	7.3%	5.5%	5.2%	5.2%	7.4%	5.7%	4.9%	4.8%
Vegetables	9.2%	9.8%	9.5%	9.8%	9.1%	9.5%	9.4%	9.6%	9.4%	10.0%	9.6%	9.9%
Unclassified	3.9%	6.1%	8.2%	4.4%	4.1%	6.8%	7.4%	3.3%	3.8%	5.6%	8.6%	4.8%

Table A 4: Detailed budget share for housing

	Full sample				Tobacco user				Non-tobacco users			
	IES 1995	**IES 2000**	**IES 2005-06**	**IES 2010-11**	**IES 1995**	**IES 2000**	**IES 2005-06**	**IES 2010-11**	**IES 1995**	**IES 2000**	**IES 2005-06**	**IES 2010-11**
Actual rent	8.1%	13.1%	16.7%	13.7%	9.4%	15.0%	19.2%	18.3%	7.0%	12.2%	15.4%	12.3%
Imputed rent	35.4%	2.1%	54.8%	66.4%	31.8%	2.5%	50.6%	59.8%	38.5%	1.8%	57.0%	68.3%
Electricity	6.6%	10.3%	8.4%	7.6%	6.9%	10.7%	8.7%	8.6%	6.4%	10.0%	8.2%	7.3%
Gas	2.7%	0.4%	0.3%	0.1%	2.4%	0.4%	0.2%	0.1%	2.9%	0.4%	0.3%	0.2%
Liquid fuels	7.5%	2.2%	1.4%	0.5%	7.4%	2.3%	1.7%	0.6%	7.6%	2.1%	1.2%	0.5%
Household maintenance	20.9%	58.6%	3.0%	1.6%	23.3%	55.0%	3.3%	1.8%	18.9%	60.6%	2.9%	1.6%
Household services	4.2%	5.7%	9.3%	6.4%	4.1%	5.2%	10.1%	6.4%	4.3%	5.9%	8.9%	6.4%
Refuse	0.7%	0.8%	0.9%	0.5%	0.7%	0.8%	1.0%	0.6%	0.7%	0.8%	0.8%	0.5%
Sanitation	1.0%	0.7%	1.0%	0.5%	1.1%	0.7%	1.1%	0.6%	1.0%	0.7%	1.0%	0.4%
Solid fuels	10.4%	1.4%	1.2%	0.3%	10.5%	1.4%	1.4%	0.3%	10.4%	1.3%	1.2%	0.3%
Water	2.4%	4.8%	2.9%	2.4%	2.5%	6.1%	2.8%	2.8%	2.4%	4.2%	3.0%	2.2%

Table A 5: Detailed budget share for transport

	Full sample				Tobacco user				Non-tobacco users			
	IES 1995	**IES 2000**	**IES 2005-06**	**IES 2010-11**	**IES 1995**	**IES 2000**	**IES 2005-06**	**IES 2010-11**	**IES 1995**	**IES 2000**	**IES 2005-06**	**IES 2010-11**
Bicycles	0.3%	0.2%	0.1%	0.0%	0.4%	0.3%	0.1%	0.0%	0.3%	0.1%	0.1%	0.0%
Fuel	31.1%	30.2%	23.4%	27.3%	29.9%	30.1%	22.9%	30.1%	32.2%	30.3%	23.7%	26.5%
Vehicle maintenance	4.3%	5.4%	3.1%	2.6%	3.6%	5.2%	2.9%	2.9%	5.0%	5.5%	3.2%	2.6%
Motorcars	49.2%	34.3%	37.4%	45.0%	52.0%	34.5%	38.2%	38.9%	46.9%	34.2%	37.1%	46.6%
Motorcycles	0.6%	0.4%	0.4%	0.5%	0.4%	0.3%	0.8%	1.8%	0.8%	0.4%	0.2%	0.1%
Transport services	2.9%	2.8%	8.4%	3.1%	2.9%	3.0%	6.4%	3.5%	3.0%	2.7%	9.3%	3.0%
Air transport	1.5%	1.0%	1.4%	1.4%	1.4%	0.9%	1.0%	1.2%	1.6%	1.1%	1.6%	1.5%
Rail transport	0.3%	1.2%	1.1%	0.5%	0.4%	1.4%	1.3%	0.7%	0.3%	1.2%	1.0%	0.4%
Water transport	3.8%	19.3%	20.6%	16.2%	3.7%	19.0%	22.2%	17.2%	3.9%	19.5%	19.9%	16.0%
Spares	0.2%	0.0%	0.0%	0.1%	0.2%	0.0%	0.0%	0.1%	0.2%	0.0%	0.0%	0.1%
Other	5.7%	5.1%	4.0%	3.3%	5.4%	5.3%	4.1%	3.7%	5.9%	5.1%	4.0%	3.1%

Table A 6: Auxiliary OLS regression for validity of inclusivity criteria of adult-sex ratio in 1995, 2000, 2005/06 and 2010/11 samples

VARIABLES	(1) **IES 1995**	(2) **IES 2000**	(3) **IES 2005-06**	(4) **IES 2010-11**
adsex_ratio	34.04***	48.19***	36.56***	28.95***
	(3.442)	(3.849)	(3.131)	(3.565)
ln_hhs	26.51***	21.19***	14.38***	8.905***
	(1.861)	(2.163)	(1.833)	(1.963)
no_kid	-7.273***	-6.729***	-5.354***	-1.582*
	(0.751)	(1.222)	(0.836)	(0.954)
ageadult_mean	-0.646***	-0.509***		-0.206***
	(0.0639)	(0.0768)		(0.0635)
Male	25.80***	5.565**	-0.150	9.727***
	(1.626)	(2.802)	(2.376)	(2.035)
Coloured	48.78***	66.35***	58.17***	77.31***
	(3.332)	(6.099)	(4.657)	(4.778)
Indian	24.30***	51.44***	37.88***	50.94***
	(4.836)	(8.685)	(7.371)	(12.28)
White	88.44***	105.1***	65.51***	73.34***
	(3.292)	(6.729)	(5.394)	(6.046)
Urban	12.62***	13.07***	7.783***	2.150
	(1.612)	(2.233)	(1.739)	(1.341)
EC	-7.916**	-46.44***	-1.379	-52.71***
	(3.615)	(6.645)	(5.282)	(5.174)
NC	-16.29***	-46.21***	-5.113	-56.67***
	(4.452)	(7.634)	(4.937)	(5.950)
FS	-10.67***	-48.59***	-0.996	-36.42***
	(3.794)	(6.496)	(5.641)	(5.238)
KZN	5.940	-45.33***	-12.90***	-53.21***
	(3.959)	(6.721)	(4.689)	(5.237)
NW	4.778	-27.49***	-3.513	-46.99***
	(4.241)	(6.944)	(5.292)	(5.462)
GA	19.93***	-28.31***	5.631	-37.93***
	(4.774)	(6.983)	(5.568)	(5.617)
MP	-4.241	-43.21***	-9.671*	-43.71***
	(4.378)	(6.624)	(5.153)	(5.335)
LMP	-3.711	-44.69***	-10.11*	-53.23***
	(4.093)	(6.861)	(5.329)	(5.052)
Constant	4.815	43.50***	-2.100	47.33***
	(5.334)	(7.325)	(4.943)	(6.184)
Observations	28,517	19,751	15,967	24,826
R-squared	0.104	0.100	0.084	0.107

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Reference groups are as follows: female, black African, rural and WC (Western Cape). Province abbreviations are as follows: EC-Eastern Cape, NC – Northern Cape, FS- Free State, KZN- Kwazulu-Natal, NW-Northwestern, GA-Gauteng, MP -Mpumalanga, LMP-Limpopo.

Table A 7: Auxiliary OLS regression for validity of inclusivity criteria of adult-sex ratio in 1995, 2000, 2005/06 and 2010/11 samples

VARIABLES	**IES 1995**	**IES 2000**	**IES 2005-06**	**IES 2010-11**	**IES 1995**	**IES 2000**	**IES 2005-06**	**IES 2010-11**
ln_asset	-0.0818***	-0.162***	-0.204***	-0.180***	0.521*	0.267***	0.250***	0.306***
	(0.0238)	(0.0112)	(0.00704)	(0.00985)	(0.301)	(0.0598)	(0.0303)	(0.0654)
ln_hhs	0.476***	0.569***	0.344***	0.252***	8.477***	7.062***	5.279***	3.821***
	(0.0736)	(0.0226)	(0.0128)	(0.0195)	(1.711)	(0.338)	(0.216)	(0.340)
no_kid	-0.0740**	-0.0720***	-0.00331	0.0149	-1.316*	-0.791***	-0.0231	0.187
	(0.0327)	(0.00764)	(0.00481)	(0.00937)	(0.713)	(0.121)	(0.0822)	(0.165)
ageadult_mean	-0.00359	0.00342***		-0.00233***	-0.0901	0.0388***		-0.0333***
	(0.00220)	(0.000762)		(0.000704)	(0.0573)	(0.0107)		(0.0116)
Male	0.199***	0.0234	0.00661	0.0352**	4.829***	0.560**	0.0405	0.538*
	(0.0688)	(0.0178)	(0.0112)	(0.0177)	(1.618)	(0.273)	(0.190)	(0.309)
Coloured	0.0447	0.0163	-0.234***	-0.0605	1.662	0.929	-3.095***	-0.991
	(0.121)	(0.0663)	(0.0248)	(0.0457)	(2.618)	(1.099)	(0.454)	(0.826)
Indian	0.221	0.408*	0.113	0.684***	8.073	9.977***	1.181	13.46***
	(0.139)	(0.222)	(0.127)	(0.120)	(4.989)	(1.916)	(2.835)	(2.707)
White	0.723***	0.715	-0.107	0.679***	16.15***	14.97	-7.367	15.27***
	(0.0823)	(0.471)	(0.217)	(0.203)	(1.813)	(9.775)	(8.111)	(5.079)
Urban	-0.176***	0.200***	0.0565***	0.0454**	-4.213***	3.497***	0.771***	0.903**
	(0.0652)	(0.0256)	(0.0135)	(0.0214)	(1.387)	(0.404)	(0.237)	(0.374)
EC	0.145	-0.415***	-0.0310	-0.0788	3.188	-5.909***	-1.196**	-1.786*
	(0.119)	(0.0491)	(0.0296)	(0.0523)	(2.708)	(0.912)	(0.535)	(0.924)
NC	-0.0933	-0.643***	-0.167***	-0.234***	0.189	-6.311***	-3.277***	-3.597***
	(0.160)	(0.0933)	(0.0312)	(0.0580)	(4.093)	(1.839)	(0.593)	(0.998)
FS	-0.564***	-0.734***	-0.129***	-0.200***	-13.24***	-9.442***	-2.031***	-3.280***
	(0.122)	(0.0547)	(0.0330)	(0.0617)	(2.911)	(0.990)	(0.591)	(1.100)
KZN	-0.00683	-0.303***	-0.00247	-0.00596	1.045	-4.242***	-1.510***	-0.806
	(0.111)	(0.0472)	(0.0287)	(0.0517)	(2.810)	(0.898)	(0.520)	(0.922)
NW	0.0664	-0.287***	-0.125***	-0.176***	4.817	-1.074	-2.295***	-3.066***
	(0.138)	(0.0561)	(0.0340)	(0.0565)	(3.279)	(1.048)	(0.621)	(0.997)
GA	0.0959	-0.110**	0.159***	0.0507	3.397	-1.014	2.218***	0.0716
	(0.112)	(0.0531)	(0.0332)	(0.0517)	(2.767)	(1.022)	(0.598)	(0.913)
MP	0.0564	-0.214***	-0.0462	-0.0376	-1.430	-0.973	-1.484**	-0.869
	(0.131)	(0.0496)	(0.0327)	(0.0555)	(3.129)	(0.976)	(0.590)	(1.003)
LMP	0.465***	-0.437***	-0.0589*	-0.128**	10.68***	-5.662***	-1.334**	-2.361**
	(0.155)	(0.0519)	(0.0316)	(0.0536)	(3.497)	(0.972)	(0.578)	(0.952)
Constant	8.428***	6.599***	7.048***	7.302***	68.01***	42.41***	47.94***	51.51***
	(0.188)	(0.0638)	(0.0305)	(0.0605)	(4.570)	(1.062)	(0.547)	(1.075)
Observations	942	9,091	12,062	5,795	648	5,953	9,058	4,298
R-squared	0.253	0.164	0.230	0.195	0.289	0.159	0.180	0.119

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Reference groups are as follows: female, black African, rural and WC (Western Cape). Province abbreviations are as follows: EC-Eastern Cape, NC – Northern Cape, FS- Free State, KZN- Kwazulu-Natal, NW-Northwestern, GA-Gauteng, MP -Mpumalanga, LMP-Limpopo.

Table A 8: Statistics for instrument validity – full sample, urban and rural samples

		Full sample			Urban			Rural		
Survey	**Variable**	$p_t t$	**lnM**	**lnM2**	$p_t t$	**lnM**	**lnM2**	$p_t t$	**lnM**	**lnM2**
1995	Shea Partial R^2	0.0007	0.0892	0.0928	0.0004	0.0704	0.0740	0.0007	0.1168	0.1190
	Partial R^2	0.0043	0.2724	0.2762	0.0035	0.2496	0.2529	0.0007	0.2614	0.2681
	F-statistic	42.6	3679.6	3751.8	20.0	1868.2	1901.5	3.1	1488.0	1540.3
	P-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0248	0.0000	0.0000
2000	Shea Partial R^2	0.0001	0.0011	0.0009	0.0000	0.0005	0.0004	0.0000	0.0003	0.0003
	Partial R^2	0.0092	0.2444	0.2626	0.0077	0.2455	0.2634	0.0043	0.1981	0.2063
	F-statistic	59.8	2089.9	2301.3	31.9	1345.6	1478.2	10.0	572.7	602.8
	P-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2005-06	Shea Partial R^2	0.0015	0.0428	0.0392	0.0013	0.0361	0.0339	0.0022	0.0447	0.0442
	Partial R^2	0.0120	0.1897	0.1954	0.0126	0.2133	0.2184	0.0034	0.1199	0.1219
	F-statistic	64.1	1238.8	1285.0	39.4	834.0	859.4	7.5	301.0	306.7
	P-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000
2010-11	Shea Partial R^2	0.0000	0.0019	0.0015	0.0000	0.0009	0.0006	0.0011	0.0050	0.0053
	Partial R^2	0.0117	0.1307	0.1369	0.0117	0.1467	0.1523	0.0020	0.0595	0.0623
	F-statistic	98.2	1242.8	1311.7	62.7	914.2	955.1	5.9	186.1	195.6
	P-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0005	0.0000	0.0000

An *F-statistic* of less than 10 is indicative of a weak instrument, in the case of a single endogenous regressor. For equations with more than one endogenous regressor, *Shea's partial R^2* can be used instead of the F-critical value. In the latter case, while there is no consensus on how low of a value indicates a problem, higher values would indicate relatively stronger instruments and would likely exhibit less bias.

Table A 9: Statistics for instrument validity – Tercile sub-samples

		Tercile 1			Tercile 2			Tercile 3		
Survey	**Variable**	$p_t t$	**lnM**	**lnM2**	$p_t t$	**lnM**	**lnM2**	$p_t t$	**lnM**	**lnM2**
1995	Shea Partial R2	0.0003	0.0004	0.0003	0.0012	0.0005	0.0004	0.0002	0.0572	0.0597
	Partial R2	0.0061	0.0345	0.0335	0.008	0.0289	0.0286	0.0002	0.1383	0.1445
	F-statistic	20.8	121.36	117.95	26.44	97.47	96.57	0.57	502.94	529.18
	P-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.6353	0.0000	0.0000
2000	Shea Partial	0.0014	0.0189	0.0212	0.0001	0.0002	0.0002	0.0002	0.0572	0.0597
	Partial R2	0.0038	0.0354	0.046	0.0049	0.0425	0.045	0.0002	0.1383	0.1445
	F-statistic	8.68	83.41	109.47	10.87	97.57	103.45	0.57	502.94	529.18
	P-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.6353	0.0000	0.0000
2005_06	Shea Partial	0.0025	0.0054	0.005	0.0049	0.0002	0.0002	0.0002	0.008	0.0074
	Partial R2	0.0025	0.0253	0.0236	0.0065	0.012	0.0116	0.0028	0.0829	0.0846
	F-statistic	4.91	51.34	47.97	11.49	21.13	20.58	4.39	140.1	143.26
	P-value	0.0021	0.0000	0.0000	0.0000	0.0000	0.0000	0.0043	0.0000	0.0000
2010_11	Shea Partial	0.0003	0.0001	0.0001	0.0004	0.0004	0.0004	0.0001	0.0084	0.0084
	Partial R2	0.0015	0.0084	0.0087	0.0005	0.0072	0.0073	0.0013	0.0591	0.0613
	F-statistic	4.41	25.72	26.56	1.51	20.56	20.79	3.27	151.89	157.97
	P-value	0.0042	0.0000	0.0000	0.2099	0.0000	0.0000	0.0204	0.0000	0.0000

An *F-statistic* of less than 10 is indicative of a weak instrument, in the case of a single endogenous regressor. For equations with more than one endogenous regressor, *Shea's partial R^2* can be used instead of the F-critical value. In the latter case, while there is no consensus on how low of a value indicates a problem, higher values would indicate relatively stronger instruments and would likely exhibit less bias.

Table A 10: Three Stage-Least square (3SLS) Regression 1995 IES Full Sample (Equation 5 specification)

VARIABLES	(1) bsfoodexp	(2) bstobexp	(3) bsalcexp	(4) bsclthsexp	(5) bseduexp	(6) bshseexp	(7) bstransexp	(8) bscommexp	(9) bshealthexp	(10) bshhgdsexp	(11) bsrecexp
Pq	0.000112	0.000342***	0.000300***	1.82e-05	3.78e-05	-0.00120***	7.74e-05	-2.41e-06	-2.58e-05	-0.000165***	4.59e-05*
	(0.000111)	(2.58e-05)	(3.60e-05)	(5.11e-05)	(2.42e-05)	(0.000201)	(5.36e-05)	(3.25e-06)	(3.40e-05)	(5.96e-05)	(2.51e-05)
lnM	0.321***	0.0127*	0.0303***	0.205***	-0.00304	-0.910***	-0.107***	0.00568***	0.0837***	0.230***	0.0189***
	(0.0183)	(0.00697)	(0.00679)	(0.00735)	(0.00340)	(0.0380)	(0.0124)	(0.000454)	(0.00463)	(0.00895)	(0.00429)
lnM2	-0.0230***	-0.00158***	-0.00212***	-0.0114***	0.000408**	0.0461***	0.00794***	-0.000300***	-0.00427***	0.0117***	-0.000611**
	(0.00104)	(0.000412)	(0.000404)	(0.000418)	(0.000201)	(0.00223)	(0.000758)	(2.69e-05)	(0.000271)	(0.000515)	(0.000252)
ln_hhs	-0.0269***	-0.00576***	-0.0129***	0.00183	-0.00511***	0.0830***	-0.0148***	-0.000135*	0.00938***	-0.00888***	-0.00483***
	(0.00262)	(0.000631)	(0.000960)	(0.00117)	(0.000530)	(0.00445)	(0.00127)	(7.04e-05)	(0.000880)	(0.00133)	(0.000658)
no_kid	-0.00132	0.00103***	0.00200***	0.00170***	-0.000724***	-0.0128***	0.00262***	-3.75e-05	0.00163***	-0.000283	0.000356
	(0.00101)	(0.000211)	(0.000314)	(0.000497)	(0.000205)	(0.00174)	(0.000470)	(2.59e-05)	(0.000294)	(0.000514)	(0.000238)
ageadult_mean	-0.000142	9.61e-05***	8.80e-05***	0.000528***	-0.000122***	0.00126***	-0.000430***	1.72e-05***	-9.21e-05***	0.000493***	0.000170***
	(0.000106)	(2.45e-05)	(3.29e-05)	(4.63e-05)	(1.87e-05)	(0.000195)	(5.21e-05)	(2.90e-06)	(3.08e-05)	(5.44e-05)	(2.22e-05)
Male	-0.00591	1.31e-05	-0.000336	0.00449**	-0.00236***	0.00224	0.00612***	-0.000175*	0.00569***	0.00667***	-5.31e-05
	(0.00382)	(0.000854)	(0.00117)	(0.00193)	(0.000905)	(0.00683)	(0.00176)	(0.000105)	(0.00107)	(0.00216)	(0.000865)
Coloured	-0.0278***	-0.00775***	-0.0127***	-0.0139***	-0.00671***	-0.0750***	-0.00288	-0.000560***	-0.00350**	-0.0101***	-0.000585
	(0.00577)	(0.00147)	(0.00203)	(0.00274)	(0.00123)	(0.0110)	(0.00283)	(0.000182)	(0.00173)	(0.00314)	(0.00131)
Indian	-0.0191***	-0.00337**	-0.0119***	-0.0233***	-0.00461***	-0.126***	-0.0301***	-0.00148***	-0.00114	-0.0369***	-0.00397***
	(0.00437)	(0.00133)	(0.00164)	(0.00205)	(0.00121)	(0.00875)	(0.00326)	(0.000165)	(0.00153)	(0.00259)	(0.00133)
White	-0.0148**	-0.0166***	-0.0246***	-0.0421***	-0.00374**	-0.227***	-0.0297***	-0.00148***	-0.0173***	-0.0450***	-0.00134
	(0.00750)	(0.00199)	(0.00268)	(0.00334)	(0.00163)	(0.0141)	(0.00387)	(0.000239)	(0.00240)	(0.00402)	(0.00168)
Urban	-0.00936***	-0.00198***	-0.00500***	-0.00399***	-0.00138***	-0.0162***	-0.000357	-0.000906***	-0.00413***	-0.00863***	-0.00258***
	(0.00208)	(0.000478)	(0.000695)	(0.000961)	(0.000428)	(0.00371)	(0.000997)	(5.21e-05)	(0.000633)	(0.00117)	(0.000472)
EC	-0.0228***	-0.00236***	-0.00365***	-0.00895***	-0.000701	-0.0295***	-0.00151	-0.000152	-0.00414***	-0.00200	-0.00447***

	(0.00325)	(0.000902)	(0.00119)	(0.00150)	(0.000730)	(0.00617)	(0.00189)	(9.86e-05)	(0.000943)	(0.00177)	(0.000881)
NC	-0.00802*	0.00182	0.000125	-0.00382*	-0.00108	-0.0444***	-4.14e-05	8.51e-05	-0.00292**	-0.00290	-0.00329***
	(0.00466)	(0.00130)	(0.00150)	(0.00198)	(0.000665)	(0.00809)	(0.00259)	(0.000135)	(0.00122)	(0.00207)	(0.000998)
FS	-0.0608***	-0.00207**	-0.00244*	-0.00466***	-0.000942	-0.0252***	-0.00395*	-0.000241**	-0.00723***	-0.0128***	-0.000997
	(0.00356)	(0.000996)	(0.00127)	(0.00172)	(0.000701)	(0.00686)	(0.00206)	(0.000120)	(0.00108)	(0.00189)	(0.000855)
KZN	-0.00909***	-0.00428***	-0.00227*	-0.000781	-0.000495	-0.00351	-0.000768	8.53e-06	-0.00607***	-0.00159	-0.00350***
	(0.00341)	(0.000976)	(0.00128)	(0.00153)	(0.000779)	(0.00670)	(0.00208)	(0.000113)	(0.00101)	(0.00183)	(0.000908)
NW	-0.0396***	-0.00439***	-0.00163	-0.0131***	-0.00123	-0.0404***	-0.00341	-4.01e-05	-0.00680***	-0.00404**	-0.00384***
	(0.00374)	(0.00114)	(0.00145)	(0.00183)	(0.000819)	(0.00758)	(0.00230)	(0.000128)	(0.00111)	(0.00196)	(0.000935)
GA	-0.000211	-0.00421***	-0.00175	-0.0128***	-0.00138	-0.000316	-0.00306	-0.000421***	-0.0129***	-0.00822***	-0.00291***
	(0.00384)	(0.00125)	(0.00156)	(0.00181)	(0.000919)	(0.00793)	(0.00242)	(0.000139)	(0.00130)	(0.00204)	(0.00101)
MP	-0.0172***	-0.00443***	-0.000852	-0.00295	-0.00181**	-0.0104	-0.00494**	3.47e-05	-0.00892***	-0.0213***	-0.00220**
	(0.00369)	(0.00119)	(0.00151)	(0.00187)	(0.000732)	(0.00746)	(0.00234)	(0.000113)	(0.00118)	(0.00218)	(0.00102)
LMP	-0.00548	-0.00217**	-0.00321**	5.34e-05	-8.80e-05	-0.00234	-0.00414*	3.42e-05	-0.00367***	-0.00544**	-0.00492***
	(0.00425)	(0.00108)	(0.00137)	(0.00203)	(0.000910)	(0.00735)	(0.00228)	(0.000124)	(0.00114)	(0.00226)	(0.00103)
Constant	0.785***	0.00894	0.0919***	0.825***	0.00398	4.601***	0.376***	-0.0257***	0.379***	1.011***	0.0885***
	(0.0788)	(0.0289)	(0.0279)	(0.0317)	(0.0141)	(0.159)	(0.0506)	(0.00187)	(0.0196)	(0.0379)	(0.0179)
Observations	29,512	29,512	29,512	29,512	29,512	29,512	29,512	29,512	29,512	29,512	29,512

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Reference groups are as follows: female, black African, rural, and WC (Western Cape). Province abbreviations are as follows: EC - Eastern Cape, NC - Northern Cape, FS - Free State, KZN - Kwazulu-Natal, NW - North Western, GA - Gauteng, MP - Mpumalanga, LMP - Limpopo.

Table A 11: Three Stage-Least square (3SLS) Regression 1995 IES Urban Sample (Equation 5 specification)

VARIABLES	(1) bsfoodexp	(2) bstobexp	(3) bsalcexp	(4) bsclthsexp	(5) bseduexp	(6) bshseexp	(7) bstransexp	(8) bscommexp	(9) bshealthexp	(10) bshhgdsexp	(11) bsrecexp
pq	0.000121	0.000318***	0.000319***	3.19e-05	4.38e-05	-0.00129***	3.42e-05	-2.81e-06	-0.000110**	-6.40e-05	3.52e-05
	(0.000124)	(3.06e-05)	(4.71e-05)	(5.91e-05)	(3.15e-05)	(0.000229)	(6.93e-05)	(4.59e-06)	(4.35e-05)	(7.04e-05)	(3.21e-05)
lnM	0.195***	0.00170	0.0211**	0.206***	-0.00426	-0.759***	-0.119***	0.00776***	0.111***	0.240***	0.0200***
	(0.0225)	(0.00925)	(0.00923)	(0.00979)	(0.00480)	(0.0486)	(0.0165)	(0.000725)	(0.00680)	(0.0110)	(0.00546)
lnM2	-0.0160***	-0.000986*	-0.00159***	-0.0115***	0.000456	0.0381***	0.00883***	-0.000415***	-0.00568***	-0.0125***	-0.000617**
	(0.00125)	(0.000535)	(0.000542)	(0.000551)	(0.000279)	(0.00281)	(0.000976)	(4.20e-05)	(0.000386)	(0.000620)	(0.000313)
ln_hhs	0.0362***	-0.00591***	-0.0145***	-0.00385**	-0.00586***	0.0709***	-0.0163***	1.41e-05	-0.00534***	-0.00729***	-0.00525***
	(0.00378)	(0.000899)	(0.00152)	(0.00181)	(0.000918)	(0.00683)	(0.00212)	(0.000139)	(0.00129)	(0.00198)	(0.00105)
no_kid	0.000645	0.00128***	0.00243***	0.00198***	-0.00103***	-0.0150***	0.00249***	-5.36e-05	0.00118***	0.000513	0.000100
	(0.00142)	(0.000338)	(0.000489)	(0.000731)	(0.000331)	(0.00259)	(0.000795)	(4.98e-05)	(0.000457)	(0.000785)	(0.000361)
ageadult_mean	-3.29e-05	0.000123***	0.000185***	0.000429***	-0.000150***	0.000723**	-0.000557***	3.14e-05***	-8.97e-05*	0.000379***	0.000183***
	(0.000145)	(3.74e-05)	(5.59e-05)	(6.90e-05)	(3.33e-05)	(0.000287)	(8.60e-05)	(5.15e-06)	(5.21e-05)	(8.58e-05)	(3.71e-05)
male	-0.00145	-0.00139	-0.00246	-0.000419	-0.00348***	-0.0249***	-0.00870***	-0.000360**	-0.00939***	-0.000909	-0.000550
	(0.00488)	(0.00116)	(0.00171)	(0.00247)	(0.00130)	(0.00880)	(0.00251)	(0.000163)	(0.00156)	(0.00297)	(0.00121)
Coloured	-0.0215***	-0.00798***	-0.0185***	-0.0192***	-0.00797***	-0.112***	-0.00373	-0.000564*	-0.00746***	-0.0212***	-0.00110
	(0.00737)	(0.00185)	(0.00287)	(0.00363)	(0.00183)	(0.0138)	(0.00417)	(0.000292)	(0.00253)	(0.00417)	(0.00185)
Indian	-0.0227***	-0.00264*	-0.0140***	-0.0288***	-0.00439***	-0.130***	-0.0276***	-0.00133***	-0.000282	-0.0439***	-0.00249*
	(0.00466)	(0.00141)	(0.00193)	(0.00248)	(0.00142)	(0.00987)	(0.00360)	(0.000197)	(0.00179)	(0.00265)	(0.00147)
White	-0.0150	-0.0154***	-0.0293***	-0.0433***	-0.00359	-0.239***	-0.0267***	-0.00122***	-0.0233***	-0.0504***	-0.000333
	(0.00963)	(0.00246)	(0.00386)	(0.00464)	(0.00239)	(0.0182)	(0.00555)	(0.000385)	(0.00343)	(0.00545)	(0.00242)
EC	-0.0135***	-0.000722	-0.000645	-0.00780***	-0.00102	-0.00205	-0.000367	-0.000153	-0.00540***	-0.00791***	-0.00299***
	(0.00370)	(0.00100)	(0.00146)	(0.00182)	(0.000906)	(0.00736)	(0.00217)	(0.000127)	(0.00133)	(0.00220)	(0.00100)
NC	-0.00629	-0.000274	-0.00236	-0.00958***	-0.00163**	-0.0307***	-0.00232	2.30e-05	-0.00267	-0.00474*	-0.00217*

	(0.00506)	(0.00123)	(0.00180)	(0.00225)	(0.000815)	(0.00997)	(0.00280)	(0.000174)	(0.00172)	(0.00245)	(0.00127)
FS	-0.0318***	-0.00238**	0.00305*	-0.00535***	-0.00173**	-0.0196**	0.00292	0.000164	-0.00395***	-0.00771***	-0.000442
	(0.00395)	(0.00113)	(0.00164)	(0.00184)	(0.000858)	(0.00810)	(0.00237)	(0.000153)	(0.00132)	(0.00223)	(0.00107)
KZN	-0.00649*	-0.00355***	0.000339	-0.00570***	-0.000259	-0.00301	-0.000452	2.06e-06	-0.00908***	-0.00963***	-0.00192*
	(0.00390)	(0.00112)	(0.00164)	(0.00199)	(0.000993)	(0.00815)	(0.00265)	(0.000161)	(0.00148)	(0.00221)	(0.00113)
NW	-0.0231***	-0.00482***	-0.000915	-0.00369	-0.00118	-0.00328	-0.00594*	-0.000201	-0.0120***	-0.00522**	-0.00263**
	(0.00489)	(0.00145)	(0.00217)	(0.00253)	(0.00126)	(0.0104)	(0.00321)	(0.000190)	(0.00180)	(0.00265)	(0.00130)
GA	-0.00611	-0.00361***	-0.000589	-0.0139***	-0.000781	-0.00888	-0.000957	-0.000430**	-0.0136***	-0.00845***	-0.00207*
	(0.00412)	(0.00135)	(0.00193)	(0.00222)	(0.00109)	(0.00917)	(0.00285)	(0.000176)	(0.00154)	(0.00240)	(0.00120)
MP	-0.0271***	-0.00493***	-0.00234	-0.00883***	-0.00121	0.0162	-0.00559*	-0.000299*	-0.00808***	-0.0170***	-0.00316**
	(0.00463)	(0.00154)	(0.00200)	(0.00235)	(0.00105)	(0.0101)	(0.00311)	(0.000160)	(0.00181)	(0.00277)	(0.00128)
LMP	-0.00257	-0.00156	-0.000520	-0.00409	-0.000768	-0.0185	-0.00902**	-0.000533**	-0.00855***	-0.00735*	-0.00340**
	(0.00623)	(0.00164)	(0.00234)	(0.00317)	(0.00171)	(0.0118)	(0.00405)	(0.000260)	(0.00227)	(0.00376)	(0.00169)
Constant	-0.242**	0.0578	-0.0594	-0.823***	0.0136	3.941***	0.424***	-0.0345***	0.503***	1.035***	-0.0934***
	(0.0991)	(0.0394)	(0.0387)	(0.0428)	(0.0203)	(0.208)	(0.0682)	(0.00305)	(0.0294)	(0.0477)	(0.0234)
Observations	16,875	16,875	16,875	16,875	16,875	16,875	16,875	16,875	16,875	16,875	16,875

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Reference groups are as follows: female, black African, rural and WC (Western Cape). Province abbreviations are as follows: EC - Eastern Cape, NC - Northern Cape, FS - Free State, KZN - Kwazulu-Natal, NW - North Western, GA - Gauteng, MP - Mpumalanga, LMP - Limpopo.

Table A 12: Three Stage-Least square (3SLS) Regression 1995 IES Rural Sample (Equation 5 specification)

VARIABLES	(1) bsfoodexp	(2) bstobexp	(3) bsalcexp	(4) bscltheexp	(5) bseduexp	(6) bshseexp	(7) bstransexp	(8) bscommexp	(9) bshealthexp	(10) bshhgdsexp	(11) bsrecexp
pq	-0.000117 (0.000250)	0.000438*** (6.35e-05)	0.000314*** (8.25e-05)	3.18e-05 (0.000101)	2.62e-05 (4.30e-05)	-0.00107*** (0.000393)	0.000135 (9.33e-05)	-6.31e-07 (4.10e-06)	0.000121* (6.41e-05)	-0.000330*** (0.000128)	8.81e-05* (5.05e-05)
lnM	0.441*** (0.0362)	0.00313 (0.0189)	0.0311* (0.0159)	0.208*** (0.0127)	-0.00290 (0.00561)	-0.961*** (0.0710)	-0.0560*** (0.0196)	0.00348*** (0.000602)	0.0603*** (0.0101)	0.205*** (0.0210)	0.0213*** (0.00795)
lnM2	-0.0296*** (0.00207)	-0.00104 (0.00113)	-0.00222** (0.000936)	-0.0116*** (0.000720)	0.000440 (0.000329)	0.0482*** (0.00416)	0.00456*** (0.00120)	-0.000180*** (3.50e-05)	-0.00312*** (0.000592)	-0.00986*** (0.00123)	-0.000903* (0.000465)
ln_hhs	-0.0126*** (0.00367)	-0.00475*** (0.000831)	-0.00984*** (0.00121)	-0.00162 (0.00148)	-0.00428*** (0.000534)	-0.0905*** (0.00550)	-0.0117*** (0.00138)	-0.000251*** (5.61e-05)	-0.0118*** (0.00109)	-0.0101*** (0.00180)	-0.00370*** (0.000738)
no_kid	-0.00246 (0.00161)	-0.00112*** (0.000319)	-0.00170*** (0.000463)	-0.00184*** (0.000683)	-0.000411 (0.000260)	-0.0128*** (0.00239)	-0.00254*** (0.000565)	8.60e-08 (2.29e-05)	-0.00225*** (0.000407)	-0.000743 (0.000774)	-0.000633** (0.000302)
ageadult_mean	-0.000355*** (0.000136)	2.38e-05 (2.93e-05)	2.54e-06 (4.03e-05)	-0.000559*** (5.05e-05)	-6.89e-05*** (1.84e-05)	-0.00171*** (0.000222)	-0.000231*** (5.61e-05)	-1.97e-06 (2.39e-06)	-0.000116*** (3.39e-05)	-0.000471*** (6.87e-05)	-0.000144*** (2.37e-05)
male	-0.00189 (0.00708)	-0.000406 (0.00165)	-0.000359 (0.00218)	-0.00695** (0.00304)	-0.00116 (0.00139)	-0.0195* (0.0108)	-0.00207 (0.00253)	3.51e-06 (0.000116)	-0.000793 (0.00166)	-0.0136*** (0.00358)	-0.000528 (0.00142)
Coloured	-0.0383*** (0.00805)	-0.00528** (0.00227)	-0.000237 (0.00287)	-0.00163 (0.00323)	-0.00291*** (0.000991)	-0.00607 (0.0126)	-0.00559 (0.00401)	-0.000194 (0.000130)	-0.00253 (0.00208)	-0.00435 (0.00391)	-0.000401 (0.00186)
Indian	-0.0428** (0.0181)	-0.00279 (0.00333)	-0.0107** (0.00421)	-0.00757 (0.00730)	-0.00517** (0.00250)	-0.0479* (0.0276)	-0.0400*** (0.0135)	-0.000219*** (0.000501)	-0.00791 (0.00498)	-0.00274 (0.0108)	-0.0102** (0.00435)
White	-0.00321 (0.0100)	-0.0168*** (0.00492)	-0.0156*** (0.00487)	-0.0317*** (0.00422)	-0.00450** (0.00206)	-0.164*** (0.0205)	-0.0525*** (0.00630)	-0.000246*** (0.000241)	-0.0107*** (0.00368)	-0.0404*** (0.00671)	-0.00610** (0.00267)
EC	-0.0622*** (0.0115)	-0.000789 (0.00306)	-0.00993** (0.00423)	-0.00315 (0.00422)	-0.00216 (0.00168)	-0.0657*** (0.0185)	-0.00669 (0.00579)	4.65e-05 (0.000200)	-0.00306 (0.00286)	-0.0111* (0.00586)	-0.00708*** (0.00270)
NC	-0.0346***	-0.00417	-0.00838**	-0.00914**	-0.000692	-0.0380**	-0.00809	9.54e-05	-0.00248	-0.00893*	-0.00671***

	(0.0110)	(0.00335)	(0.00379)	(0.00417)	(0.00138)	(0.0162)	(0.00561)	(0.000177)	(0.00237)	(0.00517)	(0.00229)
FS	-0.136***	0.000427	0.000358	0.0260***	0.00157	-0.0181	0.0122**	0.000248	0.0122***	0.0120**	-0.00277
	(0.0115)	(0.00297)	(0.00405)	(0.00435)	(0.00161)	(0.0184)	(0.00562)	(0.000192)	(0.00304)	(0.00581)	(0.00259)
KZN	-0.0179*	0.00269	0.00757*	0.000109	0.00296*	0.00988	0.00612	3.42e-05	0.00317	-0.0137***	-0.00658**
	(0.0104)	(0.00301)	(0.00398)	(0.00376)	(0.00153)	(0.0169)	(0.00567)	(0.000187)	(0.00258)	(0.00529)	(0.00256)
NW	-0.0888***	0.000433	0.00708*	0.0177***	0.00278	0.0971***	0.00602	-5.49e-05	0.00287	-0.0126**	-0.00605**
	(0.0119)	(0.00307)	(0.00420)	(0.00419)	(0.00180)	(0.0188)	(0.00589)	(0.000208)	(0.00277)	(0.00583)	(0.00271)
GA	-0.0332***	0.00483	0.00915**	0.00359	0.00481***	0.0550***	0.0130*	0.000332	0.0200***	0.00193	-0.00802***
	(0.0108)	(0.00410)	(0.00449)	(0.00395)	(0.00182)	(0.0184)	(0.00729)	(0.000241)	(0.00446)	(0.00584)	(0.00303)
MP	-0.0400***	0.00163	0.00247	0.0193***	-0.000540	0.0202	0.0101*	0.000269	0.00988***	0.0165***	-0.00263
	(0.0112)	(0.00317)	(0.00438)	(0.00419)	(0.00161)	(0.0184)	(0.00576)	(0.000211)	(0.00304)	(0.00603)	(0.00275)
LMP	-0.0283**	0.000825	-0.00730	0.00909*	0.00197	0.0196	0.00966	-2.54e-05	0.00294	-0.00578	-0.00650**
	(0.0127)	(0.00328)	(0.00455)	(0.00482)	(0.00195)	(0.0197)	(0.00621)	(0.000221)	(0.00311)	(0.00659)	(0.00289)
Constant	1.251***	0.0442	-0.0896	0.842***	-0.00281	4.845***	0.172**	-0.0153***	0.264***	0.915***	-0.0923***
	(0.152)	(0.0763)	(0.0651)	(0.0536)	(0.0230)	(0.293)	(0.0794)	(0.00248)	(0.0414)	(0.0859)	(0.0326)
Observations	12,637	12,637	12,637	12,637	12,637	12,637	12,637	12,637	12,637	12,637	12,637

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Reference groups are as follows: female, black African, rural and WC (Western Cape). Province abbreviations are as follows: EC - Eastern Cape, NC - Northern Cape, FS - Free State, KZN - Kwazulu-Natal, NW - North Western, GA - Gauteng, MP - Mpumalanga, LMP - Limpopo.

Table A 13: Three Stage-Least square (3SLS) Regression 1995 IES Full Sample (Equation 7 specification)

VARIABLES	(1) bsfoodexp	(2) bstobexp	(3) bsalcexp	(4) bsclthsexp	(5) bseduexp	(6) bshseexp	(7) bstransexp	(8) bscommexp	(9) bshealthexp	(10) bshhgdsexp	(11) bsrecexp
pq	-0.000201	0.000383***	0.000601***	-0.000187	0.000106	-0.00165**	0.000187	-9.83e-06	-0.000176	-0.000589**	8.67e-05
	(0.000379)	(6.48e-05)	(0.000170)	(0.000162)	(7.11e-05)	(0.000738)	(0.000159)	(9.71e-06)	(0.000112)	(0.000250)	(7.21e-05)
lnM	2.954***	0.466***	0.355***	1.098***	-0.0724***	-5.165***	-0.527***	0.0186***	0.216***	0.829***	-0.0124
	(0.166)	(0.0391)	(0.0688)	(0.0689)	(0.0252)	(0.355)	(0.0720)	(0.00439)	(0.0456)	(0.0944)	(0.0258)
lnM2	-0.169***	-0.0244***	-0.0179***	-0.0608***	0.00529***	0.263***	0.0348***	-0.000934***	-0.0102***	-0.0460***	0.00237
	(0.00951)	(0.00231)	(0.00411)	(0.00398)	(0.00146)	(0.0205)	(0.00425)	(0.000256)	(0.00262)	(0.00565)	(0.00152)
tob_dum	8.003***	1.465***	1.040***	2.729***	-0.174***	-13.63***	-1.135***	0.0433***	0.475***	1.816***	-0.0463
	(0.445)	(0.108)	(0.189)	(0.186)	(0.0669)	(0.962)	(0.186)	(0.0120)	(0.124)	(0.253)	(0.0677)
tob_lnM	-1.803***	-0.308***	-0.221***	-0.611***	0.0484***	2.896***	0.292***	-0.00873***	-0.0888***	-0.410***	0.0226
	(0.103)	(0.0256)	(0.0450)	(0.0432)	(0.0157)	(0.225)	(0.0443)	(0.00281)	(0.0286)	(0.0602)	(0.0159)
tob_lnM2	0.100***	0.0154***	0.0103***	0.0340***	-0.00350***	-0.146***	-0.0188***	0.000437***	0.00416**	0.0240***	-0.00219**
	(0.00589)	(0.00153)	(0.00273)	(0.00250)	(0.000902)	(0.0131)	(0.00260)	(0.000165)	(0.00163)	(0.00369)	(0.000935)
ln_hhs	-0.0339***	0.00227***	-0.00745***	-0.00332***	0.00612***	0.0512***	-0.0132***	0.000109	-0.00952***	-0.0113***	0.00381***
	(0.00263)	(0.000520)	(0.00121)	(0.00122)	(0.000567)	(0.00512)	(0.00113)	(6.69e-05)	(0.000785)	(0.00178)	(0.000574)
no_kid	9.93e-05	-0.000477***	-0.00132***	-0.00129***	0.000842***	-0.00779***	-0.00244***	4.72e-05**	-0.00154***	-0.000191	-0.000218
	(0.00102)	(0.000168)	(0.000403)	(0.000484)	(0.000200)	(0.00196)	(0.000428)	(2.37e-05)	(0.000281)	(0.000596)	(0.000218)
ageadult_mean	-0.000337*	6.53e-05**	-0.000152*	-0.000630***	0.000109***	-0.00149***	0.000424***	-1.33e-05***	0.000165***	-0.000645***	0.000171***
	(0.00192)	(3.33e-05)	(8.48e-05)	(8.02e-05)	(3.20e-05)	(0.000369)	(8.19e-05)	(4.48e-06)	(5.38e-05)	(0.000125)	(3.44e-05)
male	-0.0116***	-0.000305	-0.00470***	-0.000471	-0.00155***	-0.00242	-0.00628***	-0.000367***	-0.00191***	-0.00124	-8.29e-05
	(0.00228)	(0.000343)	(0.000800)	(0.00110)	(0.000427)	(0.00400)	(0.000870)	(5.72e-05)	(0.000581)	(0.00136)	(0.000421)
Coloured	-0.0400***	-0.00485***	-0.0107***	-0.00961***	-0.00669***	-0.0489***	-0.00227	-0.000658***	-0.00479**	-0.00739	-0.000280
	(0.00698)	(0.00136)	(0.00351)	(0.00319)	(0.00130)	(0.0145)	(0.00308)	(0.000200)	(0.00205)	(0.00483)	(0.00136)
Indian	-0.0239***	-0.00191	-0.0106***	-0.0217***	-0.00444***	-0.113***	-0.0303***	-0.00152***	-0.00167	-0.0362***	-0.00425***

	(0.00530)	(0.00128)	(0.00264)	(0.00243)	(0.00131)	(0.0108)	(0.00344)	(0.000176)	(0.00183)	(0.00357)	(0.00138)
White	0.00932	-0.0147***	-0.0359***	-0.0290***	-0.00624*	0.215***	-0.0266***	-0.00197***	-0.0264***	-0.0235*	-0.000529
	(0.0196)	(0.00358)	(0.00915)	(0.00831)	(0.00366)	(0.0391)	(0.00835)	(0.000521)	(0.00583)	(0.0130)	(0.00367)
Urban	-0.0188***	7.92e-05	-0.00579***	-0.00817***	-0.00117*	-0.00104	-0.000362	-0.00105***	-0.00665***	-0.00135***	-0.00284***
	(0.00334)	(0.000608)	(0.00139)	(0.00146)	(0.000617)	(0.00635)	(0.00145)	(8.16e-05)	(0.000996)	(0.00206)	(0.000641)
EC	-0.0248***	-0.00254***	-0.00383*	-0.00959***	-0.000821	-0.0314***	-0.00212	-0.000153	-0.00419***	-0.00157	-0.00435***
	(0.00391)	(0.000858)	(0.00201)	(0.00180)	(0.000805)	(0.00791)	(0.00205)	(0.000109)	(0.00124)	(0.00276)	(0.000928)
NC	-0.0203***	-1.69e-05	-0.00186	-0.00921***	-0.000405	-0.0292**	-0.00139	7.26e-05	-0.000239	-0.00369	-0.00315**
	(0.00655)	(0.00141)	(0.00322)	(0.00279)	(0.00103)	(0.0127)	(0.00316)	(0.000170)	(0.00179)	(0.00435)	(0.00125)
FS	-0.0759***	-0.00457***	-0.00390	-0.00178	-0.000277	-0.00389	-0.00538**	-5.17e-05	-0.00407**	-0.00532	-0.000953
	(0.00515)	(0.00108)	(0.00258)	(0.00239)	(0.000972)	(0.0105)	(0.00256)	(0.000141)	(0.00159)	(0.00356)	(0.00110)
KZN	-0.0148**	-0.00395***	-0.00726***	-0.00322	-0.000305	-0.00380	-0.000620	-0.000183	-0.00984***	-0.00946**	-0.00363***
	(0.00603)	(0.00112)	(0.00262)	(0.00263)	(0.00121)	(0.0120)	(0.00300)	(0.000174)	(0.00187)	(0.00377)	(0.00124)
NW	-0.0367***	-0.00369***	-0.00196	-0.0117***	-0.00118	-0.0347***	-0.00355	-9.33e-05	-0.00773***	-0.00239	-0.00371***
	(0.00479)	(0.00106)	(0.00229)	(0.00225)	(0.000918)	(0.00982)	(0.00251)	(0.000142)	(0.00149)	(0.00300)	(0.00102)
GA	-0.00802	-0.00339**	-0.00529	-0.0171***	-0.000653	-0.00587	-0.00241	-0.000589***	-0.0161***	-0.0152***	-0.00305**
	(0.00685)	(0.00143)	(0.00323)	(0.00299)	(0.00134)	(0.0139)	(0.00336)	(0.000204)	(0.00215)	(0.00432)	(0.00141)
MP	-0.0149**	-0.00585***	-0.00893***	-0.00640*	-0.00315**	-0.0263*	-0.00408	-0.000158	-0.0130***	-0.0309***	-0.00283*
	(0.00755)	(0.00146)	(0.00336)	(0.00332)	(0.00143)	(0.0153)	(0.00376)	(0.000204)	(0.00236)	(0.00467)	(0.00150)
LMP	-0.00880	-0.00300**	-0.00998***	-0.00355	-0.00122	-0.00789	-0.00350	-0.000217	-0.00749***	-0.0141***	-0.00538***
	(0.00626)	(0.00122)	(0.00271)	(0.00276)	(0.00122)	(0.0122)	(0.00303)	(0.000176)	(0.00185)	(0.00397)	(0.00123)
Constant	12.51***	2.142***	1.599***	4.830***	0.258**	24.60***	2.022***	0.0900***	1.093***	3.694***	-0.0247
	(0.715)	(0.165)	(0.290)	(0.295)	(0.106)	(1.516)	(0.300)	(0.0187)	(0.195)	(0.397)	(0.109)
Observations	29,512	29,512	29,512	29,512	29,512	29,512	29,512	29,512	29,512	29,512	29,512

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Reference groups are as follows: female, black African, rural and WC (Western Cape). Province abbreviations are as follows: EC - Eastern Cape, NC - Northern Cape, FS - Free State, KZN - Kwazulu-Natal, NW - North Western, GA - Gauteng, MP - Mpumalanga, LMP - Limpopo.

Table A 14: Three Stage-Least square (3SLS) Regression 1995 IES Urban Sample (Equation 7 specification)

VARIABLES	(1) bsfoodexp	(2) bstobexp	(3) bsalcexp	(4) bsclthsexp	(5) bseduexp	(6) bshseexp	(7) bstransexp	(8) bscommexp	(9) bshealthexp	(10) bshhgdsexp	(11) bsrecexp
pq	-0.000198	0.000433***	0.000855*	-0.000136	0.000156	-0.00245	-4.07e-06	-1.73e-05	-0.000558	-0.000366	6.17e-05
	(0.000630)	(0.000161)	(0.000459)	(0.000271)	(0.000151)	(0.00153)	(0.000289)	(2.28e-05)	(0.000355)	(0.000391)	(0.000138)
lnM	2.642***	0.538***	0.471***	1.108***	-0.0962*	-5.201***	0.660***	0.0387***	0.357***	0.983***	-0.0116
	(0.273)	(0.0717)	(0.166)	(0.116)	(0.0500)	(0.664)	(0.120)	(0.00921)	(0.129)	(0.137)	(0.0435)
lnM2	-0.150***	-0.0280***	-0.0235**	-0.0612***	0.00664**	0.262***	-0.0422***	-0.00206***	-0.0189***	-0.0547***	0.00237
	(0.0152)	(0.00409)	(0.00930)	(0.00646)	(0.00279)	(0.0373)	(0.00687)	(0.000519)	(0.00731)	(0.00768)	(0.00252)
tob_dum	7.469***	1.723***	1.443***	2.745***	-0.240*	-14.27***	1.490***	0.0977***	0.796**	2.227***	-0.0453
	(0.760)	(0.207)	(0.487)	(0.328)	(0.143)	(1.887)	(0.330)	(0.0266)	(0.372)	(0.394)	(0.119)
tob_lnM	-1.657***	-0.362***	-0.308***	-0.611***	0.0624*	3.006***	-0.373***	-0.0208***	-0.162*	-0.503***	0.0227
	(0.173)	(0.0479)	(0.112)	(0.0747)	(0.0328)	(0.431)	(0.0765)	(0.00607)	(0.0855)	(0.0903)	(0.0276)
tob_lnM2	0.0912***	0.0180***	0.0142**	0.0339***	-0.00436**	-0.149***	0.0229***	-0.00112***	0.00928*	0.0289***	-0.00215
	(0.00946)	(0.00272)	(0.00616)	(0.00411)	(0.00177)	(0.0239)	(0.00427)	(0.000339)	(0.00478)	(0.00499)	(0.00159)
ln_hhs	-0.0388***	-0.00245**	-0.00721***	-0.00422**	-0.00723***	-0.0403***	0.0161***	-2.10e-05	-0.00877***	-0.00905***	-0.00447***
	(0.00385)	(0.00102)	(0.00268)	(0.00165)	(0.000937)	(0.00907)	(0.00174)	(0.000134)	(0.00214)	(0.00222)	(0.000839)
no_kid	-0.000430	0.000612	0.00173*	0.00163**	-0.00119***	-0.00952**	0.00209**	-7.70e-05	0.00103	0.000432	-0.000107
	(0.00168)	(0.000384)	(0.00105)	(0.000783)	(0.000377)	(0.00373)	(0.000824)	(5.80e-05)	(0.000832)	(0.000947)	(0.000356)
ageadult_mean	-0.000317	7.55e-05	0.000268	-0.000545***	-0.000129*	0.00105	-0.000592***	2.36e-05**	-0.000252	-0.000504***	0.000194***
	(0.000303)	(8.00e-05)	(0.000228)	(0.000131)	(7.04e-05)	(0.000759)	(0.000143)	(1.07e-05)	(0.000175)	(0.000194)	(6.56e-05)
male	-0.00395	0.000475	0.00286	-0.00182	-0.00256***	0.0151*	0.00729***	-0.000516***	0.00461**	-0.00356	-0.000550
	(0.00363)	(0.000872)	(0.00241)	(0.00175)	(0.000915)	(0.00848)	(0.00156)	(0.000126)	(0.00193)	(0.00232)	(0.000782)
Coloured	-0.0305**	-0.00635*	-0.0197**	-0.0157***	-0.00854***	-0.100***	0.00376	-0.000766	0.0112	-0.0176*	-0.000913
	(0.0130)	(0.00344)	(0.00978)	(0.00564)	(0.00302)	(0.0326)	(0.00605)	(0.000480)	(0.00752)	(0.00785)	(0.00274)
Indian	-0.0283***	-0.00175	-0.0145***	-0.0266***	-0.00480***	-0.124***	0.0270***	-0.00144***	0.00211	-0.0417***	0.00246
	(0.00639)	(0.00193)	(0.00477)	(0.00321)	(0.00183)	(0.0163)	(0.00408)	(0.000259)	(0.00377)	(0.00395)	(0.00171)
White	-0.00884	-0.0174**	-0.0521***	-0.0323**	-0.00852	-0.266***	0.0297*	-0.00208	-0.0465***	-0.0342	-8.59e-05
	(0.0343)	(0.00882)	(0.0257)	(0.0148)	(0.00825)	(0.0842)	(0.0157)	(0.00128)	(0.0196)	(0.0217)	(0.00751)
Urban	-	-	-	-	-	-	-	-	-	-	-

EC	-0.0215***	-0.00236	0.00138	-0.0109***	0.00144	0.0146	0.000191	-0.000332*	0.00223	0.00487	-0.00323***
	(0.00543)	(0.00144)	(0.00383)	(0.00245)	(0.00127)	(0.0128)	(0.00257)	(0.000190)	(0.00313)	(0.00326)	(0.00124)
NC	-0.0201**	-0.000828	0.00796	-0.0153***	-8.66e-06	-0.0247	-0.00238	-0.000336	-0.00528	-0.00210	-0.00213
	(0.0101)	(0.00242)	(0.00690)	(0.00440)	(0.00216)	(0.0238)	(0.00457)	(0.000366)	(0.00562)	(0.00565)	(0.00214)
FS	-0.0344***	-0.00225	0.00479	0.00649***	-0.00126	-0.0214	0.00312	9.26e-05	0.00215	0.00616**	-0.000318
	(0.00553)	(0.00151)	(0.00384)	(0.00238)	(0.00122)	(0.0136)	(0.00274)	(0.000199)	(0.00293)	(0.00314)	(0.00121)
KZN	-0.00939	-0.00506*	-0.00742	-0.00746*	-0.00157	0.0127	0.00159	0.000214	0.0158***	0.0132**	-0.00194
	(0.00937)	(0.00271)	(0.00755)	(0.00413)	(0.00234)	(0.0252)	(0.00489)	(0.000360)	(0.00564)	(0.00606)	(0.00202)
NW	-0.0187*	-0.00569*	-0.00498	-0.00150	-1.52e-06	0.0130	0.00686	0.000391	0.0175***	-0.00164	-0.00275	
	(0.0108)	(0.00313)	(0.00897)	(0.00468)	(0.00258)	(0.0294)	(0.00539)	(0.000390)	(0.00667)	(0.00683)	(0.00239)
GA	-0.0125	-0.00432	-0.00722	0.0169***	-0.000636	-0.000121	0.00182	0.000671	0.0202***	0.0130**	-0.00221
	(0.0107)	(0.00321)	(0.00890)	(0.00482)	(0.00266)	(0.0285)	(0.00537)	(0.000413)	(0.00667)	(0.00725)	(0.00237)
MP	-0.0199	-0.00680*	-0.0142	-0.00508	-0.00341	0.0366	0.00815	6.76e-05	0.0189**	0.0235**	-0.00327
	(0.0129)	(0.00357)	(0.00999)	(0.00552)	(0.00304)	(0.0321)	(0.00620)	(0.000459)	(0.00762)	(0.00837)	(0.00271)
LMP	-0.00269	-0.00476	-0.0105	-0.00320	-0.00250	0.0115	0.0105*	0.000702	0.0154**	0.0109	-0.00386
	(0.0115)	(0.00295)	(0.00807)	(0.00518)	(0.00283)	(0.0276)	(0.00579)	(0.000453)	(0.00634)	(0.00693)	(0.00252)
Constant	-11.28***	-2.478***	-2.128***	-4.884***	0.375**	24.91***	2.590***	-0.181***	-1.736***	-4.345***	-0.0308
	(1.189)	(0.306)	(0.711)	(0.503)	(0.214)	(2.879)	(0.511)	(0.0400)	(0.556)	(0.591)	(0.186)
Observations	16,875	16,875	16,875	16,875	16,875	16,875	16,875	16,875	16,875	16,875	16,875

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Reference groups are as follows: female, black African, rural and WC (Western Cape). Province abbreviations are as follows: EC - Eastern Cape, NC - Northern Cape, FS - Free State, KZN - Kwazulu-Natal, NW - North Western, GA - Gauteng, MP - Mpumalanga, LMP - Limpopo.

Table A 15: Three Stage-Least square (3SLS) Regression 1995 IES Rural Sample (Equation 7 specification)

VARIABLES	(1) bsfoodexp	(2) bstobexp	(3) bsalcexp	(4) bsclthsexp	(5) bseduexp	(6) bshseexp	(7) bstransexp	(8) bscommexp	(9) bshealthexp	(10) bshhgdsexp	(11) bsrecexp
pq	-0.00168	0.000325**	0.000489	-0.000457	5.80e-05	0.000520	0.000319	-5.91e-06	0.000126	-0.00110	0.000152
	(0.00197)	(0.000142)	(0.000436)	(0.000617)	(0.000155)	(0.00190)	(0.000404)	(1.58e-05)	(0.000227)	(0.00103)	(0.000199)
lnM	3.447***	0.377***	0.200*	1.134***	-0.0686*	-4.844***	-0.270**	0.00506	0.171***	0.643***	0.00628
	(0.551)	(0.0657)	(0.118)	(0.161)	(0.0381)	(0.591)	(0.122)	(0.00518)	(0.0666)	(0.296)	(0.0522)
lnM2	-0.196***	-0.0199***	-0.0106	-0.0620***	0.00516**	0.243***	0.0195***	-0.000176	-0.00722*	-0.0329*	0.00109
	(0.0327)	(0.00388)	(0.00704)	(0.00956)	(0.00231)	(0.0348)	(0.00735)	(0.000309)	(0.00395)	(0.0176)	(0.00314)
tob_dum	9.261***	1.201***	0.526*	2.879***	-0.169*	-12.54***	-0.570*	0.00804	0.397**	1.429*	-0.0155
	(1.374)	(0.172)	(0.291)	(0.404)	(0.0942)	(1.530)	(0.297)	(0.0132)	(0.168)	(0.731)	(0.129)
tob_lnM	-2.081***	-0.254***	-0.115*	-0.639***	0.0487**	2.631***	0.157**	-0.000808	-0.0709*	-0.298*	0.0137
	(0.322)	(0.0405)	(0.0686)	(0.0944)	(0.0226)	(0.360)	(0.0716)	(0.00314)	(0.0398)	(0.172)	(0.0305)
tob_lnM2	0.117***	0.0128***	0.00544	0.0353***	-0.00353**	-0.133***	-0.0112**	-1.60e-05	0.00244	0.0166	-0.00168
	(0.0204)	(0.00245)	(0.00435)	(0.00597)	(0.00143)	(0.0222)	(0.00453)	(0.000196)	(0.00245)	(0.0109)	(0.00197)
ln_hhs	0.0138	-0.00261***	-0.00644**	0.00239	0.00492***	0.0719***	-0.00902***	-0.000238**	-0.00980***	-0.0149**	-0.00238*
	(0.0128)	(0.000987)	(0.00288)	(0.00422)	(0.00118)	(0.0120)	(0.00276)	(0.000101)	(0.00162)	(0.00686)	(0.00135)
no_kid	-0.000977	0.000435*	0.00124*	0.000697	-0.000461	-0.00582*	0.00237***	-9.60e-06	0.00186***	-0.00103	0.000480
	(0.00318)	(0.000229)	(0.000676)	(0.00100)	(0.000291)	(0.00348)	(0.000679)	(2.72e-05)	(0.000392)	(0.00159)	(0.000336)
ageadult_mean	-0.000638	2.55e-05	7.99e-05	-0.000645***	-5.86e-05	0.00183***	-0.000171	-3.30e-06	-9.46e-05	-0.000703**	-0.000117*
	(0.000666)	(4.61e-05)	(0.000144)	(0.000213)	(5.07e-05)	(0.000658)	(0.000144)	(5.63e-06)	(7.76e-05)	(0.000356)	(6.65e-05)
male	-0.0159***	0.000436	0.00506***	0.00253	-0.000844	-0.0114	0.00475***	-9.07e-05*	0.000115	0.000289	0.000774
	(0.00578)	(0.000430)	(0.00124)	(0.00201)	(0.000625)	(0.00707)	(0.00136)	(5.36e-05)	(0.000814)	(0.00316)	(0.000670)
Coloured	0.0394**	-0.00335**	0.00352	-0.00105	-0.00248**	-0.0215	0.00780*	0.000180	-0.000821	-0.00182	0.00161
	(0.0166)	(0.00167)	(0.00362)	(0.00500)	(0.00112)	(0.0153)	(0.00451)	(0.000152)	(0.00201)	(0.00822)	(0.00207)
Indian	0.0492	0.00259	-0.00487	-0.00420	-0.00326	0.00446	0.0466***	0.00231***	0.0129**	-0.00721	0.0131***
	(0.0314)	(0.00223)	(0.00650)	(0.0109)	(0.00306)	(0.0402)	(0.0148)	(0.000526)	(0.00556)	(0.0168)	(0.00494)
White	0.0583	-0.0110*	-0.0204	-0.0137	-0.00461	0.0886	0.0491***	0.00273***	0.0125	-0.0131	0.00503
	(0.0651)	(0.00569)	(0.0143)	(0.0205)	(0.00496)	(0.0607)	(0.0141)	(0.000528)	(0.00816)	(0.0353)	(0.00681)
Urban	-	-	-	-	-	-	-	-	-	-	-

EC	-0.0879*	-0.00360	-0.0101	-0.0111	0.0027	0.0968**	0.00887	-8.10e-05	0.00242	-0.0174	-0.0066
	(0.0475)	(0.00297)	(0.0109)	(0.0144)	(0.00342)	(0.0422)	(0.0107)	(0.000378)	(0.00515)	(0.0253)	(0.00466)
NC	-0.0628	-0.000555	-0.00929	0.000267	0.00108	0.00104	0.00970	4.73e-05	0.00125	-0.0151	-0.00662*
	(0.0392)	(0.00306)	(0.00882)	(0.0122)	(0.00281)	(0.0360)	(0.00930)	(0.000310)	(0.00391)	(0.0210)	(0.00367)
FS	-0.215***	-0.00850*	0.00216	0.000960	0.00228	0.0877	0.0172	-2.26e-06	0.00945	-0.0164	-0.00171
	(0.0694)	(0.00457)	(0.0158)	(0.0213)	(0.00528)	(0.0649)	(0.0148)	(0.000554)	(0.00759)	(0.0368)	(0.00665)
KZN	-0.0198	-0.00378*	-0.0111*	-0.000503	0.00317	0.0156	0.00567	0.000103	0.00250	-0.00657	-0.00718**
	(0.0286)	(0.00203)	(0.00664)	(0.00819)	(0.00207)	(0.0251)	(0.00748)	(0.000246)	(0.00314)	(0.0154)	(0.00311)
NW	-0.123**	-0.00335	0.00507	-0.0282	0.00356	0.133**	0.00970	-0.000132	0.00268	-0.0255	-0.00498
	(0.0584)	(0.00358)	(0.0134)	(0.0179)	(0.00443)	(0.0539)	(0.0127)	(0.000457)	(0.00619)	(0.0310)	(0.00562)
GA	-0.00529	-0.00170	-0.0127	0.0130	0.00513*	0.0123	0.0120	0.000520	0.0212***	0.0191	-0.00848*
	(0.0380)	(0.00380)	(0.00867)	(0.0116)	(0.00300)	(0.0379)	(0.00981)	(0.000357)	(0.00555)	(0.0206)	(0.00473)
MP	-0.0502*	-0.00506**	-0.00774	0.0158*	-0.000784	0.0496*	0.00818	0.000293	0.00754**	0.0234	-0.00403
	(0.0293)	(0.00212)	(0.00670)	(0.00869)	(0.00200)	(0.0254)	(0.00741)	(0.000256)	(0.00335)	(0.0158)	(0.00324)
LMP	-0.0468	-0.00305	-0.0116	0.00316	0.00199	0.0557*	0.00899	-2.08e-05	0.000806	-0.00290	-0.00747**
	(0.0359)	(0.00227)	(0.00825)	(0.0107)	(0.00259)	(0.0323)	(0.00865)	(0.000293)	(0.00384)	(0.0193)	(0.00367)
Constant	-14.71***	-1.726***	-0.855*	-5.039***	0.225	-23.37***	0.952*	-0.0287	-0.870***	-3.042**	-0.0858
	(2.382)	(0.280)	(0.508)	(0.700)	(0.161)	(2.545)	(0.512)	(0.0220)	(0.285)	(1.274)	(0.224)
Observations	12,637	12,637	12,637	12,637	12,637	12,637	12,637	12,637	12,637	12,637	12,637

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Reference groups are as follows: female, black African, rural and WC (Western Cape). Province abbreviations are as follows: EC - Eastern Cape, NC - Northern Cape, FS - Free State, KZN - Kwazulu-Natal, NW - North Western, GA - Gauteng, MP - Mpumalanga, LMP - Limpopo.

3. CHAPTER 3: TOBACCO TAX INCREASES AND THE EVOLUTION OF TOBACCO-INDUCED CROWDING-OUT

3.1 Introduction

The 1990s marked a transformative period for South Africa, characterised by the end of apartheid and the establishment of a democratic government, which created new opportunities for public health and economic reform. The post-apartheid government was motivated to address social and health inequalities exacerbated by years of systemic racial discrimination, and tobacco control became an integral part of these efforts. This decade saw South Africa's tobacco-control policies gaining momentum, spurred by rising awareness of the health and economic consequences of tobacco use. Against a backdrop of high smoking prevalence, the government implemented a series of legislative and fiscal measures aimed at reducing tobacco consumption, ultimately setting a precedent for other African nations.

Among South Africa's various tobacco-control strategies, tobacco taxation emerged as a cornerstone, especially in the period between 1994 and 2005 (Linegar & van Walbeek, 2018). Recognising the potential of excise taxes to both deter consumption and generate revenue, the government initiated substantial tax increases on tobacco products. These increases were not only intended to elevate the price of tobacco and reduce demand but also to mitigate the economic costs associated with tobacco-related diseases. The period witnessed successive tax hikes. This significantly raised the price of cigarettes, contributing to a decline in smoking rates across the country. The commitment to high excise taxes demonstrated the government's resolve in leveraging fiscal policies for public health outcomes, a strategy that laid the groundwork for sustained tobacco control (Linegar & van Walbeek, 2018).

Despite the evident successes of these tobacco tax policies in curbing smoking rates, an important area of inquiry remains largely unexplored: the economic impact of increased tobacco expenditures on household budgets. A key focus of this chapter is the phenomenon of tobacco-induced crowding-out—specifically, the extent to which household spending on tobacco products displaces essential expenditures on goods and services such as food, education, and healthcare. This effect is particularly pronounced among low-income households, where budget constraints are more stringent, and the diversion of funds toward tobacco can severely limit investment in crucial areas of family welfare.

While previous research has documented crowding-out effects, most studies have relied on static analyses using cross-sectional data, providing only a snapshot of household spending patterns at a single point in time. Such an approach risks overlooking the dynamic nature of these effects as households adjust their spending in response to ongoing price changes and tax policy adjustments. For example, while earlier studies successfully employed instrumental variable techniques to address potential endogeneity in tobacco expenditures and income, they did not account for the temporal shifts in spending behaviour that have occurred in South Africa following significant tax hikes and a notable decline in smoking prevalence between 1995 and 2010.

The current chapter addresses this limitation by applying a modified DiD method to repeated cross-sectional data, enabling us to analyse aggregate changes in household spending over time. While cross-sectional studies typically capture data from different samples at various points, repeated cross-sectional data allows us to observe aggregate shifts in household spending patterns across multiple years, even though the same households are not tracked over time. This methodological choice is particularly important for understanding the evolving crowding-out effects resulting from tobacco taxation. By integrating a propensity score weighting scheme with the DiD approach`(Abadie, 2005), the study offers a broader view of the crowding-out effect in the context of South Africa's tobacco tax policy.

The contributions of this chapter are thus threefold. First, it adds to the literature on behavioural responses to tobacco taxation by investigating how budget allocations shift in response to tax-induced price increases (Adda & Cornaglia, 2010, Xu et al., 2013). Second, it offers insights into the broader impact of fiscal policies on household expenditure, providing a valuable extension of previous work that has focused more narrowly on single time points (Caro et al., 2020, Sacks, Kwon & Backholer, 2021, Sassi et al., 2018). Finally, this study makes a critical methodological contribution by employing a difference-in-differences (DiD) approach adapted to repeated cross-sectional data, combining DiD with a weighting scheme based on propensity scores (Abadie, 2005). This approach enhances robustness against selection bias and extends the DiD methodology used in prior studies as for instance in Nyagwachi, Chelwa & van Walbeek (2020).

Beyond its scholarly implications, this chapter offers valuable insights for assessing the ongoing effectiveness of tobacco tax policies over time. It highlights potential unintended consequences, such as more pronounced expenditure shifts in response to higher cigarette

prices, which is critical for policymakers to consider. By examining these changes over time, policymakers can better respond to evolving contexts and ensure that policy objectives are being met effectively and equitably.

3.2 Relevant literature

The DiD method is widely considered a workhorse for estimating the treatment effect of various policies in applied social science research. Its popularity has been increasing and the intricacies of the method are evolving rapidly. A review of six economic journals by Bertrand, Duflo & Mullainathan (2004) found 92 papers using the method over the period between 1990 and 2000. De Chaisemartin & D'Haultfœuille (2020) reviewed 337 papers published in the American Economic Review between 2010 and 2012 that also used this method. Most recently, another review found that over 750 articles from top finance and accounting journals used the DiD design between 2000 and 2019 (Baker, Larcker & Wang, 2021).

From these reviews, it is clear that the DiD method gained traction post-2000, prompting a focused exploration of its application to tobacco-tax increases from 2010 onwards. This period is chosen to reflect the most recent advancements in the method, as the use and understanding of DiD have rapidly progressed. The literature reveals a predominance of studies conducted in Europe (Ben Lakhdar, Vaillant & Wolff, 2016, Lopez-Nicolas, Badillo-Amador & Cobacho-Tornel, 2013) and North America (Abouk et al., 2021, Callison & Kaestner, 2014, Hawkins & Baum, 2014, Hawkins, Bach & Baum, 2018, Hawkins, Kull & Baum, 2019, Manivong, Harper & Strumpf, 2017, Pesko, Courtemanche & Maclean, 2020, Rozema & Ziebarth, 2017, Saffer et al., 2020, Yu, H., Engberg & Scharf, 2019). Two studies were conducted in Asia (Jin & Cho, 2021, Zhang & Zheng, 2020), one study was done in South America (Gallego, J. M. et al., 2020) and two in Africa (Chelwa, van Walbeek & Blecher, 2016, Nyagwachi, Chelwa & van Walbeek, 2020).

Tobacco-tax increases are employed by policy makers to induce price increases on tobacco products, and therefore reduce their consumption. Most studies in this review aim to determine the effect of tobacco-tax increases on tobacco consumption, for instance by estimating a smoking participation elasticity (Callison & Kaestner, 2014, Hawkins & Baum, 2014, Lopez-Nicolas, Badillo-Amador & Cobacho-Tornel, 2013, Manivong, Harper & Strumpf, 2017, Yu, H., Engberg & Scharf, 2019). However, other popular extensions include studies exploring substitution effects to other products (Abouk et al., 2021, Hawkins, Bach & Baum, 2018, Pesko, Courtemanche & Maclean, 2020, Saffer et al., 2020), consumption of illicit cigarettes

(Ben Lakhdar, Vaillant & Wolff, 2016, Gallego, J. M. et al., 2020), and expenditures at household level (Hawkins, Kull & Baum, 2019, Jin & Cho, 2021, Nyagwachi, Chelwa & van Walbeek, 2020). Less common studies have investigated the effects of tax increases on self-reported well-being (Leicester & Levell, 2016) and the probability of enrolment in public assistance programmes (Rozema & Ziebarth, 2017).

Effects on cigarette consumption: Callison & Kaestner (2014) use the synthetic control method of Abadie, Diamond & Hainmueller (2010) to estimate the association between a cigarette-tax increase and cigarette consumption in selected states of the United States (US). They estimate a participation elasticity of between -0.02 and -0.05. However, their estimates were mostly statistically insignificant. This is a similar result to the Lopez-Nicolas, Badillo-Amador & Cobacho-Tornel (2013) study in Spain. They use a DiD estimator to compare differences in prevalence between mainland Spain and the Balearic Islands before and after the introduction of a minimum tax on cigarettes in 2006. While prices increased three times faster in the treatment territory, smoking prevalence among males remained unaffected in both the short and long term. However, among women a reduction in prevalence of up to 4.3% was detected in the long term.

Hawkins & Baum (2014) focus their study on the impact of tobacco-tax increases on differences in maternal smoking during pregnancy. Their study found that for every US$1.00 increase in cigarette tax, mothers with low education smoked between 14 and 22 fewer cigarettes monthly. In contrast, Manivong, Harper & Strumpf (2017) looked at the impact of cigarette-tax changes on the smoking outcomes of youths (aged 15-18 years). They found that a CAD$1.00 increase in cigarette tax was associated with a 0.2 percentage point decline in smoking prevalence, and 0.3 decrease in the mean number of cigarettes smoked in the past week. Despite this, the authors note that, owing to pre-existing declines in smoking even among provinces with stable cigarette taxes, it is possible that other common factors played a greater role in the overall declines observed in their study.

Yu, H., Engberg & Scharf (2019) take a different approach from previous studies. They augment their DiD analysis by incorporating generalised propensity scores to estimate the relative impact of three state-level tobacco-control policies on overall adult smoking rates and among adult subgroups in the US. They found that, compared to two other policies (tobacco-control spending and smoke-free air laws), state cigarette taxation was the only policy that significantly impacted smoking. An increase in taxes of $0.68 in constant 2014 dollars led to

0.25 percentage point drop in the adult smoking rate in the general population. Population subgroups significantly affected by cigarette taxes included females, non-Hispanic whites, adults aged 51 years and older, and adults with more than a high school education.

Effects on consumption of illicit cigarettes: Two studies considered the effect of tax increases on illicit cigarette consumption (Ben Lakhdar, Vaillant & Wolff, 2016, Gallego, J. M. et al., 2020). Ben Lakhdar, Vaillant & Wolff (2016) exploit a natural experiment resulting from tax-induced cigarette price increases in France between 2001 and 2004. They compared changes in cigarette sales between border and non-border *départments* to identify the causal effect of price increases. The study estimated that sales were about 20% higher in border *départments* than in non-border *départments* between 2002 and 2007. Although cross-border purchases are not illegal, this suggests higher tax avoidance among smokers, who found cross-border purchases more financially attractive.

In Colombia, Gallego, J. M. et al. (2020) estimated the short-term impact of a doubling of the tobacco excise taxes on the consumption of illicit cigarettes. The authors compared variations in the level of penetration of illicit cigarettes between consumers of low-priced versus higher priced cigarettes. Among smokers of low-priced cigarettes, the likelihood of consuming illicit cigarettes increased between 5.1 and 5.4 percentage points compared to smokers of higher-priced cigarettes. However, this was only statistically significant at the 10% level. In general, they concluded that, since the change in illicit cigarette penetration was much lower than estimates found in developed countries, increases in the illicit market did not represent a significant hinderance to tobacco-tax increases in Colombia.

Substitution effects: Product substitution is a major concern for tobacco control, especially where tobacco taxes are not uniformly applied or increased. Hawkins, Bach & Baum (2018) explore the relationship between changes in cigarette tax and smoke-free legislation in the US, on the one hand, and changes in the consumption of smokeless tobacco and cigar use on the other. The authors found that both chewing tobacco taxes and cigar taxes had no effect on smokeless tobacco use or cigar use. However, a 10% increase in cigarette taxes was associated with a 1.0 percentage point increase in smokeless tobacco use, and a 1.5 percentage point increase in cigar use among males.

Interest is also growing regarding the relationship between emerging tobacco and nicotine products and traditional tobacco products. For instance, some researchers have turned their attention to the impact cigarette-tax increases may have on e-cigarette use (Pesko,

Courtemanche & Maclean, 2020), while the effect of e-cigarette taxes on youth tobacco smoking is also of interest to others (Abouk et al., 2021). In the first study, Pesko, Courtemanche & Maclean (2020) implemented a two-way fixed-effects model and event study. They found that, in the US, a $1.00 increase in the traditional cigarette tax led to reductions in both current and everyday use of traditional cigarettes by 0.3 and 0.4 percentage points. However, they found that this level of tax increase also led to an increase in both current and everyday e-cigarette use by 0.3 and 0.1 percentage points, suggesting the two products are economic substitutes. Abouk et al. (2021) explore this substitution issue further, focusing on youths in the US. They also use a two-way fixed-effects model, accounting for spatial and temporal heterogeneity, to determine the impact of e-cigarette taxes on youth tobacco use. They found similar results to those of Pesko, Courtemanche & Maclean (2020); while an increase in e-cigarette taxes acted as a significant deterrent to e-cigarette consumption among the youth, with an estimated tax elasticity of between -0.06 and -0.21, it also led to significantly higher cigarette use. The tax elasticity in this latter case ranged from 0.12 for current cigarette use to 0.34 for half-pack-a-day use.

Zhang & Zheng (2020) examine the effect of increasing cigarette excise taxes on regular alcohol consumption in mainland China. They estimate a two-way fixed-effects model using propensity score matching with DiD (Heckman, J. et al., 1998). The study found that the regular drinking rate of smokers decreased significantly after the cigarette-tax increase, indicating that tobacco and alcohol are complements. Smokers were roughly twice as likely to reduce their regular alcohol intake as non-smokers.

Effects on expenditure at household-level: At the household level, tobacco expenditure affects household expenditures on other commodities, and this remains an area of on-going research. Hawkins, Kull & Baum (2019) examine the associations between cigarette taxes and smoke-free legislation, on the one hand, and household cigarette expenditure, on the other, in the US. They found that, overall, a US$1.00 cigarette tax increase reduced household cigarette expenditure by 1.5 percentage points. However, they found heterogeneity in the results. Among households at or above the poverty level, every increase in the cigarette tax was estimated to increase cigarette expenditure by US$10.73 per quarter while among households below the poverty level, it was estimated to increase cigarette expenditure by US$4.72 per quarter. The budget share on cigarettes was estimated to increase by 0.11% for households above the poverty level and by 0.14% for households below the poverty level.

A recent contribution in this area from South Korea focused on the effect of cigarette price increases on fresh food expenditure in low-income households (Jin & Cho, 2021). Using matched DiD (Abadie & Imbens, 2006), the authors exploit the sharp increase in cigarette taxes in South Korea in 2015. Their study notes that the poorest households spent relatively more on cigarettes after the tax increase by foregoing fresh food expenditures.

Nyagwachi, Chelwa & van Walbeek (2020) expand their household expenditure analysis to incorporate the effects of both tobacco and alcohol control policies on patterns of consumption and household expenditure crowding-out effects. Using a matched DiD method (Heckman, J., Ichimura & Todd, 1997, Heckman, J. et al., 1998), they analyse the evolution of the crowding-out effects of tobacco and alcohol expenditures in Kenya. Their key result was that tobacco expenditure crowded-in alcohol expenditures. Thus, although the tobacco expenditure budget share decreased 0.4 percentage points between 2005/06 and 2015/16, the budget share for alcohol expenditure increased 1.3 percentage points over the same period. They attributed the resultant crowding-out of expenditures on food items such as vegetables, bread, and cereals among tobacco-consuming households to the net increase in the combined alcohol and tobacco budget shares.

Less researched effects: A study by Leicester & Levell (2016) explores the notion that tobacco-control policies may improve the welfare of smokers who have time-inconsistent preferences by helping them to reduce their intake, or quit. Using a panel of British smokers, they implement a DiD approach to compare the effects of cigarette taxes and smoking bans on the subjective well-being of probable smokers and non-smokers.

Rozema & Ziebarth (2017) estimate the effect of cigarette-tax changes on cigarette expenditure and on food-stamp program enrolment in the US. They found that a US$1.00 increase in cigarette taxes raised the probability of food-stamp take-up among unenrolled, eligible households that reported cigarette expenditure by 3.2 percentage points. In contrast, households that reported no cigarette expenditure were unaffected by the cigarette-tax increase.

3.3 Difference-in-difference: Conceptual framework

As noted earlier, the DiD technique is used extensively in the literature to estimate causal effects at an aggregate level (Blundell & Macurdy, 1999, Blundell & Costa Dias, 2009, Imbens

& Wooldridge, 2009). The point of departure for this approach is the Rubin Causal Model[11] (Rubin, 1974), developed further by Holland (1986).

Following the potential outcome notation, let w_{ijt} be our outcome of interest defined as household *i*'s budget share for expenditure category *j* at time-period *t*. The population is observed pre-treatment (*t=0*) and post-treatment *(t=1*), resulting in:

$$w_{ijt} = w_{ijt}^{0} + (w_{ijt}^{1} - w_{ijt}^{0})\, D_{it} \quad (1)$$

where w_{ijt}^{0} is the budget share that household *i* would attain on expenditure category *j* at time-period *t* in the absence of treatment, w_{ijt}^{1} represents the budget share that household *i* would attain on expenditure category *j* at time-period *t* if exposed to treatment, and D_{it} is the treatment status of household *i* at time-period *t* taking a value of 1 if treated and 0 otherwise, since households are exposed to treatment only at *t=1*, $D_{i0} = 0$ for all *i*.

Given the above setup, the average effect of the treatment is estimated by comparing the change in the average outcome in the treated units to the change in the average outcome in the control units, defined by:

$$ATE = w_{ijt}^{1} - w_{ijt}^{0} \quad (2)$$

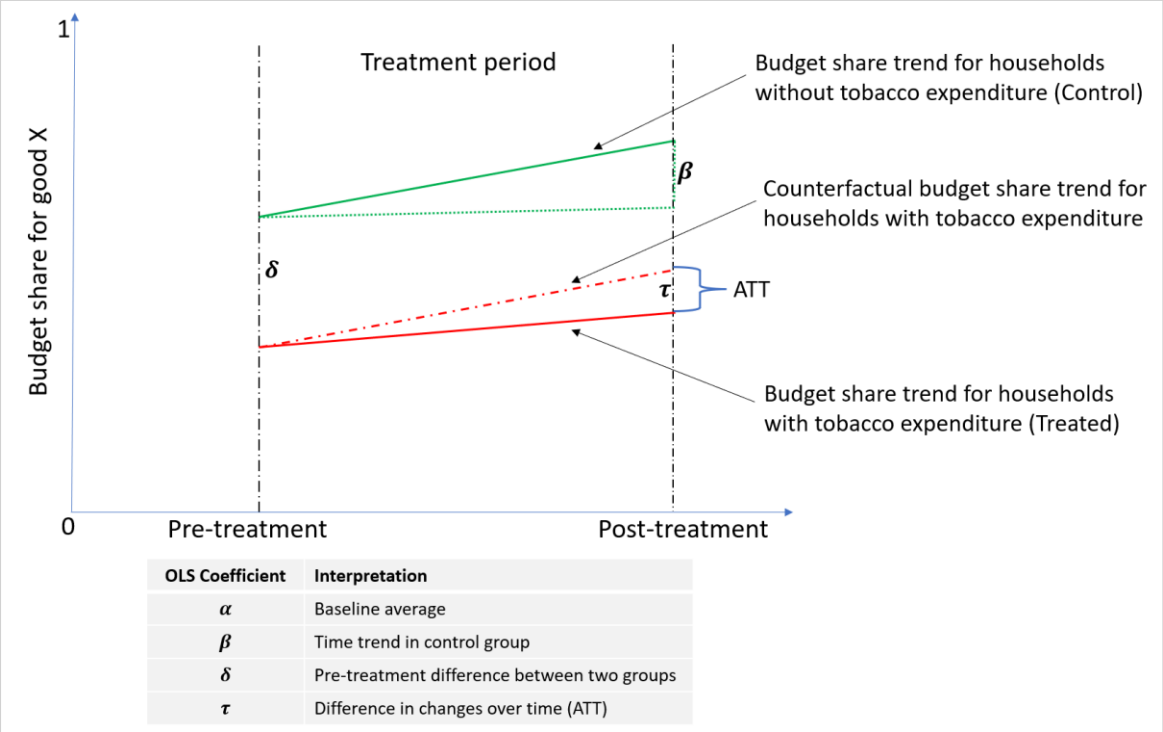
The ATE thus represents the average treatment effect that would be experienced by all households in the population under study, had they all been treated. Of course, only one outcome in Equation 2 is ever observed, which presents the classical counterfactual problem in the program evaluation literature (Angrist & Pischke, 2009). However, the average treatment effect on the treated (ATT), a related estimand, can often be more interesting and useful to estimate. The ATT refers to the average treatment effect experienced in the sub-population of households that received the treatment. Since the interest of this study is on the effect that the treatment (tobacco-tax increase) has on household budget shares among treated households, the ATT is the more useful estimand. Fortunately, it can readily be estimated via an ordinary least squares (OLS) linear regression model as follows:

$$w_{it} = \alpha + \beta t + \delta D_{it} + \tau(D_{it} \times t) + \varepsilon_{it} \quad (3)$$

[11] For a textbook treatment of the RCM, see for instance, Imbens, G.W. & Rubin, D.B. 2015. *Causal inference in statistics, social, and biomedical sciences.* Cambridge University Press.

Where t, w_{it} and D_{it} are as defined earlier, further, ε_{it} is the error term, while α, β, δ and τ are all coefficients to be estimated in the model. The coefficient τ is referred to as the ATT. Note that in the interest of making the notation less onerous, I drop the *j* subscript used earlier with the understanding that the equation would apply to any expenditure category.

Figure 3.1: Difference in difference conceptual model



OLS Coefficient	Interpretation
α	Baseline average
β	Time trend in control group
δ	Pre-treatment difference between two groups
τ	Difference in changes over time (ATT)

Conventional estimation of the ATT in the DiD approach is premised on the validity of the so-called "parallel trends" (sometimes called "common trends") assumption. In short, I assume that the average outcome for the treated and control groups would have moved in parallel if the treatment had not occurred (Angrist & Pischke, 2009, Lechner, 2011). Figure 3.1 above illustrates this point well.

When implemented well, tobacco-tax increases should lead to an increase in the price of tobacco products. Faced with an increased price of tobacco products, and assuming income levels remain unchanged, households that reported tobacco consumption before the tax increase have three primary options at their disposal: first, they could stop consuming tobacco products altogether; second, they could reduce their consumption of tobacco products which could result in more, less, or the same amount of expenditure on tobacco; third, they could maintain their level of consumption and increase their level of expenditure on tobacco products. In the DiD framework used for this chapter, the first option will have implications for identifying an appropriate counterfactual as the control group increases over time with the

inclusion of former smokers. The second and third options will have implications for the treatment group, as households either maintain existing expenditure displacements of other goods and services or adjust these displacements to find a new consumption equilibrium. The main aim of this analysis is thus to develop an understanding of how crowding-out changed over time, using this DiD approach, while being cognisant of these changes.

3.4 Empirical strategy

To fully appreciate the empirical strategy, it is useful first to understand the evolution of tax increases in South Africa between 1994 and 2010. The South African government began implementing large tobacco-tax increases after 1994. In that year, the excise tax burden as a proportion of the retail price of cigarettes was approximately 21% while the total tax burden, including value-added tax (VAT) was 33% (Linegar & van Walbeek, 2018). The government targeted a total tax threshold of 50% with increases implemented over an unspecified period, beginning in 1994. Between 1995 and 2000 the total tax burden increased by 10 percentage points from 36.4% to 46.4%. The tax burden increased to 51.5% in 2005/6. This was further increased to 52% in 2006 and maintained by inflation-adjusted increases of between three to four per cent at a time.

As this study considers tax increases to be the treatment, the period of interest is restricted to between 1995 and 2005-06. This affords an opportunity to adopt a quasi-experimental design with a basic two-group and two-time-period framework. However, a stepwise approach is used, in which the 1995 and 2000 surveys are analysed first, followed by a comparison of the 2000 and 2005-06 surveys. A potentially important distinction in this regard is that of exposure time and "dose" between the two comparisons. Dose in this case refers to the percentage increase in the real price of cigarettes due to increases in the excise tax rate between the two periods.

In the first comparison, all households in the baseline year (1995) are subject to a relatively modest excise-tax burden and are faced with relatively low real cigarette prices. Households that spend on cigarettes experience repeated "treatments" in the form of tobacco-price increases over the first five-year interval between 1995 and 2000. Households in the 2000 survey thus act as the "post-intervention" observations. As described earlier, the total tax burden increased by about 10 percentage points, and the real price of cigarettes increased by 67% over this period; this is the treatment for the first comparison. The next step compares households from the 2005/06 survey to those from the 2000 survey. Thus, households in the year 2000 are now considered as the baseline, despite having already experienced prior tax increases. The total tax

burden over this second period increased by just over five percentage points, and the real price increased by 22%.

Strictly speaking, the analysis of the impact of the treatment in the second comparison could be regarded as a kind of "additional impact", over and above the tax increases already experienced. Ideally, however, in order to determine the impact in this incremental sense, one would have to follow the same households through time from 1995 to 2000 and then to 2005-06. This requirement is not viable, as there is no way of telling if the same households are used across repeated cross-sectional surveys. In this context, therefore, the second comparison effectively ignores the fact that the households experienced any prior tobacco tax increases.

The ensuing analysis was conducted using three specifications. First, a simple OLS regression, as presented in Equation 3, is estimated. Next, Equation 3 is augmented to include a vector of control variables as follows:

$$w_{it} = \alpha + \beta t + \delta D_{it} + \tau(D_{it} \times t) + \theta X' + \varepsilon_{it} \qquad (4)$$

where X includes control variables such as household size, number of children in the household, mean age of adults in the household, income, an alcohol use dummy, gender, race, and province dummies while θ is the corresponding set of parameters to be estimated.

Importantly, the fundamental parallel trends assumption may be violated because various other factors could influence the variations in outcomes. To address this concern, the third stage of this analysis employs a semi-parametric approach that incorporates covariates. A common method used for this purpose in the literature is the Matched Difference-in-Differences (MDiD) estimator (Heckman, J., Ichimura & Todd, 1997, Heckman, J. et al., 1998). This estimator uses a non-parametrically calculated propensity score to match the treated and control groups before applying the standard DiD procedure described earlier.

However, it is important to note a potential drawback of this estimator, which arises from its use of the propensity score matching (PSM) technique in the first stage. King, Gary & Nielsen (2019) show that for datasets that are initially balanced on covariates, pruning observations via PSM may lead to increased imbalance and possibly bias the effect estimate of interest – a phenomenon they refer to as the "PSM paradox". An alternative estimator proposed by Abadie (2005) also calculates a propensity score non-parametrically, but uses a direct weighting scheme, akin to the Horvitz-Thompson statistic (Horvitz & Thompson, 1952), to estimate the effect of treatment on the treated. Although both the MDiD and Abadie (2005) estimators avoid

estimation errors due to functional form misspecification, only the latter has the additional advantage of avoiding the potential bias associated with the PSM paradox.

The assumptions and formal presentation of how DiD is used to show the change in crowding-out on households that consumed tobacco is presented in line with Abadie (2005). Furthermore, although Abadie (2005) original presentation (2005) distinguished three data types, namely (1) repeated outcomes, (2) repeated cross-sections, and (3) multilevel treatments, the discussion below follows only his exposition of repeated cross-sections, as that constitutes the data structure used for this chapter. The presentation also picks up from the two-group, two-period setting discussed earlier. The ATT is identified under two main assumptions. The first is a conditional parallel trends assumption which, as for the usual parallel trend assumption in DiD models discussed earlier, states that, conditional on covariates, the average outcomes for treated and control would have followed parallel paths, had there not been treatment. The second assumption is that the support of the propensity score constitutes a subset of the support of the propensity score of the untreated. In simpler terms, this second assumption implies that the range of propensity scores (probabilities of receiving treatment) for the group that received the treatment must fall entirely within the range of propensity scores for the group that did not receive the treatment. This ensures that there is some overlap in the propensity scores between the treated and untreated groups, allowing for a meaningful comparison. A further assumption, specific to the repeated cross-section presentation, is that conditional on t=0 (t=1), the data are independent and identically distributed from the distribution of outcomes for the control (treated), treatment, and covariates. In other words, at the baseline, the probability of selecting a household into the control group is equally distributed, while this is true for households in the treatment group in the second time-period. If all three assumptions hold, the ATT is identified as follows (Abadie, 2005, Chang, 2020):

$$E\left[\frac{P(D=1|X)}{P(D=1)}\cdot\varphi_0\cdot W\right] = E[w_{i1} - w_{i0}\,|D=1] \tag{5}$$

where

$$\varphi_0 = \frac{T_i - \lambda_0}{\lambda_0\cdot(1-\lambda_0)}\cdot\frac{D_i - P(D_i=1|X_i)}{P(D_i=1|X_i)\cdot P(D_i=0|X_i)} \tag{6}$$

$\lambda_0 \equiv P(T_i = 1)$ represents the proportion of observations sampled in the post-treatment period. Equation 5 computes a weighted average budget share (W) adjusted using weights derived

from the ratio of the propensity score $[P(D = 1|X)]$ to the probability of treatment $[P(D = 1)]$. In addition, it includes an extra term (φ_0) that accounts for the sample sizes of treated and untreated groups in the post-treatment period.

As this is a case of repeated cross-sections, I was unable to use the Stata user-written programme, *absdid* (Houngbedji, 2016).[12] Instead, I manually calculated the sample analogue of *phi* (φ_0) following Equation 7 below. To do this, I first estimated the propensity score using logistic regression. I then plugged in the appropriate values in the following equation:

$$\hat{\varphi}_i = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{T_i - \hat{\lambda}}{\hat{\lambda}\cdot\left(1 - \hat{\lambda}\right)} \cdot \frac{D_i - \hat{g}(X_i)}{\hat{g}(X_i)\cdot(1 - \hat{g}(X_i))}\right] \tag{7}$$

Where $\hat{\lambda}$ is the proportion of observations sampled in the post-treatment period and $\hat{g}(X_i)$ is the propensity. The result of Equation 7 was then used to adjust the normal household survey weight variable in the dataset, before estimating the usual DiD with and without covariates.[13]

Additional methodological considerations

1. Endogeneity of the treatment group

In conducting a DiD analysis, ensuring the exogeneity of the treatment group is a critical assumption. This exogeneity assumption implies that treatment assignment should not be influenced by factors related to the outcome, beyond the intervention itself. However, in this analysis, the potential endogeneity of the treatment group assignment warrants careful consideration. While treatment and control groups were defined based on observable characteristics, there may still be unobserved factors influencing group membership that could introduce bias if they are correlated with the outcome.

To address this potential endogeneity, matching was applied prior to estimating the DiD model, with the aim of making the treatment and control groups more comparable based on observable characteristics. Specifically, groups were matched by geographical location such as urban or rural as well as province, household size and composition, (e.g. single versus multi-person

[12] Houngbedji (2016) wrote a Stata programme called *absdid* that calculates the DiD estimate based on Abadie (2005). However, it is applicable in the context of repeated outcomes and is best suited in a panel data context.
[13] Remember Abadie's DiD is a direct weighting scheme. By imposing this weighting scheme over the whole sample, I hope the distributions of covariates among treated and untreated are the same. This is the same premise on which propensity score matching is based. I used the *diff* command in Stata with the adjusted household weights to generate the outputs presented later.

households, and the proportion of adults and children in a household) and various demographic variables such as gender and race. Matching improves balance between the groups by controlling for observed confounders and can help reduce bias from these factors. However, matching alone does not fully address endogeneity concerns, particularly because it relies on the assumption that all relevant confounders are observed and can be accounted for. In practice, it is often difficult to confirm that all confounding variables are observed, as certain unmeasured factors could still differ systematically between the groups.

2. Matching

In addition to endogeneity concerns, it is important to consider the limitations of matching with respect to the parallel trends assumption. The DiD approach assumes that in the absence of treatment, the outcome trajectories of the treatment and control groups would follow parallel trends over time. While matching can make the groups more similar at baseline, it does not guarantee that they will exhibit parallel trends post-treatment. This assumption remains a foundational but untestable aspect of the DiD model; thus, it is important to assess whether observable pre-treatment trends in the outcome variable align between the matched treatment and control groups as a proxy check. In this study, we test for the likelihood of this assumption holding by visually inspecting trends in mean budget shares between treated and control households before tobacco-tax increases were introduced. This is followed up by a more formal test (Rambachan & Roth, 2019, Ryan et al., 2019) that takes advantage of the multiple time-period in this dataset.

3. The parallel trends assumption

While matching was employed to improve comparability between groups based on observable characteristics, it does not guarantee the parallel trends assumption holds, especially in an environment like South Africa, where external factors may influence groups differently. Several reasons for deviations from parallel trends are pertinent here.

First, other nationwide anti-tobacco campaigns and policy changes during the study period likely affected groups unevenly. Policies such as stricter advertising restrictions, price hikes, or intensified public health messaging could disproportionately impact current tobacco users (treatment group) by encouraging cessation or reduced consumption, while non-users would be less directly influenced, potentially leading to divergent trends.

Second, economic fluctuations—including shifts in employment, inflation, and disposable income—may influence consumption behaviours differently. A decline in disposable income, for instance, could lead tobacco users to reduce or quit consumption, while non-users are less likely to alter their behaviour, thereby disrupting parallel trends.

Third, improved healthcare access or greater health awareness (e.g., increased availability of cessation programs, heightened awareness of smoking's health risks) may differentially affect tobacco users. Access to cessation support might lead more treatment group members to quit, affecting trends in outcomes for this group in ways that do not similarly impact the control group.

Fourth, changing cultural and social norms around smoking during the period may have influenced treatment and control groups unevenly. For instance, rising stigma or shifts in social attitudes towards smoking could more strongly discourage tobacco users, prompting behavioural changes not observed in non-users, thereby violating the parallel trends assumption.

Lastly, tobacco industry responses, such as price adjustments, targeted promotions, or new product introductions, could further disrupt parallel trends. A price reduction or promotion aimed at retaining current users, for example, would likely have a larger impact on the treatment group, while non-users remain unaffected.

4. Standard errors

Finally, a note on the calculation of standard errors: ideally, standard errors in this context would need to account for clustering, following Bertrand, Duflo & Mullainathan (2004) and Cameron & Miller (2015). However, the datasets used were inconsistent in providing a suitable variable that could be used as the clustering variable.[14] Finally, the estimations were bootstrapped, to ensure the robustness of the estimated results and appropriately account for uncertainty. The procedure involved 1,000 replications of the estimation process, with a fixed random seed (12345) to ensure reproducibility. For each bootstrap iteration, the dependent variables—*food*, *tobacco*, *alcohol*, *clothing*, *education*, *housing*, *transportation*, *communication*, *health*, *household goods*, *recreation*, and *miscellaneous expenditures*—were

[14] For instance, although the 1995 IES has a *district* variable with 358 unique values that could be an appropriate clustering variable and the 2000 survey has a variable called *primary sampling unit (PSU) number* with 2,956 unique values, the 2005-06 survey has neither district nor PSU variables.

regressed on the interaction term of time and the treatment dummy, as well as a set of covariates.

3.5 Data

Data for this chapter were obtained from South Africa's Income and Expenditure Surveys (IESs) conducted in 1995, 2000 and 2005-06. The surveys are nationally representative and were conducted by Statistics South Africa every five years using a two-stage stratified sampling design. Several important changes in the survey collection methods and tools occurred over the period (Yu, D., 2008, Yu, D., 2016), so that ensuring comparability of the surveys over time is a non-trivial task. Further complicating this are various data-quality issues, including changes in sampling procedures (Kerr & Wittenberg, 2015, Machemedze, Kerr & Wittenberg, 2014) and the presence of outliers and influential data points on key variables (Wittenberg, 2017a, Wittenberg, 2017b). These issues are discussed in an earlier chapter and addressed appropriately.

Despite these challenges, the final datasets used are comparable once a meticulous protocol of data cleaning is followed. A rich array of information is available in the datasets that includes geographic, demographic, socioeconomic, and household level expenditure data. These allow for a decent level of disaggregation and analysis of the topic.

3.5.1 Descriptive data

The descriptive statistics for the final datasets are provided in Table 3.1. The data show an expected decline in the proportion of households that reported tobacco consumption between 1995 and 2010. About 46.7% of households reported tobacco consumption in 1995, 34% in 2000, 37.4% in 2005/06, and 23.6% in 2010/11. This is in line with reported declines in tobacco consumption from various surveys over this period (Reddy et al., 2015, Vellios & Van Walbeek, 2016).

Table 3.1: Descriptive statistics from the Income and Expenditure Surveys (IES) 1995, 2000 and 2005-06

	IES 1995			IES 2000			IES 2005-06			IES 2010-11		
Variable	**Tob**	**No Tob**	**Full 1995**	**Tob**	**No Tob**	**Full 2000**	**Tob**	**No Tob**	**Full 2005-06**	**Tob**	**No Tob**	**Full 2010-11**
Number of households	8,751	9,546	18,297	9,181	16,906	26,087	8,157	12,852	21,009	6,030	18,803	24,833
Tobacco consuming households (%)	100.0	0.0	46.7	100.0	0.0	34.0	100.0	0.0	37.4	100.0	0.0	23.6
Non-tobacco consuming households (%)	0.0	100.0	53.3	0.0	100.0	66.0	0.0	100.0	62.6	0.0	100.0	76.4
Households in rural areas (%)	42.2	45.7	44.1	30.3	35.6	33.8	31.7	37.1	35.1	25.3	35.1	32.8
Households in urban areas (%)	57.8	54.3	55.9	69.7	64.4	66.2	68.3	62.9	64.9	74.7	64.9	67.2
Mean monthly tobacco expenditure*	150.1	0.0	70.2	173.6	0.0	59.1	116.8	0.0	43.6	178.0	0.0	42.0
Mean monthly household expenditure*	7,112.5	7,396.2	7,263.6	4,860.9	4,193.1	4,420.4	5,230.6	6,117.0	5,785.8	7,604.7	7,762.7	7,725.5
Mean household size	4.5	4.3	4.4	3.9	3.9	3.9	3.9	3.8	3.8	3.7	3.9	3.9
Mean number of kids	1.4	1.5	1.4	1.1	1.3	1.3	1.2	1.3	1.2	1.0	1.3	1.2
Mean number of adults	3.1	2.8	3.0	2.8	2.5	2.6	2.9	2.6	2.7	2.9	2.8	2.8
Mean age of kids#	7.5	7.6	7.5	7.3	7.4	7.4	0.0	0.0	0.0	6.9	7.0	7.0
Mean age of adults#	37.5	38.6	38.1	37.7	37.1	37.3	0.0	0.0	0.0	38.3	37.5	37.7
African-headed households (%)	66.2	73.4	70.0	70.2	82.4	78.3	72.5	80.0	77.2	64.1	80.4	76.6
Coloured-headed households (%)	12.3	4.7	8.3	14.0	5.0	8.1	12.8	4.8	7.8	18.6	5.4	8.5
Indian-headed households (%)	2.9	2.4	2.6	2.8	2.2	2.4	2.9	2.3	2.5	2.5	2.6	2.6
White-headed households (%)	18.7	19.4	19.1	12.9	10.3	11.2	11.9	12.9	12.5	14.8	11.6	12.3
Female-headed households (%)	20.0	41.1	31.3	41.9	52.8	49.1	41.7	50.5	47.2	27.2	43.0	39.3
Male-headed households (%)	80.0	58.9	68.7	58.1	47.2	50.9	58.3	49.5	52.8	72.8	57.0	60.7

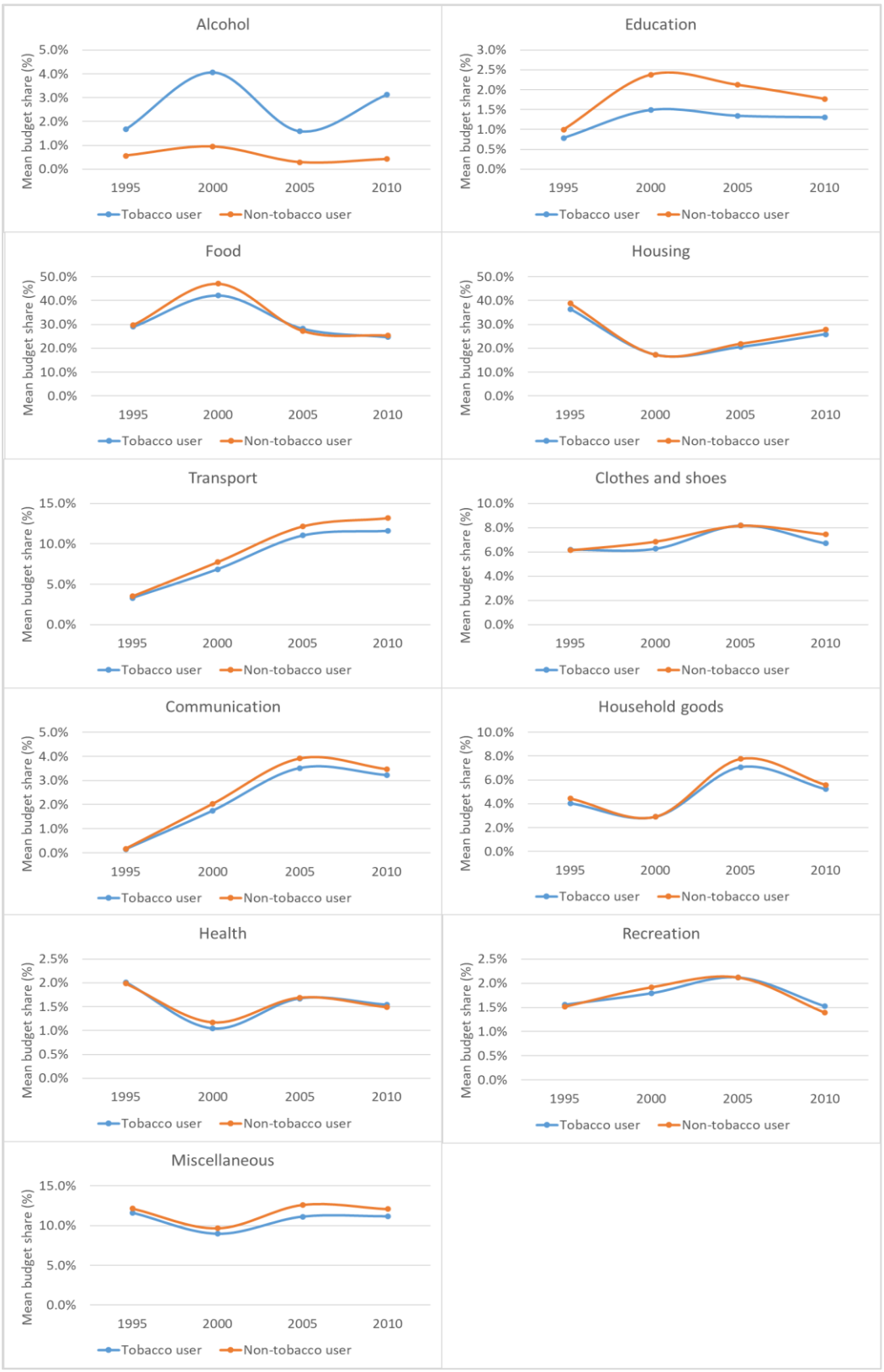
*Expenditures prior to the IES 2010/2011 were all inflated to March 2011. # Age in the 2005-06 IES was captured as a categorical variable, hence no mean was calculated.

3.5.2 Testing for violations of the parallel trends assumption

The key identifying assumption in DiD is that the budget shares in the treated households would have had similar trends to the budget shares in control households in the absence of treatment. A simple way to check the validity of this assumption is to inspect visually trends in mean budget shares between treated and control households before tobacco-tax increases were introduced – that is, comparing the mean budget shares of various expenditure categories for households with tobacco expenditure and of those without. This is shown in Figure 3.2 below for all expenditure categories defined earlier over the period 1995 to 2010. Note that, although 2010 was not used in the analysis, it is included here in order to provide a more complete picture of the trends. In Figure 3.2, all categories evidently violate the parallel trends assumption, except for miscellaneous expenditures.

This method of visual inspection, however, does not consider the change in composition among the non-tobacco-spending households. Instead, in each subsequent year after 1995, former smokers are included in the pool of non-smokers, and may bias the trends towards the smokers. For instance, if there are systematic differences between never-smokers (households that never took up tobacco expenditure), and former smokers (households that no longer incur tobacco expenditures) such that former smokers are similar to current smokers (households that currently incur tobacco expenditures), then the presence of former smokers in the pool of non-tobacco-spending households will make the average in this pool similar to that of tobacco-spending households. At best, a visual comparison of the mean budget shares would only provide a crude – possibly biased – determination of the likelihood of parallel trends. Unfortunately, isolating the trend of households that could be classified as never-smokers throughout the period is not possible without the use of a panel dataset to track the same households from one period to the next.

Figure 3.2: Trends in mean budget shares disaggregated by tobacco use



To address the limitation of the visual inspection method, I borrow an approach from recent developments in the multiple time-period DiD literature that involves a more formal test

(Rambachan & Roth, 2019, Ryan et al., 2019). In its typical form, the test would include lags and leads of the treatment effect in an OLS regression. However, the starting point is first to generalise the common-trends assumption to include all time periods. Thus, for all $t \in \{1, \dots, \bar{t}\}$, let

$$[E(w_{it}(0)|D_{it} = 1, T_i = t) - E(w_{it}(0)|D_{it} = 1, T_i = t - 1)] = [E(w_{it}(0)|D_{it} = 0, T_i = t) - E(w_{it}(0)|D_{it} = 0, T_i = t - 1)] \quad (8)$$

In other words, Equation 8 means that the evolution of the mean outcome of the treatment and control groups between each pair of consecutive time periods is the same. In fact, by extension, the treatment and control groups experience the same evolution of the outcomes from period 0 to t, for any t. If we further assume that treatment happens at time t_0, where $1 < t_0 < \bar{t} - 1: 1\{D_{it} = 1, T_i \geq t_0\}$, then the common trends assumption shown by Equation 8 has testable implications.

To test the validity of the parallel-trends assumption, I first estimate the following OLS regression:

$$w_{it} = \alpha + \sum_{t=1}^{4} \beta_t t + \delta D_{it} + \sum_{t=1}^{4} \gamma_{1t} D_{it} t + \varepsilon_{it} \quad (9)$$

where the parameters w_{it}, D_{it}, t and ε_{it} are as earlier defined and α, β, δ and γ are coefficients to be estimated.

Importantly, for Equation 8's assumption to be valid, the coefficient of γ must be zero. Thus, $\gamma_{1t} = 0$ for $t < t_0$ and $\gamma_{1t} = E(w_{it}(1) - w_{it}(0)|D_{it} = 1, T_i = t)$ for $t \geq t_0$. The coefficient γ can be used to test for the parallel trends assumption with the null hypothesis that $\gamma_{1t} = 0$ and an alternative hypothesis that $\gamma_{1t} \neq 0$. If the null hypothesis, $\gamma_{1t} = 0$ for $t < t_0$ is rejected, this indicates a violation of the parallel-trends assumption, and to fail to reject the null implies the parallel-trends assumption holds.

Table 3.2: Test for violation of Parallel Trends Assumption

	F-stat	p-value	Interpretation
Food	51.12	0.00	PTA violated
Alcohol	509.14	0.00	PTA violated
Clothes & shoes	14.07	0.00	PTA violated
Education	30.07	0.00	PTA violated
Housing	10.37	0.00	PTA violated
Transport	13.93	0.00	PTA violated
Communication	20.85	0.00	PTA violated
Health	**2.29**	**0.08**	**PTA not violated**
Household goods	**1.89**	**0.13**	**PTA not violated**
Recreation	11.54	0.00	PTA violated
Miscellaneous	**2.58**	**0.05**	**PTA not violated**

PTA= Parallel Trends Assumption

Table 3.2 presents the results of the joint F-tests performed after running the regression with Equation 9's specification using the four IES's of 1995, 2000, 2005/06 and 2010/11. The results show that the budget share on household goods, and to a lesser extent health and miscellaneous expenditures, all fail to reject the null hypothesis. This implies that all budget share categories violated the parallel-trends assumption, except for the three categories highlighted (health, household goods, and miscellaneous). As noted in the visual inspection done earlier, the miscellaneous category was not in violation of the parallel-trends assumption.

Both the visual inspection and the formal tests confirm the initial suspicions about the violation of parallel trends in the analysis conducted. It thus strengthens the motivation for the use of Abadie (2005) reweighting procedure to correct for this violation.

3.5.3 Reweighting the data using Abadie's 2005 procedure

An abrogation of the parallel-trends assumption means that DiD is unable to identify the ATT without some remedial measure. As explained earlier, this chapter uses Abadie (2005) DiD method to do this. It employs a direct weighting procedure that uses a propensity score estimator to ensure that the distribution of covariates is the same in both control and treatment groups. Once the reweighting procedure is implemented, one expects the distribution of various descriptive statistics to be similar between tobacco-user (our treated group) and non-tobacco-user households (our control). This procedure is conducted in Stata (StataCorp).

Table 3.3 presents the difference between these households, both prior to and after the reweighting procedure was implemented. In effect, this is an assessment of balance between treatment and control groups and effectively rules out these factors as confounders in determining the ATT, as shown in the DiD results presented in later sections. From the table it can be seen that, in nearly all the statistics assessed, the difference is much smaller after the reweighting procedure. In fact, in some cases, the difference loses its statistical significance after reweighting. This demonstrates that the reweighting procedure worked as intended, by making the appropriate adjustments in covariates between treated and control groups, and in so doing, making a stronger case to conclude that any differences in outcomes may be attributed to the treatment.

It is important to note that while reweighting aims to reduce differences in covariates, achieving statistical insignificance across all variables is not necessary. The key objective is to minimise these differences to the extent that they no longer meaningfully bias the estimation of treatment effects. In this case, the reduction in differences after reweighting provides confidence in the robustness of the analysis and the reliability of the estimated treatment effects.

To further ensure the robustness of the DiD analysis, a bootstrapping procedure is employed. In this study, bootstrapping is implemented with 1,000 repetitions, which allows for the calculation of more reliable standard errors for the estimated treatment effects (ATT). This approach accounts for potential sampling variability and provides a measure of the precision of the estimates.

The bootstrapping procedure is conducted in Stata (StataCorp) and is applied to the Abadie-weighted DiD estimator. Each iteration of the bootstrap replicates the entire estimation process, including the reweighting procedure using propensity scores and the calculation of ATT. The resulting bootstrapped standard errors are robust to heteroskedasticity and provide more accurate confidence intervals for the treatment effects, even when sample sizes are moderate or when the underlying data exhibit non-normality.

Table 3.3: Difference in descriptive statistics between tobacco user and non-tobacco user households before and after reweighting procedure

	1995 vs. 2000						2000 vs. 2005/06					
	Pre-reweighting			Post-reweighting			Pre-reweighting			Post-reweighting		
Variable	**Diff**	**t-stat**	**p-value**	**Diff**	**t-stat**	**p-value**	**Diff**	**t-stat**	**p-value**	**Diff**	**t-stat**	**p-value**
Households in rural areas	-3.0	33.5	0.0	1.5	-0.6	0.5	-5.2	39.8	0.0	0.3	-2.9	0.0
Households in urban areas	3.0	26.7	0.0	-1.5	3.4	0.0	5.2	29.7	0.0	-0.3	-2.4	0.0
Proportion of multi-person households	3.1	22.5	0.0	-2.1	-3.7	0.0	1.9	21.6	0.0	-2.1	-4.8	0.0
Proportion of single-person households	-3.1	34.6	0.0	2.1	4.2	0.0	-1.9	40.1	0.0	2.1	-1.7	0.1
Average monthly tobacco expenditure	160.8	-90.6	0.0	161.5	-88.0	0.0	142.1	-63.4	0.0	139.2	-61.8	0.0
Average monthly household expenditure	488.3	-3.0	0.0	-618.0	2.6	0.0	-284.7	1.7	0.1	-989.59	3.9	0.0
Average monthly household income	467.3	-3.2	0.0	-558.7	3.7	0.0	-659.0	2.7	0.0	-507.3	2.3	0.0
Average household size	0.2	-6.2	0.0	0.0	-1.4	0.2	0.1	-3.3	0.0	0.1	-2.5	0.0
Percentage of children in household	-3.0	12.3	0.0	0.0	3.2	0.0	-3.9	12.8	0.0	-1.5	4.1	0.0
Percentage of adults in household	3.0	-12.3	0.0	0.0	-3.2	0.0	4.0	-12.9	0.0	1.5	-4.1	0.0
Average age of adults in household	-0.1	0.6	0.6	-0.2	1.8	0.1	0.7	-3.5	0.0	0.3	-1.2	0.2
Average age of household head	-0.5	2.8	0.0	-0.7	3.5	0.0	0.1	-0.2	0.9	-0.7	1.4	0.2
Proportion of African-headed households	-10.6	48.7	0.0	1.6	0.6	0.6	-9.6	50.9	0.0	1.6	-4.5	0.0
Proportion of coloured-headed households	8.2	-18.8	0.0	-0.2	1.2	0.2	8.4	-9.5	0.0	-0.6	-0.2	0.9
Proportion of Indian-headed households	0.6	3.3	0.0	0.1	-0.3	0.8	0.6	3.6	0.0	0.2	-1.2	0.2
Proportion of white-headed households	1.9	9.5	0.0	-1.5	3.6	0.0	0.5	11.5	0.0	-1.2	0.8	0.5
Proportion of female-headed households	-17.9	53.1	0.0	2.3	-2.3	0.0	-9.8	43.8	0.0	0.3	-2.7	0.0
Proportion of male-headed households	17.9	9.0	0.0	-2.3	4.3	0.0	9.8	21.8	0.0	-0.3	-2.3	0.0
Proportion in Western Cape	5.8	-5.3	0.0	-0.2	1.2	0.2	5.9	0.6	0.6	-1.1	0.4	0.7
Proportion in Eastern Cape	-1.6	22.2	0.0	-0.7	3.1	0.0	-0.5	19.9	0.0	0.7	-2.9	0.0
Proportion in Northern Cape	1.4	-7.2	0.0	-0.1	1.0	0.3	1.6	-3.7	0.0	0.0	-1.3	0.2
Proportion in Free State	3.3	-1.1	0.0	0.0	1.0	0.3	3.6	2.3	0.0	0.2	-1.6	0.1
Proportion in Kwazulu-Natal	-5.4	30.8	0.0	0.2	0.8	0.4	-6.9	31.3	0.0	0.3	-1.8	0.1
Proportion in North-West	2.0	5.3	0.0	0.8	-1.7	0.1	1.8	7.1	0.0	0.2	-1.5	0.1
Proportion in Gauteng	0.2	15.2	0.0	-0.2	1.3	0.2	0.1	19.0	0.0	-0.4	-1.1	0.3
Proportion in Mpumalanga	-1.0	17.6	0.0	-0.2	1.5	0.1	-1.1	19.3	0.0	-0.1	-0.8	0.4
Proportion in Limpopo	-4.7	29.3	0.0	0.4	-0.8	0.4	-4.5	26.9	0.0	0.2	-1.7	0.1

Note that Diff is calculated as the difference in means is calculated for tobacco HH minus mean for non-tobacco HH.

The descriptive statistics for pooled samples are presented in Tables 3.4 and 3.5. Table 3.4 reports on the pooled sample from the 1995 and 2000 Income and Expenditure Surveys (IES), while Table 3.5 focuses on the pooled sample from the 2000 and 2005/06 IES datasets. Each table compares households that report tobacco spending with those that do not, before and after the reweighting procedure using Abadie (2005)'s method.

In Table 3.4, approximately two-thirds of households are urban-based, with minimal variation between tobacco-spending and non-tobacco-spending households both before and after reweighting. For instance, 63.2% of tobacco-spending households and 60.2% of non-tobacco-spending households resided in urban areas before reweighting, which shifted to 63.6% and 65.1% respectively after reweighting. While these differences appear minor, the high statistical significance of differences prior to reweighting ($p < 0.001$) underscores potential structural disparities between these groups that may complicate causal inference.

Household composition also varied, with multi-person households dominating across both groups (82.2%–87.1%), yet single-person households constituted a significantly larger share of non-tobacco-spending households prior to reweighting (17.8%) compared to tobacco-spending households (14.7%). After reweighting, the gap narrowed (12.9% vs. 15.0%), suggesting that the reweighting process partially aligns group characteristics.

One notable demographic difference relates to race. African-headed households comprised 78.6% of non-tobacco-spending households prior to reweighting, compared to 68.0% of tobacco-spending households. This aligns with previous findings that tobacco use among African households is lower than among other racial groups (Reddy, Meyer-Weitz & Yach, 1996, Reddy et al., 2015). After reweighting, these proportions converged (66.5% vs. 68.1%), though the persistence of racial disparities suggests unobserved factors may still be influencing outcomes.

The regional distribution of households mirrored South Africa's population density, with higher concentrations in Gauteng, KwaZulu-Natal, and the Western Cape and lower densities in the Northern Cape, Free State, and North-West provinces. Statistically significant differences in regional proportions between groups prior to reweighting diminished post-reweighting, indicating improved comparability.

Table 3.5, summarising the 2000–2005/06 pooled data, reflects similar patterns to Table 3.4, though with minor differences. Average household size was slightly smaller (3.9 for tobacco-using households versus 3.8 for non-using households), and the proportion of female-headed

households rose notably, exceeding 40% in this period compared to approximately 30% in the earlier dataset.

Table 3.4: Descriptive statistics for comparison for pooled sample of 1995 and 2000 Income and Expenditure Surveys

	Pre-reweighting				Post-reweighting			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variable	Tob	No Tob	t-stat	p-val	Tob	No Tob	t-stat	p-val
Number of households	23,376	32,411			23,376	32,411		
Households in rural areas	36.8	39.8	33.5	0.0	36.4	34.9	-0.6	0.5
Households in urban areas	63.2	60.2	26.7	0.0	63.6	65.1	3.4	0.0
Proportion of multi-person households	85.3	82.2	22.5	0.0	85.0	87.1	-3.7	0.0
Proportion of single-person households	14.7	17.8	34.6	0.0	15.0	12.9	4.2	0.0
Average monthly tobacco expenditure	160.8	0.0	-90.6	0.0	161.5	0.0	-88.0	0.0
Average monthly household expenditure	6,248	5,760	-3.0	0.0	6,173	6,791	2.6	0.0
Average monthly household income	7,425	6,957	-3.2	0.0	7,361	7,920	3.7	0.0
Average household size	4.21	4.1	-6.2	0.0	4.2	4.2	-1.4	0.2
Percentage of children in household	22.0	25.0	12.3	0.0	22.0	23.0	3.2	0.0
Percentage of adults in household	78.0	75.0	-12.3	0.0	78.0	77.0	-3.2	0.0
Average age of adults in household	37.6	37.7	0.6	0.6	37.6	37.9	1.8	0.1
Average age of household head	46.0	46.5	2.8	0.0	45.9	46.6	3.5	0.0
Proportion of African-headed households	68.0	78.6	48.7	0.0	68.1	66.5	0.6	0.6
Proportion of coloured-headed households	13.1	4.9	-18.8	0.0	13.1	13.3	1.2	0.2
Proportion of Indian-headed households	2.9	2.3	3.3	0.0	2.9	2.7	-0.3	0.8
Proportion of white-headed households	16.1	14.2	9.5	0.0	15.9	17.5	3.6	0.0
Proportion of female-headed households	30.0	48.0	53.1	0.0	30.7	28.4	-2.3	0.0
Proportion of male-headed households	70.0	52.0	9.0	0.0	69.3	71.6	4.3	0.0
Proportion in Western Cape	13.4	7.5	-5.3	0.0	13.4	13.6	1.2	0.2
Proportion in Eastern Cape	12.9	14.5	22.2	0.0	12.8	13.5	3.1	0.0
Proportion in Northern Cape	2.7	1.3	-7.2	0.0	2.7	2.7	1.0	0.3
Proportion in Free State	9.1	5.9	-1.1	0.0	9.1	9.1	1.0	0.3
Proportion in Kwazulu-Natal	15.4	20.8	30.8	0.0	15.3	15.2	0.8	0.4
Proportion in North-West	8.9	6.9	5.3	0.0	8.9	8.1	-1.7	0.1
Proportion in Gauteng	24.6	24.4	15.2	0.0	24.8	25.0	1.3	0.2
Proportion in Mpumalanga	5.5	6.5	17.6	0.0	5.5	5.7	1.5	0.1
Proportion in Limpopo	7.5	12.2	29.3	0.0	7.5	7.0	-0.8	0.4

While reweighting improved balance across observable characteristics, it is critical to remember that matching alone cannot address all potential violations of key assumptions, such

as parallel trends or the influence of unobserved confounders. For example, systematic differences in unobserved characteristics, such as household attitudes toward health, regional policy enforcement, or economic shocks, may persist post-reweighting and introduce bias into the analysis.

Moreover, the descriptive statistics highlight shifts in characteristics, such as gender composition and household size, which may reflect broader socio-economic changes unrelated to tobacco expenditure. These factors could interact with treatment effects in unforeseen ways, complicating causal interpretation.

The persistence of significant differences in characteristics such as racial composition and urban-rural distribution after reweighting suggests that some underlying disparities remain unaddressed. These disparities, coupled with the dynamic policy and economic context of South Africa during the study period, necessitate caution in interpreting results. Policy implications derived from these findings must consider these limitations and be framed within the broader socio-economic and policy landscape, acknowledging the potential for residual confounding effects.

Table 3.5: Descriptive statistics for comparison pooled sample of 2000 and 2005-06 Income and Expenditure Surveys

	Pre-reweighting				Post-reweighting			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variable	Tob	No Tob	t-stat	p-val	Tob	No Tob	t-stat	p-val
Number of households	17,413	29,923			17,413	29,923		
Households in rural areas	31.1	36.3	39.8	0.0	31.2	30.8	-2.9	0.0
Households in urban areas	68.9	63.7	29.7	0.0	68.8	69.2	-2.4	0.0
Proportion of single person households	82.2	80.3	21.6	0.0	82.3	84.4	-4.8	0.0
Proportion of non-single person households	17.9	19.7	40.1	0.0	17.7	15.6	-1.7	0.1
Average monthly tobacco expenditure	142.1	0.0	-	0.0	139.2	0.0	-	0.0
Average monthly household expenditure	5,158	5,442	63.4	0.1	5,182	6,172	61.8	0.0
Average monthly household income	7,135	7,793	1.7	0.0	7,216	7,723	3.9	0.0
Average household size	3.9	3.8	2.7	0.0	3.9	3.8	2.3	0.0
Percentage of children in household	20.4	24.4	-3.3	0.0	20.4	21.9	-2.5	0.0
Percentage of adults in household	79.6	75.7	12.8	0.0	79.6	78.1	4.1	0.0
Average age of adults in household	37.7	37.1	-	0.0	37.7	37.5	-4.1	0.0
Average age of household head	43.1	43.0	12.9	0.0	43.1	43.7	-1.2	0.2
Proportion of African-headed households	71.4	80.9	-3.5	0.0	71.5	69.9	1.4	0.2
Proportion of Coloured-headed households	13.3	4.9	-0.2	0.9	13.2	13.8	-4.5	0.0
Proportion of Indian-headed households	2.9	2.2	50.9	0.0	2.9	2.6	-0.2	0.9
Proportion of White-headed households	12.5	11.9	-9.5	0.0	12.4	13.7	-1.2	0.2
Proportion of female-headed households	41.8	51.5	3.6	0.0	41.8	41.5	0.8	0.5
Proportion of male-headed households	58.2	48.5	11.5	0.0	58.3	58.6	-2.7	0.0
Proportion in Western Cape	13.7	7.8	43.8	0.0	13.6	14.7	-2.3	0.0
Proportion in Eastern Cape	13.2	13.7	0.6	0.6	13.3	12.7	0.4	0.7
Proportion in Northern Cape	3.1	1.5	21.8	0.0	3.2	3.1	-2.9	-
Proportion in Free State	9.1	5.5	-3.7	0.0	9.2	9.1	-1.3	0.2
Proportion in Kwazulu-Natal	13.8	20.7	2.3	0.0	13.8	13.5	-1.6	0.1
Proportion in North-West	8.4	6.6	31.3	0.0	13.8	13.5	-1.8	0.1
Proportion in Gauteng	25.8	25.7	7.1	0.0	8.4	8.2	-1.5	0.1
Proportion in Mpumalanga	5.8	6.9	19.0	0.0	25.6	25.9	-1.1	0.3
Proportion in Limpopo	7.1	11.5	19.3	0.0	5.8	5.9	-0.8	0.4
			26.9		7.1	7.0	-1.7	0.1

3.6 Empirical results

The DiD results showing the average treatment effect on the treated (ATT) are presented in Tables 3.6 to 3.8. In each of these tables, two panels are presented, a left panel (columns 1 – 4) showing the transition from 1995 to 2000 and a right one (columns 5 – 8) showing transitions from 2000 to 2005/06. The DiD coefficients can be interpreted from two different perspectives. First, following the impact evaluation literature, the coefficients are equivalent to the average treatment effect of the increase in tobacco taxes on the budget share of the respective expenditure category among tobacco-spending households. Another way to look at these DiD coefficients is to consider that they tell us the direction of crowding-out or crowding-in between the first period and the second for tobacco-spending households. Since the budget shares are expressed as a percentage, the DiD estimates obtained represent the percentage point difference in mean budget shares between any two consecutive periods. Of course, mathematically, this is simply calculated as the difference in the difference of the mean budget shares between tobacco spenders and non-spenders prior to and after the tobacco-tax increase.

The results in the tables thus reflect three things: (1) the budget share difference as a measure of the level of crowding-out at baseline; (2) the post-intervention budget share difference as a measure of crowding-out after the tax increase; and (3) the DiD estimate of the ATT as a measure of how the baseline crowding-out changed in the second period among tobacco users. Bear in mind that the estimates for (1) and (2) in this case detect crowding-out if the coefficient is negative and statistically significant. Furthermore, to determine unambiguously the direction of crowding-out, the DiD coefficient must be interpreted relative to the first period coefficients and contingent on statistical significance. For instance, if the DiD coefficient is not statistically significant, then there is no reason to believe crowding-out changed between the two periods. This is indicated as "No change" in the tables. On the other hand, the DiD coefficient can be interpreted as an increase in crowding-out if two conditions are met: (1) crowding-out was detected in the first period; and (2) the DiD coefficient is negative and statistically significant.[15]

[15] The mathematical intuition here is as follows: Since DiD= (Period 2 difference – Period 1 difference), if the Period 1 difference is negative (implying crowding-out in the first period) and the DiD coefficient is also negative, it follows that the Period 2 difference must be negative and greater than the Period 1 difference in absolute terms. This would imply crowding-out increased in the second period. On the other hand, if the DiD coefficient is positive, then the Period 1 level of crowding-out would be higher in absolute terms than the Period 2 figure. Of course, in this second scenario, it is possible that Period 2 may still detect crowding-out (implying a negative coefficient that would be less than that of Period 1 in absolute terms), it may shift to crowding-in (implying a positive coefficient), or may tend towards zero (implying no statistically significant difference between the tobacco consuming and non-consuming households). Following this logic, a positive (negative) DiD coefficient implies a decrease (increase) in crowding-out in the second period, or possibly a complete elimination of it. The same logic is extrapolated if we start with crowding-in in the first period.

Extending this logic further, if crowding-out is detected in the first period and the DiD coefficient is positive and statistically significant, then this can be interpreted as a decrease in crowding-out. Alternatively, if the first period detected crowding-in, then the opposite is true, namely that a positive and statistically significant DiD coefficient would imply an increase in crowding-in for that expenditure category, while a negative statistically significant coefficient would imply a decrease in crowding-in.

3.6.1 Full sample analysis

Results from the analysis of the full sample of households are shown in Table 3.6. In 1995, tobacco-consuming households experienced crowding-out on education, transport, health, household goods and recreation. All these categories continued to experience crowding-out in the subsequent periods. Between 1995 and 2000, education, transport and recreation all showed evidence of increased crowding-out, while health and household goods showed no change. Communication expenditure was crowded-out in the second period, but not the first. Alcohol, housing and, to a lesser extent, communication were all crowded-in during the first period. Crowding-out on transport increased by 0.47 percentage points, while crowding-out on education and recreation increased by 0.37 and 0.25 percentage points. There was increased crowding-in on alcohol (0.57 percentage points) and, interestingly, communication switched from crowding-in to crowding-out. Although not statistically significant in the first instance, there is some evidence to suggest that food also switched from crowding-out to crowding-in over the relevant period.

The evolution of crowding-out from 2000 to 2005/06 is shown in the right panel of Table 3.6. The baseline numbers in this part of the table closely reflect the post-intervention figures in the left panel of the same table. The difference is that the number of observations in the two analyses differ between the two periods (i.e., n = 55,777 in the 1995-2000 period and n = 47,336 in the 2000-2005/6 period). It is also important to keep in mind that smoking rates were declining over the entire period. Consequently, the absolute level of crowding-out or crowding-in during the second period is expected to be influenced by this decline, which may result in a reduction compared to the first period. In the subsequent survey period (2005/06), education, housing, transport, household goods, and miscellaneous expenditures all registered some degree of crowding-out. However, apart from the last category, there was a decrease in the level of crowding-out in the second period compared to the first. The level of crowding-in on alcohol also registered a decline.

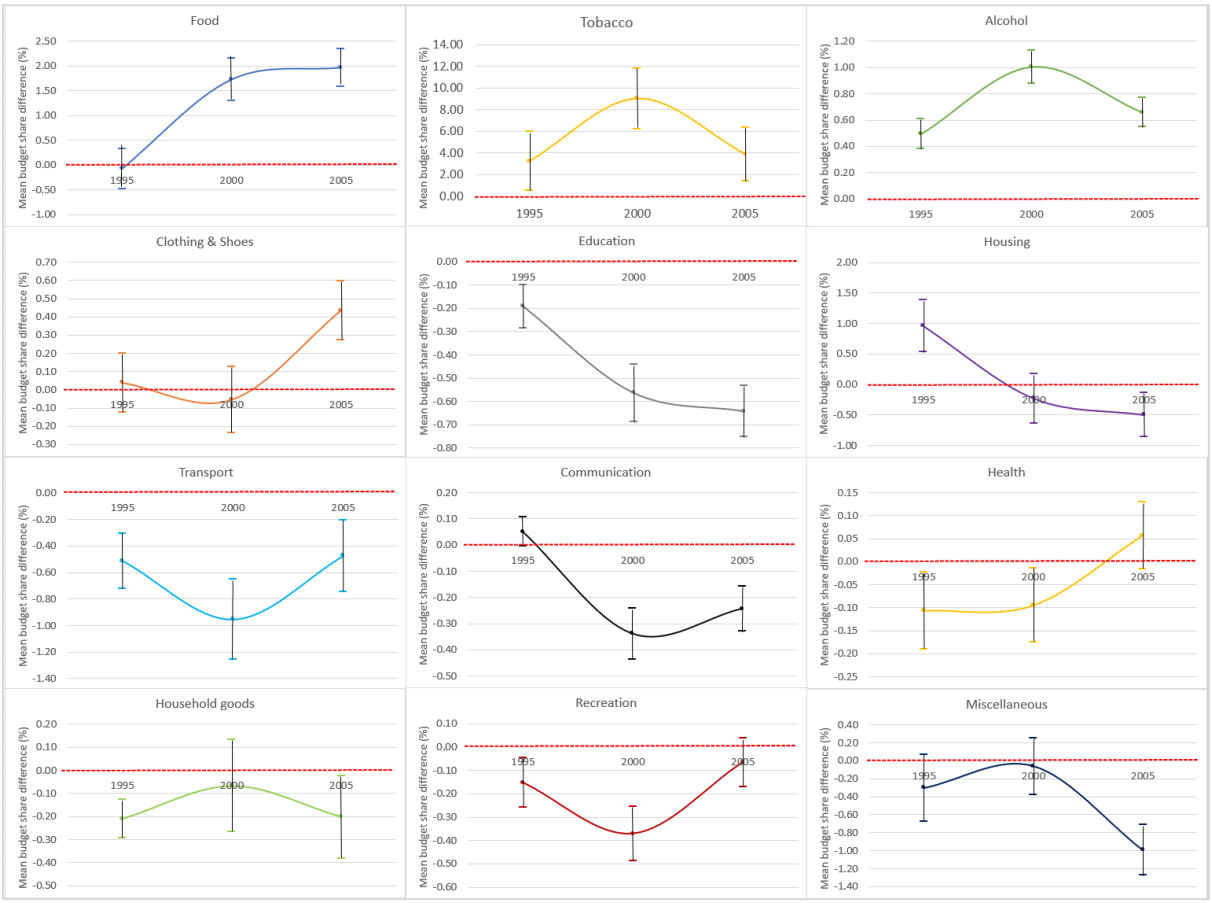
Table 3.6: Crowding-out transitions from 1995 to 2000 IES and from 2000 to 2005/06 IES using all samples

	1995 to 2000				2000 to 2005-06			
	Baseline difference	**Post-intervention difference**	**Diff-in-diff**	**Impact on crowding-out/in**	**Baseline difference**	**Post-intervention difference**	**Diff-in-diff**	**Impact on crowding-out/in**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Food	-0.04	1.47***	1.68***	Increased CI	1.82***	1.97 ***	0.26	No change
	(0.21)	(0.19)	(0.38)		(0.31)	(0.20)	(0.45)	
Tobacco	3.29***	9.28***	5.71**	Increased CI	8.99***	3.91***	-5.07*	Decreased CI
	(0.11)	(1.30)	(2.80)		(2.61)	(1.26)	(2.66)	
Alcohol	0.49***	1.03***	0.57***	Increased CI	1.01***	0.66***	-0.34**	Decreased CI
	(0.05)	(0.05)	(0.13)		(0.12)	(0.06)	(0.16)	
Clothing & shoes	0.013	-0.05	-0.047	No change	-0.07	0.44***	0.50**	Increased CI
	(0.10)	(0.08)	(0.16)		(0.13)	(0.08)	(0.20)	
Education	-0.21***	-0.54	-0.37***	Increased CO	-0.55***	-0.64***	-0.11	No change
	(0.05)	(0.05)	(0.10)		(0.09)	(0.06)	(0.14)	
Housing	1.00***	-0.19	-1.26***	Decreased CO	-0.25	-0.50***	-0.31	No change
	(0.29)	(0.21)	(0.44)		(0.34)	(0.19)	(0.49)	
Transport	-0.53***	-0.95***	-0.47**	Increased CO	-0.98***	-0.47***	0.50	No change
	(0.11)	(0.10)	(0.23)		(0.21)	(0.14)	(0.37)	
Communication	0.05***	-0.34***	-0.37***	Switched: CI to CO	-0.33***	-0.24***	0.05	No change
	(0.01)	(0.03)	(0.06)		(0.06)	(0.04)	(0.11)	
Health	-0.12**	-0.10**	-0.00	No change	-0.11**	0.06	0.18**	Decreased CO
	(0.06)	(0.04)	(0.08)		(0.05)	(0.04)	(0.09)	
Household goods	-0.20*	-0.04	0.12	No change	-0.10	-0.20**	-0.08	No change
	(0.11)	(0.08)	(0.15)		(0.11)	(0.09)	(0.21)	
Recreation	-0.13***	-0.38***	-0.25**	Increased CO	-0.38***	-0.07	0.30**	Decreased CO
	(0.05)	(0.05)	(0.12)		(0.10)	(0.05)	(0.13)	
Miscellaneous	-0.31	0.07	0.39	No change	-0.07	-0.99***	-0.95***	Increased CO
	(0.26)	(0.18)	(0.35)		(0.24)	(0.14)	(0.35)	
Sample size	**54,777**	**54,777**	**54,777**		**47,336**	**47,336**	**47,336**	

CO –Crowding-out; CI – Crowding-in. Bootstrapped standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

The difference in mean budget shares shown in Figure 3.3 represent crowding-out if the graph is below zero, and crowding-in if the graph is above zero. For instance, categories such education and transport that were crowded-out in the 1995 survey, based on the Table 3.6 results, have graphs that start below zero, while the graph for alcohol, which was crowded-in, is above zero. Altogether, these graphs show how crowding-out or crowding-in changed over the period.

Figure 3.3: Trends in the mean budget share difference from 1995 to 2005/06



Note: The bands indicate 95% confidence intervals. Note that values above the red dashed line are indicative of crowding-in, while those below are indicative of crowding-out. Values for the year 2000 are obtained from the baseline difference in budget shares from the 2000 to 2005/06 analysis.

The diagrams in Figure 3.3 present a range of mixed results. Between 1995 and 2005, the mean budget share for food increased by slightly more than two percentage points, marking the largest increase across all expenditure categories. This suggests, on average, that food expenditure experienced a crowding-in effect during this period. However, it is crucial to interpret this result with caution, as other factors, such as changes in household composition or income levels, may have contributed to the observed trend, and thus, the direct association with tobacco consumption alone cannot be conclusively established.

Education expenditure, on the other hand, exhibited the most consistent and negative impact among tobacco-consuming households, with crowding-out effects intensifying over time. While education spending surged during the initial period from 1995 to 2000, this upward trajectory slowed between 2000 and 2005. A plausible explanation for this could be the demographic characteristics of smoking households, such as the higher proportion of single-adult male households, which may be less likely to spend on education relative to other households. Therefore, caution is needed in attributing the observed crowding-out effect solely to tobacco consumption, as household-specific factors could influence this relationship.

Expenditure on alcohol displayed both crowding-in and crowding-out effects, with a peak observed in 2000 followed by a decrease, ultimately returning to a value close to the 1995 baseline by 2005. This fluctuation suggests that alcohol expenditure was subject to influences beyond tobacco consumption, including potential policy changes, shifts in income distribution, or other socio-economic factors during the period under consideration.

Additional categories that registered positive shifts in budget shares, indicating increased expenditure, include health, clothing and shoes, and transport, albeit to varying extents. Health expenditure increased by approximately 0.15 percentage points, transitioning from a crowding-out effect in 1995 to crowding-in in 2005. Similarly, clothing and shoes expenditure showed a more significant increase of 0.40 percentage points, although the baseline levels suggest that crowding-out was not a major issue initially. In the case of transport expenditure, a crowding-out effect was observed in 2000, but this reverted to baseline levels by 2005/06, indicating a recovery that could be influenced by broader economic conditions or policy interventions, rather than tobacco consumption alone.

3.6.2 Analysis disaggregated by urban and rural settlements

To better understand some of the nuances of these transitions, Tables 3.7 and 3.8 show the results disaggregated by rural and urban settlements. Among rural tobacco-consuming households in the 1995 survey, food, clothing and shoes, education, transport, health, household goods, and recreation all registered crowding-out. In addition, crowding-in was detected for alcohol and housing. More categories experienced crowding-out in rural areas than in urban ones. Specifically, only education, transport, recreation, and miscellaneous expenditures were crowded-out in urban areas. On the other hand, food, alcohol, and clothing and shoes all showed crowding-in among urban tobacco-consuming households in that year. Education and transport were more crowded-out among urban households (by 0.29 and 0.58

percentage points respectively) than among rural households (0.18 and 0.51 percentage points, respectively). Crowding-out on recreation was higher among rural households (0.18 percentage points) compared to urban ones (0.11 percentage points), while food was severely crowded-out in rural areas (1.12 percentage point difference), but crowded-in among urban households (0.98 percentage points).

As in the full sample result, urban households registered increased crowding-out on education, suffering at least a 0.50 percentage point increase in crowding-out between 1995 and 2000. On the other hand, transport and recreation both had initial increases in crowding-out between 1995 and 2000, followed by a substantial decrease in crowding-out between 2000 and 2005/06.

Rural households, in comparison, also experienced increased crowding-out on education of approximately 0.20 percentage points. Crowding-out on food, clothing and shoes, transport, health, and household goods all decreased over the same period. In fact, food switched from being crowded-out in 1995 (1.17 percentage point difference) to being crowded-in in 2005 (1.19 percentage point difference) among rural households. The level of crowding-in of alcohol expenditures increased between 1995 and 2000, but decreased between 2000 and 2005/06.

Table 3.7: Crowding-out transitions from 1995 to 2000 IES and from 2000 to 2005/06 IES using rural households

	1995 to 2000				2000 to 2005-06			
	Baseline difference	**Post-intervention difference**	**Diff-in-diff**	**Impact on crowding-out/in**	**Baseline difference**	**Post-intervention difference**	**Diff-in-diff**	**Impact on crowding-out/in**
Food	-1.12***	0.23	1.45**	Decreased CO	0.70	1.19***	0.54	No change
	(0.36)	(0.34)	(0.63)		(0.53)	(0.33)	(0.72)	
Tobacco	3.41***	6.51***	3.10***	Increased CI	6.36***	3.76***	-2.60***	Decreased CI
	(0.06)	(0.11)	(0.23)		(0.22)	(0.13)	(0.26)	
Alcohol	0.36***	0.81***	0.49**	Increased CI	0.66***	0.24**	-0.38	No change
	(0.08)	(0.10)	(0.22)		(0.22)	(0.10)	(0.27)	
Clothing & shoes	-0.22	0.14	0.38	Decreased CO	0.01	0.34*	0.34	No change
	(0.14)	(0.13)	(0.25)		(0.21)	(0.14)	(0.32)	
Education	-0.18***	-0.53***	-0.39***	Increased CO	-0.55***	-0.41***	0.13	No change
	(0.06)	(0.06)	(0.14)		(0.12)	(0.08)	(0.19)	
Housing	2.16***	0.42	-1.95***	Decreased CI	0.20	-0.24	-0.44	No change
	(0.42)	(0.33)	(0.60)		(0.41)	(0.26)	(0.58)	
Transport	-0.51***	-0.22 *	0.28	Decreased CO	-0.31	-0.81***	-0.52	Increased CO
	(0.13)	(0.13)	(0.25)		(0.23)	(0.18)	(0.42)	
Communication	0.01	-0.30***	-0.27***	Increased CO	-0.30***	-0.35***	-0.07	No change
	(0.01)	(0.03)	(0.07)		(0.07)	(0.06)	(0.14)	
Health	-0.14**	-0.02	0.13	Decreased CO	0.09	0.19***	0.09	No change
	(0.07)	(0.06)	(0.10)		(0.08)	(0.05)	(0.12)	
Household goods	-0.64***	-0.16	0.47**	Decreased CO	-0.17	0.34**	0.55*	Increased CI
	(0.15)	(0.13)	(0.22)		(0.17)	(0.16)	(0.31)	
Recreation	-0.18**	-0.32***	-0.11	No change	-0.31***	-0.12	0.16	Decreased CO
	(0.07)	(0.08)	(0.13)		(0.11)	(0.07)	(0.14)	
Miscellaneous	0.47	-0.05	-0.48	No change	-0.02	-0.38*	-0.41	No change
	(0.41)	(0.32)	(0.50)		(0.33)	(0.22)	(0.46)	
Sample size	**22,700**	**22,700**	**22,700**		**19,564**	**19,564**	**19,564**	

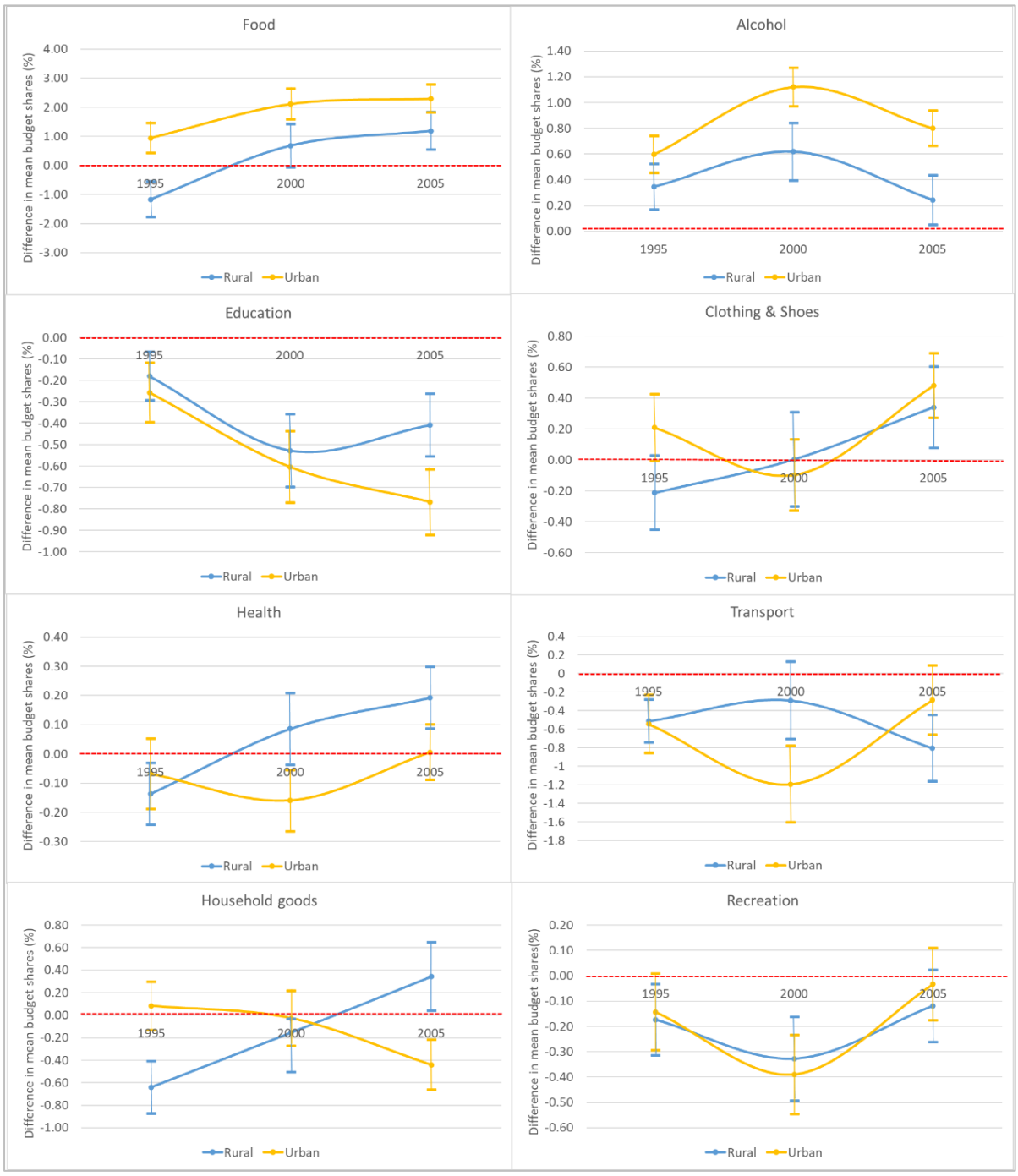
CO – Crowding-out; CI – Crowding-in. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 3.8: Crowding-out transitions from 1995 to 2000 IES and from 2000 to 2005/06 IES using urban households

	1995 to 2000				2000 to 2005-06			
	Baseline difference	**Post-intervention difference**	**Diff-in-diff**	**Impact on crowding-out/in**	**Baseline difference**	**Post-intervention difference**	**Diff-in-diff**	**Impact on crowding-out/in**
Food	0.98***	2.08***	1.27**	Increased CI	2.23***	2.30***	0.21	No Change
	(0.28)	(0.23)	(0.50)		(0.39)	(0.25)	(0.54)	
Tobacco	2.98***	10.55 ***	7.11*	Increased CI	10.15**	3.89**	-6.23	Decreased CI
	(0.42)	(2.04)	(4.24)		(4.15)	(2.00)	(4.28)	
Alcohol	0.59***	1.12***	0.55***	Increased CI	1.12***	0.80***	-0.32*	Decreased CI
	(0.07)	(0.06)	(0.15)		(0.14)	(0.07)	(0.19)	
Clothing & shoes	0.18	-0.16*	-0.34	Switched: CI to CO	-0.11	0.48***	0.59**	Increased CI
	(0.13)	(0.10)	(0.21)		(0.15)	(0.11)	(0.24)	
Education	-0.29***	-0.57***	-0.32**	Increased CO	-0.58***	-0.77***	-0.21	No Change
	(0.07)	(0.06)	(0.14)		(0.12)	(0.08)	(0.20)	
Housing	-0.15	-0.47*	-0.32	No Change	-0.42	-0.70***	-0.36	No Change
	(0.35)	(0.26)	(0.58)		(0.43)	(0.26)	(0.63)	
Transport	-0.58***	-1.26***	-0.75**	Increased CO	-1.23***	-0.29	0.94*	Decreased CO
	(0.18)	(0.14)	(0.33)		(0.28)	(0.19)	(0.52)	
Communication	0.04**	-0.35***	-0.38***	Increased CO	-0.33***	-0.20***	0.10	No Change
	(0.02)	(0.04)	(0.09)		(0.09)	(0.06)	(0.15)	
Health	-0.09	-0.13**	-0.07	No Change	-0.19***	0.01	0.21*	Decreased CO
	(0.10)	(0.05)	(0.12)		(0.07)	(0.05)	(0.11)	
Household goods	0.11	-0.03	-0.18	No Change	-0.07	-0.44***	-0.37	Increased CO
	(0.14)	(0.10)	(0.20)		(0.14)	(0.11)	(0.26)	
Recreation	-0.11	-0.39***	-0.31*	Increased CO	-0.41***	-0.03**	0.37**	Decreased CO
	(0.07)	(0.07)	(0.16)		(0.13)	(0.07)	(0.17)	
Miscellaneous	-0.68**	0.16	0.87*	Decreased CO	-0.02	-1.16***	-1.16***	Increased CO
	(0.35)	(0.22)	(0.46)		(0.30)	(0.19)	(0.44)	
Sample size	**32,077**	**32,077**	**32,077**		**27,768**	**27,768**	**27,768**	

CO – Crowding-out; CI – Crowding-in. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Figure 3.4: Mean budget share differences for selected categories disaggregated by settlement type



Note: The bands indicate 95% confidence intervals.

In Figure 3.4, the differences in mean budget shares for food, alcohol, education, clothing and shoes, transport, health, household goods and recreation are compared over time and disaggregated by urban and rural settlements. Compared to other categories, these categories were all crowded-out at the baseline (1995), and their transitions thus merit closer examination.

As noted earlier, food expenditure transitioned from being crowded-out in 1995 to being crowded-in from 2000 to 2005. The magnitude of these transitions was consistently higher among urban households, compared to rural ones. Although alcohol expenditure also experienced crowding-in, the picture is quite different. While the level of crowding-in generally trended upwards for urban households, the opposite was true for rural households, where the level of crowding-in seemed to trend downwards. In both cases, the level of crowding-in was highest in 2000.

The graph for education showed a worsening scenario of crowding-out for both rural and urban households between 1995 and 2000. Further to this, although households in both areas experienced deeper levels of crowding-out between 1995 and 2005, rural households improved marginally in the latter half of that period. Expenditure on clothing and shoes moved from crowding-out at baseline in 1995 to crowding-in by 2005, particularly for rural households. Urban households in this category experienced wider fluctuations over time, crowded-in at the baseline, then crowded-out in 2000 and crowded-in again in 2005.

Expenditures on health, household goods, transport, and recreation all had a mixture of trends between 1995 and 2005. Health and household goods showed a transition from crowding-out to crowding-in among rural households, while transport and recreation expenditures generally fluctuated in the region below the red dashed line on the graphs, indicating crowding-out.

Overall, these transitions show very eclectic trends across different expenditure types. These geographic disparities likely had different implications for urban and rural households. For instance, urban households that maintained tobacco expenditure after tax increases may have done so at the expense of education expenditure more than their counterparts in rural areas.

3.6.3 Analysis disaggregated by income terciles

In an effort to further tease out the heterogenous nature of these transitions over time, an analysis disaggregated by income level is often interesting, particularly in the context of South Africa, where income inequality has been a perennial challenge (Francis & Webster, 2019, Leibbrandt, Finn & Woolard, 2012). In Figure 3.5, results for households crowding-out/in are disaggregated by income tercile, where Tercile 1 is the lowest income group and Tercile 3 the highest.

In 1995, households in the lowest income group experienced crowding-out on food, health, recreation, and education, although the last category was only significant at the 10% level. The middle-income group, in comparison, registered crowding-out on transport, household goods,

and recreation, with the last category also only significant at the 10% level. Lastly, crowding-out was detected on education, housing, and transport in the wealthiest third of the population in that year. Comparing this with the rural-urban disaggregation presented earlier for 1995 suggests that the crowding-out observed on food and health in both rural and urban areas was likely driven by expenditure crowding-out among the lowest income group (Tercile 1). On the other hand, crowding out on education, again in both areas, was primarily driven by the wealthiest and the poorest thirds of the income gradient.

Education expenditure crowding-out persisted through 2000 and 2005 across all three income groups, while alcohol expenditure was also consistently crowded-in. In comparison, categories such as health, transport, housing, recreation, and household goods all experienced switching between crowding-out and crowding-in.

In Terciles 1 and 2, the two lowest income groups, there was a transition from crowding-out to crowding-in for food expenditure, starting from the year 2000. On the other hand, wealthier households consistently allocated a higher share of their budget to food throughout the entire period. A detailed examination of the transitions, presented in Table B7 in the annex, indicates that the shift to crowding-in was driven by an increase in the budget allocation for food among households in Terciles 1 and 2 during both 2000 and 2005. This peculiar observation suggests a potential enhancement in household food security for urban residents with lower incomes over this period.

Education expenditure also showed some interesting trends. As noted, education crowding-out generally worsened throughout the period from 1995 to 2005. While this was mainly detected among poor (T1) or rich (T3) households in 1995, in 2000 households in all income groups were affected, with higher displacements among the rural poor compared to urban ones. In 2005, the urban poor were more adversely affected by crowding-out than poor rural households.

In the middle-income group, there was no change in crowding-out on transport, household goods, or recreational expenditures. However, some additional categories were crowded-out in 2000. For instance, education, housing, clothing and shoes, and communication were all crowded-out in the 2000 survey, where the last two categories were significant at the 10% level. In the wealthiest group, crowding out decreased for housing, but increased for transport and education. Additional categories in the 2000 survey such as health and recreation also recorded crowding-out.

Figure 3.5: Mean budget share differences for selected categories disaggregated by income tercile

Note: Values above the red dashed line are indicative of crowding-in while those below are indicative of crowding-out.

An examination of alcohol expenditure revealed that crowding-in was most prominent among affluent households in both urban and rural areas in 1995. However, by 2000, the urban middle-income group made the most significant contribution to the observed crowding-in. Although all other subgroups also saw increases from their 1995 baseline shares, the rural middle-income

group was the exception, with no significant change. In 2005, the urban poor and middle-income groups made the most substantial contributions to the observed crowding-in.

3.7 Discussion

An examination of the overall results from the analysis of household budget shares suggests that households incurring tobacco expenditure adjusted their spending patterns across various expenditure categories in a non-uniform manner. This non-linearity is particularly evident in the temporal shifts of expenditure allocations: in some periods, an expenditure category may have been crowded out by tobacco spending, whereas in subsequent periods, the same category may have experienced crowding-in or remained unaffected. This variability likely reflects the inherent volatility of the underlying data and highlights the complexity of household decision-making processes in response to changes in tobacco prices. However, several key findings remained consistent across the analysis period.

First, the expenditures on education, transport, and health consistently experienced crowding-out effects throughout the period under study. Interestingly, while both health and transport expenditures exhibited a decline in the degree of crowding-out over time, the opposite trend was observed for education, which experienced an increased crowding-out effect. A more granular examination reveals that the reduction in the crowding-out effect on education was more pronounced among lower-income households, compared to their wealthier counterparts. This disparity suggests that the impact of tobacco consumption on educational expenditure is not homogenous, with lower-income households potentially adjusting their budgets differently in response to price increases.

The persistent crowding-out effect on education, particularly in urban households, warrants closer attention, especially in the context of rising educational costs relative to other goods and services. Given that education is generally considered a price-inelastic good (Gallet, 2007). the expectation would be for households to allocate a larger share of their budget to education in response to rising prices. The observed reduction in the share of expenditure allocated to education, despite rising costs, suggests a troubling trend: as households increasingly allocate funds to tobacco, children's access to education may be adversely affected. Moreover, this trend is further compounded by the broader socio-economic context of increasing income inequality, despite an overall rise in household incomes.

The second significant finding pertains to the crowding-in effect observed in alcohol expenditure. Households that allocated a portion of their budget to tobacco consistently increased their spending on alcohol. Specifically, between 1995 and 2000, households that purchased tobacco allocated 0.46 percentage points more of their budget to alcohol than non-tobacco-consuming households. Although this gap narrowed slightly to 0.38 percentage points between 2000 and 2005/6, the positive relationship between tobacco consumption and alcohol expenditure remained statistically significant. This suggests a persistent cross-product consumption effect, whereby increased tobacco expenditure is associated with greater alcohol consumption.

Regional variations were also observed, with urban households demonstrating a more pronounced crowding-in effect compared to their rural counterparts. This divergence could be attributed to differing income dynamics and consumption patterns between urban and rural households. The price trends for alcoholic beverages, as reflected in the Consumer Price Index (CPI) data, suggest that the observed crowding-in effect was not solely driven by price changes in the alcohol sector. The CPI values for alcoholic drinks generally followed the broader CPI trend, indicating that other factors—such as income growth—also played a role in shaping the relationship between tobacco and alcohol expenditure.

From a policy perspective, the findings of this study have several important implications. First, the analysis provides a nuanced understanding of the effects of tobacco tax increases on households that continue to smoke despite higher tobacco prices. The heterogeneity in the effects of tobacco expenditure on other budget categories underscores the need for targeted interventions. While there is some evidence of equity gains, particularly in the reduction of crowding-out effects on education among low-income households, the negative effects are also pronounced. Households that were unable to reduce their tobacco consumption experienced substantial decreases in expenditure on essential categories, such as transport and health. This highlights the need for complementary tobacco-control interventions, such as smoking cessation programmes, which could alleviate the adverse effects on these households.

Second, the consistent crowding-in effect observed for alcohol expenditure suggests that tobacco-control policies may have unintended consequences for alcohol consumption. The positive association between tobacco and alcohol spending implies that the two products may act as complements in household budgets. Policymakers should be mindful of these cross-

product effects when designing fiscal policies aimed at reducing the consumption of harmful goods. A more integrated approach to the regulation of tobacco and alcohol—rather than addressing each product in isolation—could lead to more effective public health outcomes.

To fully contextualise the effects of price and income elasticity on household spending behaviour, it is essential to consider these dynamics within the broader economic environment. To this end, data from Statistics South Africa were used to analyse trends in the Consumer Price Index (CPI) for selected expenditure categories from 1995 to 2006 (Statistics South Africa, 2001, Statistics South Africa, 2021), with particular focus on those categories identified as being influenced by tobacco expenditure. The analysis draws on both nominal CPI (Panel A) and real CPI (Panel B) data. Nominal CPI reflects the combined effect of category-specific price changes and inflation, whereas real CPI excludes the effects of inflation, allowing for a clearer understanding of price trends in real terms. The year 1994 serves as the base year, with an index value set to 100. Values below 100 indicate a decrease in prices compared to 1994, while values above 100 reflect an increase.

In Panel A, the overall CPI demonstrates an upward price trend for all product categories for the period considered. Nominal CPI values for the different categories are not all that helpful, because the nominal prices have been increasing for all product categories. However, the real price indices are much more interesting, because they indicate what has happened to relative prices. As a result, in the subsequent discussion, I will only refer to the real CPI values.

Between 1995 and 2005, South Africa experienced moderate but uneven economic growth, with the per capita GDP growth rate rising from 1.5% annually (1995-2000) to 2.9% (2000-2005) (The World Bank Group, 2023). This growth reflected an overall increase in average household incomes, but as Leibbrandt & Levinsohn (2014) highlight, the benefits of this income growth were not equally distributed across income groups. As higher-income groups saw larger income gains, their spending behaviour, particularly on luxury goods and services, may have been more pronounced compared to lower-income households.

This broader economic backdrop is critical in understanding the spending patterns observed in the results, particularly the crowding-out and crowding-in effects. For instance, the growth in household income should theoretically have led to increased expenditure on both essential and non-essential goods. However, the observed data suggests that increases in income were not

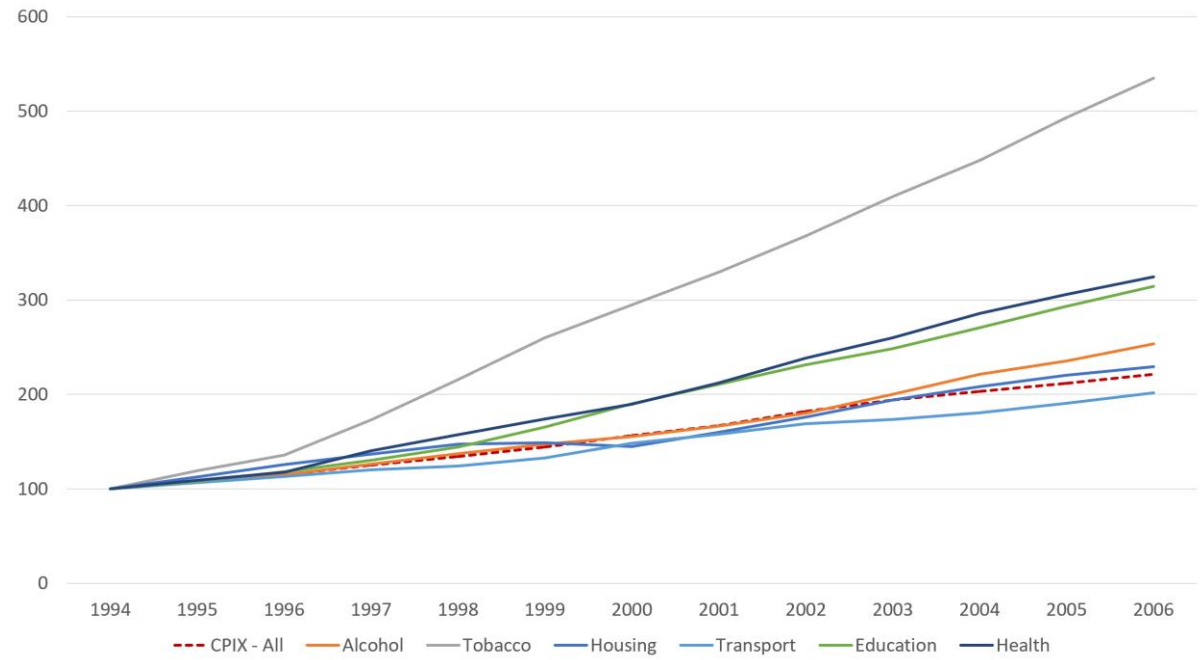
always accompanied by an increase in the budget share allocated to necessities like education and transport. This can be partially explained by the fact that these goods are necessities, with income elasticity less than 1. As income grows, absolute expenditure on necessities increases, but the budget share tends to decrease as higher-income households spend more on other categories, such as luxuries.

This effect is evident in the crowding-out observed for education, health, and transport. Despite rising incomes, these categories experienced a reduction in their share of total household expenditure, particularly among lower-income households. This pattern aligns with the economic context, where rising incomes were not equitably distributed. The increasing cost of education, for example, outpaced general inflation (Calitz & Fourie, 2016), putting pressure on household budgets and contributing to the observed crowding-out. Education faced increased crowding-out in urban areas, where education costs rose disproportionately relative to income, thus limiting the capacity of households to allocate sufficient resources towards education, even as overall incomes grew.

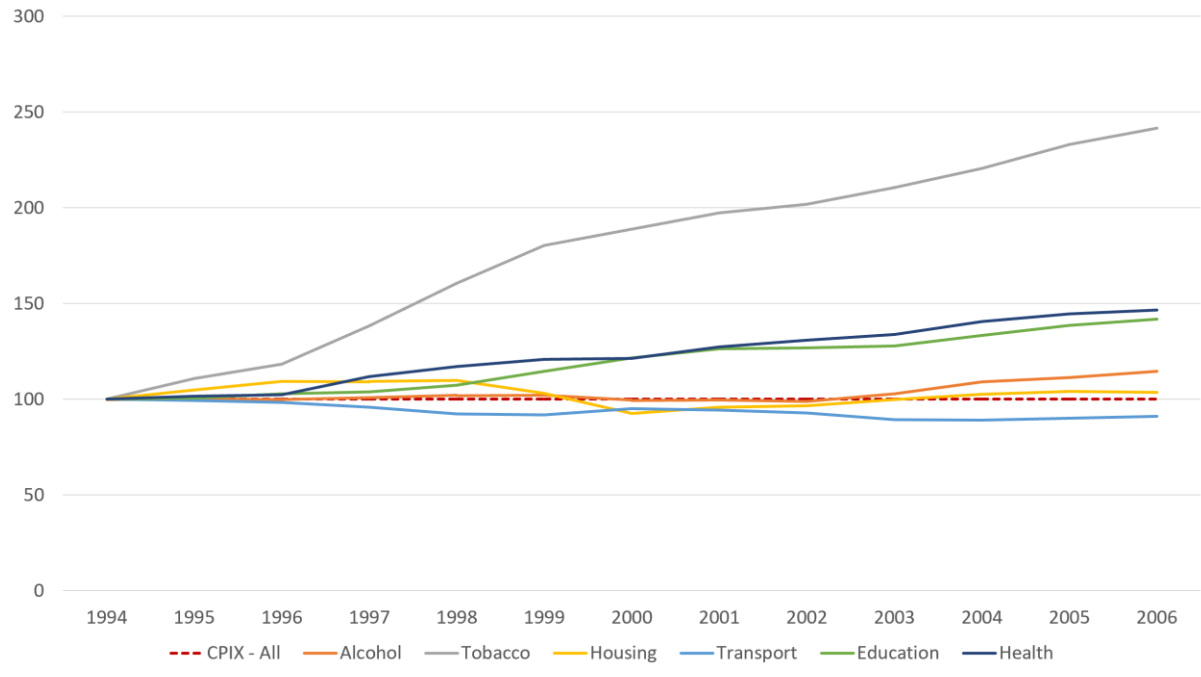
On the other hand, alcohol expenditure exhibited a crowding-in effect, with households that purchased tobacco allocating a larger share of their budget to alcohol over the same period. This is particularly significant considering that alcohol prices, as shown by the CPI, remained relatively stable compared to the price increases in tobacco, which were driven by government excise tax hikes and industry-led price adjustments. The crowding-in effect between tobacco and alcohol expenditure suggests that, as household incomes rose, there was an increase in discretionary spending on goods like alcohol, despite limited price changes in the alcohol sector. This aligns with the income elasticity concept, where luxury goods (such as alcohol) see their budget share increase more as income grows, which is consistent with the positive relationship between tobacco and alcohol expenditure observed in the results.

Figure 3.6: Consumer Price Indices for selected categories in South Africa (1994-2006)

Panel A: Nominal CPI values with 1994 as reference year



Panel B: Real CPI values with 1994 as reference year



Note: Data for Consumer Price Indices (CPI) extracted from Statistics South Africa's P0141 monthly releases over the period and annualised. CPIX refers to the CPI excluding interest rates on mortgage bonds. A complete series of CPI values were available for Metropolitan areas only.

Furthermore, the relative stability in the price of transport and housing during this period—reflected in the CPI trends—suggests that these categories were less influenced by price changes, which may explain why they were more consistently crowded-out by tobacco expenditure. Since these goods are more price-inelastic, households would be less responsive to price fluctuations and more likely to adjust their budget allocations in favour of tobacco, a relatively more price-sensitive category.

Macroeconomic factors also played a role. The fluctuations in household expenditure patterns were partly driven by the broader economic environment. The weak but positive GDP growth during the period resulted in an overall rise in incomes, but the uneven distribution of these gains had a pronounced effect on consumption patterns. The fact that tobacco expenditure persisted even amid increasing prices suggests that, for certain households, the income effect outweighed the price sensitivity. The crowding-in of alcohol expenditure, in particular, is consistent with the notion that, as incomes rose, so did the demand for other vices or luxury goods, with tobacco serving as a potential gateway for increased alcohol consumption.

Policy implications based on these findings underscore the need for a nuanced understanding of how economic factors—such as income inequality, income growth, and price changes—affect household budgets. First, the persistent crowding-out effects in essential sectors like education and health highlight a potential equity concern. Households with lower incomes, who may already be struggling with higher education costs, could be adversely affected by rising tobacco expenditures. This reinforces the need for targeted interventions, such as tobacco cessation programmes, to mitigate the negative impact on household spending in these crucial areas.

Second, the positive relationship between tobacco and alcohol expenditures suggests that tobacco control policies could have cross-product effects on alcohol consumption. Policymakers should be cognisant of these links, as aligning tobacco and alcohol control efforts might lead to more comprehensive public health benefits. For example, increased tobacco taxes might not only reduce smoking rates but also moderate alcohol consumption, highlighting the importance of a holistic approach to the regulation of harmful commodities.

3.8 Summary and conclusion

This chapter critically examined the dynamics of crowding-out and crowding-in effects in the context of rising tobacco taxes and retail prices in South Africa between 1995 and 2005. Using a DiD methodology, this study sought to elucidate how households' expenditure patterns evolved in response to changes in tobacco prices, considering both temporal (1995–2000 and 2000–2005/06) and cross-sectional variations. Acknowledging the critical role of the parallel trends assumption in the DiD framework, the study employed Abadie's semi-parametric approach (2005) to address potential violations of this assumption. Specifically, this method used a direct weighting scheme to balance covariate distributions between the control and treatment groups, thus mitigating confounding biases in the estimated treatment effect. A range of diagnostic tests were conducted to evaluate the validity of the parallel trends assumption, providing further justification for the adoption of the Abadie method.

The empirical results reveal a nuanced pattern of expenditure adjustments among tobacco-consuming households, with significant heterogeneity across different categories of household spending. In some categories, such as food, crowding-in effects were observed, with the relative share of household expenditure allocated to food increasing following the tobacco price hikes. In contrast, expenditure on education showed pronounced crowding-out effects, with a reduction in the share of spending directed towards education. These findings highlight that the impact of rising tobacco taxes on household budget allocations is not uniform across all consumption categories, indicating that the elasticity of expenditure in response to tobacco price changes varies depending on the type of expenditure.

Furthermore, the study identifies potential equity considerations. The analysis suggests that the negative impact of tobacco taxation on educational expenditure was less severe among lower-income households, implying that these households were less affected by the crowding-out of educational spending. This finding suggests that fiscal interventions such as tobacco taxes may produce differential effects across income groups, with the potential to mitigate some socio-economic disparities, although these effects remain context-specific and require further investigation.

Despite its contributions, the study has several limitations. The econometric models used in this analysis did not incorporate fixed effects, which could have controlled for unobserved

heterogeneity and potential clustering within household data (Bertrand, Duflo, & Mullainathan, 2004). The absence of fixed effects was due to the lack of consistent variables in the available datasets, which constrained the ability to fully account for these sources of variation. Moreover, while the semi-parametric DiD method is robust in addressing violations of the parallel trends assumption, it relies on the assumption that only observable covariates included in the model influence household expenditure shares. This overlooks the potential role of unobserved factors, a limitation that could be addressed through the use of panel data in future studies, which would better account for time-invariant individual characteristics and their influence on expenditure decisions.

In conclusion, although the study has several methodological limitations, it provides valuable preliminary insights into the interaction between fiscal policy, specifically tobacco taxation, and household expenditure patterns in South Africa. The findings suggest that the economic consequences of tobacco tax policies are complex and vary depending on household characteristics, particularly income level, and the nature of the goods in question. These results contribute to the growing body of literature on the effects of excise taxes on consumption behaviour, with implications for policy design aimed at balancing fiscal objectives with public health goals. Future research, using more granular data and advanced econometric techniques, will be essential to refine these findings and offer a more comprehensive understanding of the long-term socio-economic consequences of tobacco taxation.

3.9 Annex B

Table B 1: Standard difference in difference estimation without covariates – Full sample

	1995 to 2000				2000 to 2005-06			
	Baseline difference	**Post-intervention difference**	**Diff-in-diff**	**Impact on crowding-out/in**	**Baseline difference**	**Post-intervention difference**	**Diff-in-diff**	**Impact on crowding-out/in**
Food	0.00399	-0.022***	-0.0260***	Increased CO	-0.0220***	0.026***	0.0439***	Switched: CO to CI
	(0.00243)	(0.002)	(0.00339)		(0.00276)	(0.002)	(0.00375)	
Alcohol	0.0119***	0.034***	0.0223***	Increased CI	0.0342***	0.005***	-0.0197***	Decreased CI
	(0.000550)	(0.001)	(0.000769)		(0.000614)	(0.001)	(0.000834)	
Clothing & shoes	0.00278***	-0.002**	-0.00495***	Switched: CI to CO	-0.00211**	0.006***	0.00553***	Switched: CO to CI
	(0.000885)	(0.001)	(0.00124)		(0.00101)	(0.001)	(0.00137)	
Education	-0.00185***	-0.008***	-0.00626***	Increased CO	-0.00810***	-0.007***	0.000462	No change
	(0.000536)	(0.001)	(0.000749)		(0.000734)	(0.001)	(0.000999)	
Housing	-0.0112***	0.008***	0.0188***	Switched: CO to CI	0.00738***	-0.007***	-0.0141***	Increased CO
	(0.00227)	(0.002)	(0.00317)		(0.00226)	(0.002)	(0.00307)	
Transport	-0.00148	-0.006***	-0.00410**	Increased CO	-0.00548***	-0.006***	-0.00206	No change
	(0.00115)	(0.001)	(0.00160)		(0.00162)	(0.001)	(0.00221)	
Communication	-0.000226	-0.002***	-0.00171***	Increased CO	-0.00192***	-0.003***	-0.000675	No change
	(0.000307)	(0.00)	(0.000428)		(0.000519)	(0.000)	(0.000706)	
Health	0.000972**	-0.001*	-0.00171***	Switched: CI to CO	-0.000728*	0.001	0.00101*	Decreased CO
	(0.000443)	(0.00)	(0.000619)		(0.000424)	(0.000)	(0.000576)	
Household goods	-0.00265***	0.001	0.00398***	Switched: CO to CI	0.00138	-0.002*	-0.00606***	Increased CO
	(0.000846)	(0.001)	(0.00118)		(0.00106)	(0.001)	(0.00145)	
Recreation	0.000852	-0.001	-0.00149*	Increased CO	-0.000622	-0.001***	0.00120	No change
	(0.000561)	(0.001)	(0.000784)		(0.000641)	(0.001)	(0.000871)	
Miscellaneous	-0.00304	-0.002	0.00115	No change	-0.00197	-0.011***	-0.00953***	Increased CO
	(0.00207)	(0.002)	(0.00289)		(0.00179)	(0.001)	(0.00243)	
Sample size	**55,847**	**55,847**	**55,847**		**47,336**	**47,336**	**47,336**	

CO – Crowding-out; CI – Crowding-in. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table B 2: Standard difference in difference estimation with covariates – Full sample

	1995 to 2000				2000 to 2005-06			
	Baseline difference	**Post-intervention difference**	**Diff-in-diff**	**Impact on crowding-out/in**	**Baseline difference**	**Post-intervention difference**	**Diff-in-diff**	**Impact on crowding-out/in**
Food	0.0159***	0.003	-0.0131***	Decreased CI	0.00741***	0.033**	0.0257***	Increased CI
	(0.00226)	0.002	(0.00307)		(0.00255)	0.002	(0.00327)	
Alcohol	-0.000654	0.013***	0.0138***	Increased CI	0.0109***	0.002***	-0.00871***	Decreased CI
	(0.000474)	0	(0.000645)		(0.000542)	0	(0.000695)	
Clothing & shoes	0.00133	-0.002*	-0.00289**	Increased CO	0.000721	0.004***	0.00283**	Increased CI
	(0.000474)	0.001	(0.00121)		(0.00104)	0.001	(0.00134)	
Education	-0.00125**	-0.007***	-0.00549***	Increased CO	-0.00639***	-0.006***	-1.78e-05	No change
	(0.000545)	0.001	(0.000741)		(0.000773)	0.001	(0.000992)	
Housing	-0.00422*	0.013***	0.0172***	Switched: CO to CI	0.00357	-0.009***	-0.0125***	Switched: CI to CO
	(0.00229)	0.002	(0.00312)		(0.00229)	0.002	(0.00294)	
Transport	-0.00399***	-0.010***	-0.00651***	Increased CO	-0.00925***	-0.008***	0.00151	No change
	(0.00110)	0.001	(0.00150)		(0.00169)	0.002	(0.00216)	
Communication	-0.000323	-0.003***	-0.00245***	Increased CO	-0.00292***	-0.003***	-0.000169	No change
	(0.000307)	0	(0.000417)		(0.000542)	0	(0.000695)	
Health	0.000538	-0.002***	-0.00229***	Switched: CI to CO	-0.000192	0.001	0.000804	No change
	(0.000445)	0	(0.000605)		(0.000449)	0	(0.000576)	
Household goods	-0.00341***	0	0.00350***	Decreased CO	0.00162	-0.004***	-0.00605***	Increased CO
	(0.000866)	0.001	(0.00118)		(0.00113)	0.001	(0.00144)	
Recreation	-0.00107*	-0.004***	-0.00258***	Increased CO	-0.00319***	0	0.00296***	Decreased CO
	(0.000566)	0.001	(0.000769)		(0.000667)	0.001	(0.000856)	
Miscellaneous	-0.00286	-0.002***	0.000839	No change	-0.00231	-0.009***	-0.00633***	Increased CO
	(0.00206)	0.002	(0.00280)		(0.00185)	0.002	(0.00238)	
Sample size	**55,787**	**55,787**	**55,787**		**47,336**	**47,336**	**47,336**	

CO – Crowding-out; CI – Crowding-in. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table B 3: Abadie's Difference in difference estimation without covariates – Full sample

	1995 to 2000				2000 to 2005-06			
	Baseline difference	**Post-intervention difference**	**Diff-in-diff**	**Impact on crowding-out/in**	**Baseline difference**	**Post-intervention difference**	**Diff-in-diff**	**Impact on crowding-out/in**
Food	0.0142***	0.009***	-0.00540*	Decreased CI	0.0176***	0.026***	0.00823**	Increased CI
	(0.00233)	0.002	(0.00319)		(0.00258)	0.002	(0.00346)	
Alcohol	0.00147**	0.017***	0.0153***	Increased CI	0.0136***	0.005***	-0.00846***	Decreased CI
	(0.000678)	0.001	(0.000930)		(0.000767)	0.001	(0.00103)	
Clothing & shoes	0.00278***	-0.001	-0.00392***	Decreased CI	-0.000714	0.006***	0.00654***	Switched: CO to CI
	(0.000851)	0.001	(0.00117)		(0.000967)	0.001	(0.00130)	
Education	-0.00183***	-0.006***	-0.00448***	Increased CO	-0.00601***	-0.007***	-0.000566	No change
	(0.000488)	0	(0.000669)		(0.000638)	0.001	(0.000856)	
Housing	0.00825***	0	-0.00875***	Decreased CI	-0.00339	-0.006***	-0.00364	No change
	(0.00223)	0.002	(0.00306)		(0.00220)	0.001	(0.00295)	
Transport	-0.0101***	-0.009***	0.000660	No change	-0.0102***	-0.003***	0.00383*	Decreased CO
	(0.00115)	0.001	(0.00158)		(0.00160)	0	(0.00214)	
Communication	-0.000380	-0.003***	-0.00281***	Increased CO	-0.00353***	0.001	0.000717	No change
	(0.000296)	0	(0.000407)		(0.000506)	0	(0.000678)	
Health	-0.00249***	-0.001**	0.00160***	Reduced CO	-0.00105**	0.001	0.00163***	Decreased CO
	(0.000441)	0	(0.000604)		(0.000412)	0	(0.000552)	
Household goods	-0.00190**	0	0.00161	No change	-0.000980	-0.002*	-0.000780	No change
	(0.000821)	0.001	(0.00113)		(0.00103)	0.001	(0.00138)	
Recreation	-0.00317***	-0.003***	-0.000263	No change	-0.00388***	-0.001***	0.00241***	Decreased CO
	(0.000555)	0.001	(0.000761)		(0.000608)	0.001	(0.000816)	
Miscellaneous	-0.00676***	0	0.00648**	Reduced CO	-0.00157	-0.011***	-0.00992***	Increased CO
	(0.00199)	0.002	(0.00273)		(0.00167)	0.001	(0.00224)	
Sample size	**55,775**	**55,775**	**55,775**		**47,336**	**47,336**	**47,336**	

CO – Crowding-out; CI – Crowding-in. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table B 4: Crowding-out transitions from 1995 to 2000 IES and from 2000 to 2005/06 IES using Tercile 1 households

	1995 to 2000				2000 to 2005-06			
	Baseline difference	**Post-intervention difference**	**Diff-in-diff**	**Impact on crowding-out/in**	**Baseline difference**	**Post-intervention difference**	**Diff-in-diff**	**Impact on crowding-out/in**
Food	-1.47***	0.27	1.74***	Decrease CO	0.70*	1.45***	0.76	No change
	(0.35)	(0.34)	(0.49)		(0.40)	(0.34)	(0.52)	
Alcohol	0.18*	0.81***	0.63***	Increased CI	0.55***	0.17	-0.37**	Decreased CI
	(0.11)	(0.10)	(0.15)		(0.13)	(0.11)	(0.16)	
Clothing & shoes	0.01	-0.04	-0.06	No change	-0.12	0.09	0.21	No change
	(0.13)	(0.13)	(0.19)		(0.18)	(0.15)	(0.23)	
Education	-0.10*	-0.49***	-0.39***	Increased CO	-0.45***	-0.41***	0.033	No change
	(0.05)	(0.05)	(0.08)		(0.09)	(0.08)	(0.12)	
Housing	1.00***	0.14	-0.86*	Decrease CI	-0.12	0.25	0.37	No change
	(0.32)	(0.31)	(0.45)		(0.34)	(0.29)	(0.45)	
Transport	-0.10	-0.26***	-0.16	No change	-0.31	-0.46***	-0.15	No change
	(0.10)	(0.09)	(0.13)		(0.20)	(0.17)	(0.27)	
Communication	-0.01	-0.20***	-0.20***	Increased CO	-0.22***	-0.48 ***	-0.26**	Increased CO
	(0.03)	(0.03)	(0.04)		(0.08)	(0.070	(0.11)	
Health	-0.12***	-0.09*	0.04	No change	-0.03	0.13	0.16*	Decreased CO
	(0.05)	(0.05)	(0.06)		(0.06)	(0.05)	(0.08)	
Household goods	0.04	-0.07	-0.11	No change	-0.07	-0.45***	-0.37*	Increased CO
	(0.09)	(0.09)	(0.13)		(0.17)	(0.14)	(0.22)	
Recreation	-0.19**	-0.35***	-0.17	No change	-0.34***	0.17	0.51***	Decreased CO
	(0.10)	(0.09)	(0.13)		(0.10)	(0.09)	(0.13)	
Miscellaneous	0.74***	0.27	-0.47	No change	0.42*	-0.46**	-0.88***	Switched: CI to CO
	(0.23)	(0.23)	(0.32)		(0.23)	(0.19)	(0.30)	
Sample size	**19,729**	**19,729**	**19,729**		**17,492**	**17,492**	**17,492**	

CO – Crowding-out; CI – Crowding-in. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table B 5: Crowding-out transitions from 1995 to 2000 IES and from 2000 to 2005/06 IES using Tercile 2 households

	1995 to 2000				2000 to 2005-06			
	Baseline difference	**Post-intervention difference**	**Diff-in-diff**	**Impact on crowding-out/in**	**Baseline difference**	**Post-intervention difference**	**Diff-in-diff**	**Impact on crowding-out/in**
Food	-0.07	0.33	0.40	No change	0.54	2.08***	1.54***	Increased CI
	(0.34)	(0.31)	(0.46)		(0.35)	(0.31)	(0.46)	
Alcohol	0.35***	0.94***	0.59***	Increased CI	0.95***	0.70***	-0.25*	Decreased CI
	(0.11)	(0.10)	(0.15)		(0.11)	(0.10)	(0.15)	
Clothing & shoes	0.25	-0.32**	-0.57***	Increased CO	-0.28*	0.40***	0.67***	Switched: CO to CI
	(0.16)	(0.15)	(0.22)		(0.17)	(0.15)	(0.22)	
Education	-0.12	-0.32***	-0.20*	Increased CO	-0.31***	-0.47***	-0.16	No change
	(0.08)	(0.07)	(0.10)		(0.09)	(0.08)	(0.12)	
Housing	0.38	-0.18	-0.57	No change	-0.56*	-0.84***	-0.28	No change
	(0.33)	(0.31)	(0.45)		(0.32)	(0.29)	(0.43)	
Transport	-0.33**	-0.46***	-0.13	No change	-0.47**	-1.10***	-0.63**	Increased CO
	(0.14)	(0.13)	(0.19)		(0.23)	(0.20)	(0.30)	
Communication	-0.02	-0.29***	-0.27***	Increased CO	-0.23***	-0.24***	-0.00	No change
	(0.05)	(0.04)	(0.07)		(0.08)	(0.07)	(0.11)	
Health	0.09	0.01	-0.08	No change	0.10	0.09	-0.00	No change
	(0.07)	(0.07)	(0.10)		(0.07)	(0.06)	(0.09)	
Household goods	-0.33**	0.10	0.44**	Decrease in CO	0.16	-0.10	-0.26	No change
	(0.15)	(0.14)	(0.21)		(0.19)	(0.17)	(0.25)	
Recreation	-0.13*	-0.03	0.10	No change	-0.01	-0.09	-0.09	No change
	(0.08)	(0.07)	(0.11)		(0.09)	(0.08)	(0.12)	
Miscellaneous	-0.07	0.20	0.28	No change	0.11	-0.44*	-0.55	No change
	(0.34)	(0.31)	(0.46)		(0.28)	(0.25)	(0.37)	
Sample size	**18,825**	**18,825**	**18,825**		**16,194**	**16,194**	**16,194**	

CO – Crowding-out; CI – Crowding-in. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table B 6: Crowding-out transitions from 1995 to 2000 IES and from 2000 to 2005/06 IES using Tercile 3 households

	1995 to 2000				2000 to 2005-06			
	Baseline difference	**Post-intervention difference**	**Diff-in-diff**	**Impact on crowding-out/in**	**Baseline difference**	**Post-intervention difference**	**Diff-in-diff**	**Impact on crowding-out/in**
Food	1.25***	2.05***	0.79*	Increased CI	2.08***	1.33***	-0.75*	Decreased CI
	(0.30)	(0.28)	(0.41)		(0.31)	(0.29)	(0.43)	
Alcohol	0.52***	0.82***	0.30***	Increased CI	0.82***	0.52***	-0.29**	Decreased CI
	(0.08)	(0.07)	(0.11)		(0.09)	(0.08)	(0.12)	
Clothing & shoes	0.01	0.20	0.20	No change	0.28**	0.52***	0.24	No change
	(0.13)	(0.12)	(0.18)		(0.14)	(0.13)	(0.19)	
Education	-0.39***	-0.75***	-0.36**	Increased CO	-0.79***	-0.98***	-0.19	No change
	(0.11)	(0.10)	(0.15)		(0.14)	(0.13)	(0.19)	
Housing	-1.61***	-0.05	1.56***	Decreased CO	-0.11	-0.92**	-0.81	No change
	(0.41)	(0.38)	(0.55)		(0.42)	(0.40)	(0.58)	
Transport	-0.74***	-1.58***	-0.83**	Increased CO	-1.26***	0.41	1.67***	Decreased CO
	(0.26)	(0.24)	(0.36)		(0.34)	(0.33)	(0.47)	
Communication	-0.01	-0.34***	-0.33***	Increased CO	-0.29***	0.03	0.32**	Decreased CO
	(0.06)	(0.06)	(0.08)		(0.09)	(0.09)	(0.13)	
Health	0.13	-0.23**	-0.35***	Increased CO	-0.26***	-0.05	0.21*	Decreased CO
	(0.09)	(0.09)	(0.13)		(0.09)	(0.08)	(0.11)	
Household goods	0.06	0.04	-0.02	No change	-0.02	-0.08	-0.06	No change
	(0.16)	(0.15)	(0.21)		(0.17)	(0.16)	(0.24)	
Recreation	-0.03	-0.56***	-0.53***	Increased CO	-0.55***	-0.13	0.42***	Decreased CO
	(0.11)	(0.10)	(0.15)		(0.12)	(0.11)	(0.16)	
Miscellaneous	0.83**	0.40	-0.43	No change	0.10	-0.65**	-0.75*	Increased CO
	(0.37)	(0.34)	(0.50)		(0.30)	(0.29)	(0.42)	
Sample size	**17,221**	**17,221**	**17,221**		**13,650**	**13,650**	**13,650**	

CO – Crowding-out; CI – Crowding-in. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table B 7: Crowding-out transitions from 1995 to 2000 IES and from 2000 to 2005/06 IES disaggregated by rural-urban and income category

	1995						2000						2005					
	Poor		Middle		Rich		Poor		Middle		Rich		Poor		Middle		Rich	
	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural
food	-0.21	**-1.88*** **	0.41	**-1.22** **	1.32***	0.51	1.57**	0.01	0.88**	-0.18	1.98***	2.56***	1.10**	1.80***	3.13***	0.29	1.41***	0.54
	(0.66)	(0.40)	(0.44)	(0.52)	(0.33)	(0.74)	(0.62)	(0.52)	(0.44)	(0.58)	(0.33)	(0.86)	(0.50)	(0.45)	(0.40)	(0.49)	(0.32)	(0.76)
tob	4.63	3.61***	4.31** *	3.54***	2.48***	2.22***	29.12** *	7.15***	6.92** *	5.55***	4.79***	5.43***	6.39	4.52***	4.44***	3.14***	2.92***	2.80***
	(12.83)	(0.18)	(0.15)	(0.13)	(0.07)	(0.21)	(9.96)	(0.27)	(0.15)	(0.15)	(0.08)	(0.26)	(8.13)	(0.24)	(0.14)	(0.13)	(0.08)	(0.23)
alc	-0.04	0.28**	0.29**	0.42**	0.55***	0.35**	0.54***	0.58***	1.27** *	0.28	0.81***	0.79***	0.39**	-0.05	0.84***	0.45	0.55***	0.32*
	(0.19)	(0.13)	(0.14)	(0.17)	(0.09)	(0.16)	(0.19)	(0.17)	(0.13)	(0.19)	(0.10)	(0.22)	(0.16)	(0.15)	(0.12)	(0.16)	(0.09)	(0.19)
clths	0.50**	-0.23	0.29	0.20	0.15	**-0.74** **	-0.41	0.12	-0.52**	0.12	0.36**	-0.39	0.31	-0.24	0.13	0.71***	0.49***	0.95***
	(0.25)	(0.16)	(0.22)	(0.22)	(0.15)	(0.33)	(0.28)	(0.22)	(0.22)	(0.25)	(0.15)	(0.38)	(0.23)	(0.20)	(0.20)	(0.21)	(0.15)	(0.34)
edu	**-0.17***	-0.08	-0.16	-0.10	**0.40*****	**-0.41***	**-0.41****	**0.49*****	**0.28****	**0.40*****	**0.80*****	**-1.10****	-0.55***	-0.31***	-0.50***	-0.41***	-1.05***	-0.05
	(0.11)	(0.06)	(0.11)	(0.10)	(0.12)	(0.25)	(0.17)	(0.09)	(0.13)	(0.13)	(0.15)	(0.43)	(0.14)	(0.08)	(0.12)	(0.11)	(0.14)	(0.38)
hse	0.21	1.39***	-0.09	1.58***	**1.74*****	-0.88	-0.57	0.27	-0.69	-0.17	-0.22	1.08	0.31	0.33	-0.36	-1.65***	-1.25***	0.92
	(0.61)	(0.36)	(0.44)	(0.49)	(0.45)	(0.90)	(0.56)	(0.41)	(0.43)	(0.47)	(0.47)	(0.96)	(0.46)	(0.36)	(0.39)	(0.40)	(0.45)	(0.85)
trans	-0.06	-0.11	-0.29	**-0.37***	**-0.61****	**-1.54*****	-0.27	-0.32	**-0.69****	-0.01	**-1.38*****	0.28	-0.17	-0.74***	-1.59***	-0.23	-0.76**	-2.08***
	(0.21)	(0.10)	(0.20)	(0.19)	(0.29)	(0.57)	(0.33)	(0.25)	(0.30)	(0.35)	(0.38)	(0.85)	(0.28)	(0.22)	(0.27)	(0.29)	(0.36)	(0.75)
comm	0.01	-0.01	-0.01	-0.015	-0.02	0.05	**-0.23***	**-0.23****	-0.14	**-0.44*****	**-0.32*****	-0.12	-0.55***	-0.43***	-0.21**	-0.32***	-0.02	0.05
	(0.08)	(0.03)	(0.07)	(0.06)	(0.07)	(0.11)	(0.14)	(0.09)	(0.11)	(0.12)	(0.10)	(0.22)	(0.11)	(0.08)	(0.10)	(0.10)	(0.10)	(0.19)
health	-0.06	**-0.15*****	0.06	0.03	0.13	0.12	**-0.16***	0.10	-0.02	0.35***	**-0.22****	**-0.60*****	0.06	0.22***	0.14*	0.07	-0.10	0.34**
	(0.08)	(0.06)	(0.10)	(0.10)	(0.10)	(0.19)	(0.09)	(0.09)	(0.08)	(0.11)	(0.09)	(0.17)	(0.08)	(0.08)	(0.08)	(0.09)	(0.09)	(0.15)
hhgds	0.50** *	-0.16	0.01	**-0.83*****	0.19	**-0.71***	0.16	-0.36	0.15	0.16	0.01	-0.42	-0.69***	-0.19	-0.76***	-1.06***	-0.05	0.11
	(0.18)	(0.11)	(0.20)	(0.24)	(0.18)	(0.41)	(0.25)	(0.22)	(0.23)	(0.33)	(0.18)	(0.54)	(0.21)	(0.19)	(0.21)	(0.28)	(0.17)	(0.48)
rec	-0.14	-0.20*	**-0.25****	-0.01	0.02	-0.18	-0.18	**-0.48*****	-0.01	-0.07	**-0.58*****	-0.26	0.34**	-0.03	-0.09	-0.13	-0.10	-0.29
	(0.20)	(0.11)	(0.11)	(0.11)	(0.12)	(0.20)	(0.17)	(0.13)	(0.12)	(0.13)	(0.13)	(0.22)	(0.14)	(0.11)	(0.11)	(0.11)	(0.13)	(0.20)
misc	-0.54	1.15***	-0.24	0.30	0.40	3.43***	-0.026	0.80***	0.04	0.36	0.34	**-1.82****	-0.56**	-0.35	-0.73**	0.18	-0.67**	-0.82
	(0.37)	(0.30)	(0.41)	(0.55)	(0.39)	(1.01)	(0.34)	(0.30)	(0.36)	(0.43)	(0.33)	(0.85)	(0.28)	(0.27)	(0.33)	(0.36)	(0.31)	(0.75)
	7,547	**12,182**	**11,245**	**7,580**	**14,044**	**3,177**	**7,341**	**10,151**	**9,311**	**6,883**	**11,121**	**2,529**	**7,341**	**10,151**	**9,311**	**6,883**	**11,121**	**2,529**

4. CHAPTER 4: THE EFFECT OF TOBACCO AND ALCOHOL EXPENDITURE ON CHILD HEALTH OUTCOMES

4.1 Introduction

The nutritional status of children is a key determinant of their long-term health and socio-economic outcomes. Adequate nutrition in early childhood is critical for physical growth, immune function, and cognitive development (Dube et al., 2001, Foltran et al., 2011, Ismail & Suffla, 2013). Insufficient nutrition during this formative period can lead to a range of adverse outcomes, including chronic diseases, impaired educational performance, and reduced labour market productivity (Hoddinott et al., 2013, Mensch et al., 2019, Victora et al., 2008). These consequences, in turn, perpetuate cycles of poverty, as poorly nourished children are more likely to experience lower economic mobility as adults (Said-Mohamed et al., 2015).

South Africa's Global Hunger Index (GHI) score,[16] a multidimensional measure of hunger-associated undernutrition, child wasting, child stunting, and child mortality, declined from 14.5 in 2018 to 12.9 in 2021. This suggests the country may be on track to meet its Sustainable Development Goal (SDG) targets on hunger, specifically, the target to end all forms of malnutrition by 2030. However, despite this progress, more still needs to be done to further reduce the prevalence of malnutrition and improve household food security in the country (National Department of Health & ICF, 2019).

The determinants of child malnutrition are multifaceted, and much of the existing research focuses on social determinants of health, including maternal education, sanitation, and low birth weight (Akombi et al., 2017, Keino et al., 2014). A less explored but increasingly recognised factor is household expenditure on tobacco and alcohol, particularly in low-income settings. Tobacco consumption, both by pregnant women and other household members, has been linked to poorer child nutrition due to its displacement of household spending on food (Cohen, 1981). This "crowding-out" effect, whereby spending on tobacco and alcohol reduces available resources for essential goods such as food, may exacerbate malnutrition and

[16] The Global Hunger Index determines hunger on a 100-point scale, where 0 is the best possible score (no hunger) and 100 is the worst. Each country's GHI score is classified by severity, from low (less than 10) to extremely alarming (50 and above).

contribute to adverse child health outcomes (Chelwa & Van Walbeek, 2014, Chelwa & Koch, 2019, Husain et al., 2018, Jin & Cho, 2021, John, 2008, Nguyen & Nguyen, 2020, Nyagwachi, Chelwa & van Walbeek, 2020, Paraje & Araya, 2018, Pu et al., 2008, San & Chaloupka, 2016).

The current study is related to the literature on cigarette-induced crowding-out on household expenditure. However, it goes beyond merely determining the presence of crowding-out by demonstrating that its effect on food expenditure can have adverse effects on child health nutrition. To provide a thorough understanding of the issues at hand, I first examine the relevant literature on this topic and position the study within the broader field of child health research.

This study extends the existing literature by examining the impact of household tobacco and alcohol expenditure on child nutrition, specifically for children between the ages of six months and five years. Unlike prior studies that have primarily focused on the prenatal period or infant health, this study investigates a key yet under-researched period of child development. The central aim is to assess whether household expenditures on tobacco and alcohol crowd out food spending, thereby negatively affecting child health.

In doing so, this study makes several important contributions to the literature. First, it extends the "crowding-out" literature by empirically testing the effects of tobacco and alcohol expenditure on child nutrition, providing new evidence on how household consumption choices impact child health. Second, to address endogeneity concerns, the study employs an instrumental variable estimator to account for potential bias in the relationship between tobacco/alcohol expenditure and household income. This methodological approach strengthens the causal inference of the analysis. Finally, the study incorporates mediation analysis to explore the mechanisms through which tobacco and alcohol expenditure may adversely affect child health outcomes, offering a deeper understanding of how household consumption choices translate into health impacts.

This chapter is structured as follows: Section 2 reviews the relevant literature on child nutrition and household expenditure, with a focus on tobacco and alcohol consumption. Section 3 outlines the empirical strategy, including the instrumental variable approach and the use of mediation analysis. Section 4 presents the results of the analysis, followed by a discussion of the findings in Section 5. The chapter concludes with policy implications and suggestions for future research.

4.2 Literature review

Child nutrition has been a significant topic within both medical/public health and economics literature, each discipline offering distinct contributions. While the medical and public health fields primarily explore associations between various factors and child nutrition, economics tends to focus on establishing causal relationships, particularly through health production functions. These economic models offer deeper insights into the causes of health outcomes and are essential for informing policy decisions. Given the limitations of observational studies in establishing causality, this review emphasises the economics literature, particularly studies that use health production functions to understand the role of tobacco and alcohol consumption in influencing child health.

Child health production functions have long been a focal point for economists seeking to understand the determinants of child health. The seminal work of Grossman (1972), which introduced the concept of health production functions, laid the foundation for much of this research. This review considers studies published after 1972 that meet three criteria: (1) they are published in economics journals or explicitly use the child health production framework; (2) they focus on health outcomes rather than health disparities; and (3) they include tobacco or alcohol consumption as explanatory variables. Based on these criteria, 28 studies published over the past three decades were identified, with a particular focus on the impacts of tobacco and alcohol consumption on child health. The review highlights three major thematic strands: (1) the impact of maternal smoking on child health, (2) the effect of tobacco-control policies on child health, and (3) the influence of alcohol consumption on child health. This literature sets the stage for understanding how household expenditures on tobacco and alcohol may crowd out spending on food, negatively affecting child health—a central concern of this study.

The impact of maternal smoking on child health

The relationship between maternal smoking and child health is well-documented, particularly through the lens of the fetal origins hypothesis (Almond & Currie, 2011, Barker, 1990), which posits that prenatal factors significantly influence long-term health outcomes. Studies on this topic typically divide into two categories: (1) the impact of smoking during pregnancy on birth outcomes (e.g., birth weight, pre-term birth, small-for-gestational-age babies), and (2) the post-birth effects, such as respiratory problems, cognitive and behavioural issues, and increased risk of sudden infant death syndrome (SIDS). Six studies focus on the first category (Abrevaya & Dahl, 2008, Hamilton, 2001, Tominey, 2007, Walker, Tekin & Wallace, 2009, Yan, 2013, Yan,

2020), two focus on the second category (Sabia, 2008, Wehby et al., 2011), and one study addresses both types of outcomes (Dasgupta, Ghimire & Pacheco, 2022).

In the first strand of literature, several studies have found evidence of a negative impact of maternal smoking on birthweight. For instance, Hamilton (2001) used data from a randomised controlled trial in the USA and found that late-term maternal smoking decreased birthweight by up to 348g among children born to smokers compared to those of non-smokers. Similarly, Tominey (2007) used UK panel data to estimate that prenatal smoking lowered birthweight by 1.7%, but did not increase the likelihood of low birthweight (weight less than 2500g) or pre-term birth (less than 38 weeks). Abrevaya & Dahl (2008) also used panel data from the USA and found that smoking reduced birthweight throughout the distribution by between 26.2 and 82.5 grams.

Furthermore, Walker, Tekin & Wallace (2009) examined birth outcomes for mothers in the USA and found an overall negative effect of smoking on birthweight. However, their results are more nuanced when comparing teen mothers to older mothers, especially across race. For non-black teen mothers who smoked 10 or fewer cigarettes per day, their babies were 9.7 grams lighter compared to babies born to non-black smoking adults. For black teen mothers who smoked 10 or fewer cigarettes per day, their babies were 42.9 grams lighter than babies of adult black women with the same smoking intensity. This difference increased to 300 grams for babies born to women who smoked more than 20 cigarettes a day in both groups.

Two additional studies in this strand also reported negative effects of tobacco consumption on birth outcomes. Yan (2013) examined the effect of the timing of smoking cessation among pregnant women and found that cessation in the second trimester lowered birthweight by 68g and increased the risk of low birthweight by 1.2 percentage points compared to mothers who quit earlier. On the other hand, smoking throughout pregnancy lowered birthweight by 90g and also increased the risk of low birthweight by 1.2 percentage points. Yan (2020) shifted the focus to the impact of prenatal care, with the promotion of smoking cessation programmes as one of the potential mechanisms to affect birth outcomes. The study found that an additional prenatal visit was associated with a 22 grams higher birthweight, a 0.8 percentage points lower probability of low birth weight, and a 1.1 percentage points lower probability of pre-term birth.

One possible explanation for the varying magnitudes of the impact of smoking (or cessation during pregnancy) on birth outcomes could be the sensitivity of these estimates to model specification (Reichman et al., 2009). Addressing multiple sources of endogeneity poses a

challenge for researchers, and many studies are only able to control for one or two sources of endogeneity. For instance, all the studies controlled for unobserved heterogeneity by estimating either mother or sibling fixed-effects models. However, these attempts faced different limitations.

Hamilton (2001) study exploited a data structure to control for unobserved heterogeneity, but was limited to only 22% of the sample who were considered compliers. Meanwhile, Walker, Tekin & Wallace (2009) used an OLS model and a matching estimator, both of which are likely to be biased, as they do not account for unobserved heterogeneity. The fixed-effects model used in the study may have issues with measurement error in the smoking variables.

A study by Abrevaya & Dahl (2008) did not account for potential feedback or dynamic effects, specifically whether prenatal care and smoking in later pregnancies (in older mothers) may be correlated with birth outcomes in earlier pregnancies (younger mothers). However, they conducted sensitivity analyses and concluded that feedback effects were unlikely to influence their results significantly.

The literature on postpartum maternal smoking and its effects on post-birth outcomes has been studied in a small number of research papers. Sabia (2008) examined the probability of the onset of asthma among children aged one to three and found that postpartum maternal smoking led to a 2%-3% higher probability of asthma at age one and a 4%-5% higher probability by age three, with asthma rates among the former group increasing by 4 to 7 percentage points. Wehby et al. (2011) studied the rate of child neurodevelopment[17] and the risk of adverse neurological outcomes as measured by the risk for neurodevelopmental delay, based on the instrumental norms and by a measure of the child's neurodevelopment relative to the sample's average neurodevelopment. They found that maternal smoking during pregnancy increased the probability of high neurodevelopment risk by 0.07 percentage points, with larger effects for low socioeconomic groups. Dasgupta, Ghimire & Pacheco (2022) examined the effects of postpartum maternal smoking on child weight from birth to age five, and found that children of smokers weighed 530 grams less than those of non-smokers at birth. Additionally, Agee, Atkinson & Crocker (2009) found that indirect effects such as poor air quality had a negative influence on children's cognitive and behavioural health outcomes.

[17] Neurodevelopment is a term referring to the brain's development of neurological pathways that influence performance or functioning (e.g., intellectual functioning, reading ability, social skills, memory, attention or focus skills).

The techniques used in these studies deserve comment. A study by Yan (2013) used a difference-in-difference estimation technique to compare outcomes among mothers who restarted smoking within a year or two of their child's birth with outcomes among those who did not restart smoking. The results were robust to individual fixed effects, but could not account for reverse causality-induced endogeneity, which may bias the fixed-effects estimates downwards. By way of contrast, Wehby et al. (2011) and Dasgupta, Ghimire & Pacheco (2022) also accounted for endogeneity via instrumental variable estimation. Furthermore, Wehby et al. (2011) used fixed effects to account for unobserved heterogeneity, but acknowledged possible estimation bias due to self-selection into smoking based on relevant unobservables. Agee, Atkinson & Crocker (2009), in contrast to the previous studies, argued that single-equation estimations induce an omitted variable bias, as they fail to consider various positive and negative inputs that may impact aspects of a child's health. To address this issue, they employed an output-based directional distance function for a balanced panel of households, a tool commonly used in production theory to determine allocative and technical efficiency.

The impact of tobacco control policies on child health outcomes

Research in the field of maternal smoking and child health outcomes also investigates the effects of different tobacco-control policies, with a focus on maternal smoking as a mediator for the policy's impact. Although mediation analysis is often associated with the statistics and public health fields (Imai, Keele & Yamamoto, 2010, Imai et al., 2011, MacKinnon, 2012), economists have been investigating the causal mechanisms of various treatment effects for many years (Celli, 2021, Heckman, James J & Pinto, 2015). The studies are further divided into two groups, namely, (A) studies on the impact of tobacco taxes; (B) studies on the impact of smoking bans.

A. Tobacco taxes

Five research studies investigated how tobacco-tax hikes affect child health outcomes. Three were from the USA (Lien & Evans, 2005, Mellor, 2011, Simon, 2016), one from Canada (Sen & Piérard, 2011), and one from Indonesia (Block & Webb, 2009b).

Lien & Evans (2005) found that tobacco-tax increases were positively associated with birth weight and negatively associated with the risk of low birth weight. The study estimated that a 10% increase in the tax rate led to a 5% decrease in the probability of pregnant women smoking. Furthermore, the authors demonstrated that, among women who were more likely to quit smoking due to the tax increase, continued smoking was associated with a reduction in birth

weight by an average of 182 grams and doubled the likelihood of a low-weight birth. These findings were based on a pooled sample of states analysed in the study.

Sen & Piérard (2011) focused on the impact of lagged cigarette taxes on birth weight and foetal deaths. The study found that a 10% increase in lagged cigarette taxes decreased infant mortality by between 1.7% and 1.8%, but did not have any relationship to birth weight.

Mellor (2011) explored the link between increased cigarette costs, increased food expenditure, and increased calorie consumption, which affected the BMI of children whose mothers smoked. The study found that both cigarette taxes and prices had a significant positive effect on the BMI of children with smoking mothers. Specifically, a $1 increase in cigarette prices resulted in a BMI increase of between 0.18 kg/m^2 and 0.22 kg/m^2, based on a BMI baseline of 17.7 kg/m^2.

Similarly, Block & Webb (2009b) investigated the effect of cigarette expenditures on food expenditures at the household level, and its impact on child health. Their study found that, in low-income households, a sizeable portion of income was diverted to cigarette expenditure, resulting in declines in both food quantity and quality. These findings suggest that smoking may have a negative impact on child health outcomes, through reduced food intake.

Simon (2016) studied the impact of state cigarette-tax increases on school absenteeism, doctor visits, hospitalizations, and asthma rates to show that for a dollar increase in the state cigarette tax, there was a 10% decrease in sick days from school, and a 4.7% decrease in having more than two doctor visits in the previous 12 months. Moreover, a reduction was observed in hospitalisations and asthma rates. Markowitz et al. (2013) studied the effect of tax increases on human capital mechanisms, such as age, education, and income, among pregnant women. The results suggest that teenage mothers benefit the most from tobacco-control policies, with a $1.00 increase in cigarette tax increasing the probability of full-term birth by 3.1 percentage points among teenage mothers and by 0.7 percentage points among non-teenage mothers.

In a different study by Markowitz (2008), the direct effects of cigarette price and tax increases on the incidence of SIDS were examined. The results showed that a 10% increase in the real price of cigarettes resulted in 6.9% to 7.6% fewer deaths from SIDS, while a similar increase in taxes resulted in 1.6% to 1.8% fewer SIDS deaths on average. A later study explored the impact of these policies on SIDS across 23 developed countries over a 20-year period (King, C., Markowitz & Ross, 2015). The study found that a $1 increase in the price of cigarettes reduced the number of SIDS deaths by an average of 12.5 to 12.8 per year.

A study by Nonnemaker & Sur (2007) investigated the impact of tobacco prices on child health outcomes in rural Bangladesh, with a particular focus on nutritional status. The study used data from a longitudinal survey of 600 households with children aged 2–10, conducted between 1996 and 1997. Their analysis found that higher tobacco prices were positively correlated with improved child nutritional status, particularly for height-for-age and weight-for-height outcomes. Additionally, while tobacco prices were found to have no significant effect on the incidence of respiratory illness, the study underscores the broader potential of tobacco tax policies to improve child health through changes in household expenditure behaviour.

The studies discussed in this literature demonstrate the complexity of the relationships under examination, and many have addressed the concern of endogeneity from various sources in diverse ways. For instance, Lien & Evans (2005) used the tax hike as an instrument for smoking status in an instrumental variables estimation, in addition to estimating a naive OLS model. In contrast, Markowitz (2008) used a fixed-effects Poisson estimator to account for endogenous smoking variables. Other studies have addressed different issues with various methods, such as Markowitz et al. (2013), who used an ordered probit regression with state fixed-effects to address concerns about the true effect of cigarette smoking on birth outcomes, which may substantially deviate from the mean. Sen & Piérard (2011) used generalised least squares to address the limitation related to the small number of clusters in the study, while Block & Webb (2009b) used a combination of parametric and non-parametric techniques to circumvent the endogeneity problem. In contrast, Simon (2016) used an event-study design. Mellor (2011) used difference-in-differences to address the endogeneity issue.

B. Smoking bans

This review examines nine studies that investigate the impact of smoking bans on child health. These studies were conducted in various countries, including the US (Cooper & Pesko, 2017, Gao & Baughman, 2016, Williamson, 2010), Canada (McGeary et al., 2020), Germany (Coneus & Spiess, 2012), Norway (Bharadwaj, Johnsen & Løken, 2014), and Hungary (Hajdu & Hajdu, 2018).

In general, smoking bans were found to have a positive effect on birth weight. Studies by McGeary et al. (2020) and Hajdu & Hajdu (2018) both found that the ban increased birthweight by an average of 13 grams and 56 grams, respectively. Coneus & Spiess (2012) found that smoking, proxied by exposure to carbon monoxide, lowered birth weight by 289 grams on

average, while Bharadwaj, Johnsen & Løken (2014) found that children born after a smoking ban was implemented were less likely both to be born pre-term or to have very low birth weight.

Furthermore, smoking bans were found to have positive impacts among older children. Williamson (2010), for instance, found that having a smoker in the home reduced lung function by 1.4%, while having two smokers reduced it by 2.2%. Similarly, McGeary et al. (2020) noted reductions in respiratory allergies, asthma attacks, the number of emergency room visits, ear infections, and reports of poor or fair health following the introduction of a smoking ban. The implementation of smoking bans was also found to reduce the number of SIDS deaths, depending on whether the ban was in private workplaces (resulting in 3.5 fewer SIDS deaths per 1000 live births) or in restaurants (resulting in 1.4 fewer SIDS deaths per 1000 live births) (Markowitz, 2008).

However, some studies found counterintuitive results. Gao & Baughman (2016), for example, found no significant impact of a comprehensive smoking ban on several infant health outcomes. Instead, they observed a slight increase in the prevalence of low birth weight and the probability of very low birth weight among children born to young mothers. Meanwhile, Cooper & Pesko (2017) investigated the impact of indoor vaping restrictions on child health outcomes, and found no effects on birth outcomes, but they did note some evidence of a slight decrease in pre-term and very pre-term births. The study found that, following the introduction of a comprehensive indoor vaping restriction, combustible smoking participation among pregnant women increased by between 0.9 to 2 percentage points, while smoking intensity also went up by 10.3 to 12.6 more cigarettes smoked per month. A possible explanation for these counterintuitive findings was the change in smoking venue from workplaces and other public venues, where smoking bans were in effect, to homes, as suggested by Adda & Cornaglia (2010).

Most of the reviewed studies utilised a difference-in-difference framework to estimate the treatment effect of smoking bans on various child health outcomes (Bharadwaj, Johnsen & Løken, 2014, Cooper & Pesko, 2017, Hajdu & Hajdu, 2018, McGeary et al., 2020). In contrast, Coneus & Spiess (2012) employed a mother fixed-effects model to control for unobserved heterogeneity, while Williamson (2010) used a cigarette excise tax as an instrument for exposure to tobacco smoke. Gao & Baughman (2016) used OLS, but conducted a series of robustness checks for endogeneity and sample selection.

The impact of parental alcohol use on child health

Two research studies focused on the impact of parental alcohol use on child health outcomes: one from the US (Snow Jones, Miller & Salkever, 1999), and the other from Russia (Giannelli, Mangiavacchi & Piccoli, 2013).

Snow Jones, Miller & Salkever (1999) explored the impact of parental alcohol consumption, ranging from heavy/problem drinking to light/moderate drinking, on child behavioural health. The study revealed that parental alcohol use had negative effects on child health regardless of which parent drank. The study suggests that parental alcohol consumption is an input in the child health production function.

Meanwhile, Giannelli, Mangiavacchi & Piccoli (2013) examined the relationship between alcohol consumption and parental child care, which is a critical input in overall child health. The study found that alcohol consumption by men reduced the amount of time they spent on childcare. However, alcohol consumption did not affect the labour supply or wages of either men or women. On the other hand, alcohol consumption by women (or female caregiver) had no significant impact on childcare.

Snow Jones, Miller & Salkever (1999) employed OLS and two-stage least squares models to estimate the effect of parental alcohol use on child behavioural health. The study consistently found negative effects of alcohol consumption on child health, regardless of the estimation method or model specification used. However, the magnitude of the two-stage least squares estimates was larger than the OLS estimates.

On the other hand, Giannelli, Mangiavacchi & Piccoli (2013) used a seemingly unrelated regression framework to estimate the impact of alcohol consumption on parental child care. The study accounted for the correlation of errors and sample selection. The complexities of modelling the impact of alcohol consumption on child health were akin to those encountered in the tobacco consumption studies discussed earlier.

4.3 A conceptual framework for child health and nutrition

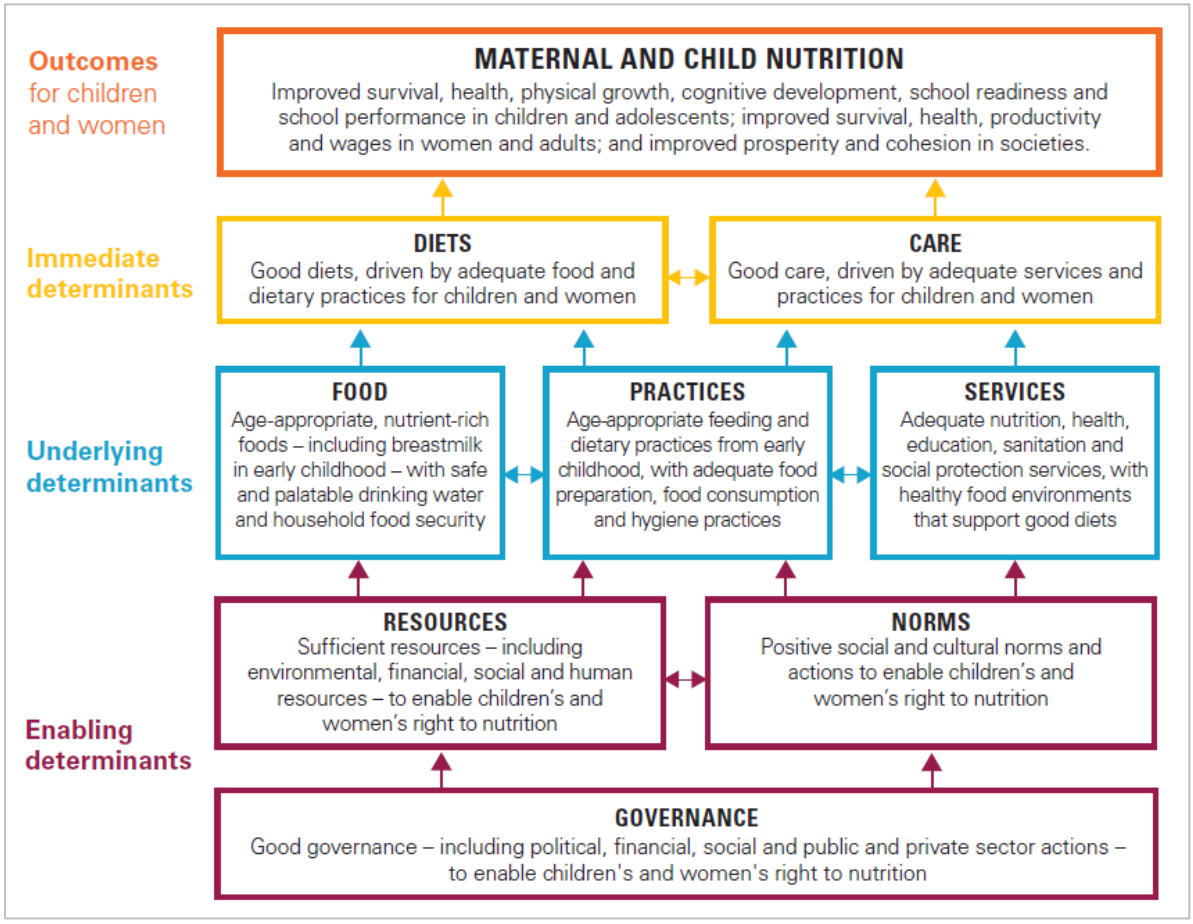
Over the years, many conceptual frameworks have been developed to help researchers and policymakers comprehend the intricate relationships that influence individual child health and nutritional status. Among these, the Mosley-Chen Framework has been historically used by social scientists in the analysis of child mortality. On the other hand, UNICEF's framework (UNICEF, 1990) and its updated version (UNICEF, 2020) have been designed to focus specifically on nutritional analysis.

The Mosley-Chen framework identifies five proximate variables that directly influence the risk of child survival (Mosley & Chen, 1984). The five variables are: (1) maternal factors such age, parity (number of children a woman has given birth to) and birth interval; (2) environmental contamination, encompassing factors such as air quality, food and water quality, soil quality, and the presence of insect vectors of disease; (3) nutritional deficiency, including the intake of calories, protein and micronutrients; (4) injury, whether accidental or intentional; and (5) personal illness control, such as medical treatment and personal preventive measures. The framework presupposes that all social and economic determinants will operate through these variables to affect child morbidity and mortality.

As an alternative framework, the UNICEF framework focuses on maternal and child nutrition and identifies three broad categories of determinants, viz.: immediate, underlying, and enabling determinants (UNICEF, 2020). As depicted in Figure 4.1, the immediate determinants of nutrition include diet and care, which interact with each other to affect nutritional outcomes. Underlying factors include the food, practices, and services available to women and children in their households, communities, and environments that facilitate good nutrition. Lastly, enabling determinants comprise the political, financial, social, cultural, and environmental conditions that promote good nutrition for women and children. These determinants are broadly categorised into governance, resources, and norms.

Alcohol and tobacco use among adults can be incorporated into the UNICEF framework in several ways. As alluded to earlier, the consumption of both substances can affect household food security by diverting potential expenditure on nutritious food towards maintaining tobacco or alcohol use. This has been explored in various studies (Chelwa & Van Walbeek, 2014, John, Ross & Blecher, 2012, Masa-Ud, Chelwa & Van Walbeek, 2020, Nguyen & Nguyen, 2020). Secondly, exposure to second-hand smoke can have adverse effects on the air quality of the child's environment, potentially increasing their risk of contracting various childhood diseases (Öberg et al., 2011). Lastly, alcohol abuse has been linked to child care lapses (Sarkola et al., 2007)), which is an immediate determinant of child nutrition in the UNICEF framework.

Figure 4.1: UNICEF Conceptual Framework on the Determinants of Maternal and Child Nutrition, 2020.



Source: UNICEF (2020)

4.4 Empirical strategy

Health status in relation to the equations from the previous section is generally not directly observed (Behrman & Deolalikar, 1988). This means that, empirically, most studies on this topic use some measure of nutritional status. The current chapter follows the literature in this respect, and begins with a detailed discussion of the dependent variables to be used in the empirical model.

4.4.1 Dependent variables

The most common measures of child nutritional status are based on a combination of anthropometric measures that consider the height and weight of a child at a given age. The three conventional measures used are the ratios of height-for-age (HAZ); weight-for-height

(WHZ); and weight-for-age (WAZ) transformed into z-scores[18] (Charmarbagwala et al., 2004). Using these three measures, the WHO categorises stunting, wasting, and undernutrition, if the value of any individual's HAZ, WHZ or WAZ measure is two standard deviations lower than a predetermined norm (World Health Organization, 1995). A value lower than three standard deviations is considered severe. The states (stunted, wasted, and undernutrition) are all broadly categorised as states of anthropometric failure.

Importantly, these different measures capture different dimensions of nutritional status. Stunting or low height-for-age is widely considered to be a measure of chronic nutritional deficiency, and thus acts as a good measure of long-run social conditions. Wasting or low weight-for-height, in contrast, is indicative of current acute nutritional deficiency, with little or no account for past insufficient food intake. Undernutrition or low weight-for-age is the most ambiguous of the three measures, reflecting both wasting and stunting. It thus does not distinguish between long-term malnutrition and short-term undernourishment.

Since the focus of this chapter is to determine the impact of household tobacco and alcohol expenditure on child health outcomes, it adopts two continuous dependent variables, namely, HAZ and WAZ suitable to capture the effect. At the outset, a baseline model is estimated using ordinary least squares regression (OLS) as shown below:

$$y = \alpha + \beta_1 Tob + \beta_2 Alc + \beta_3 C + \beta_4 H + \beta_5 X + u \qquad (1)$$

where *y* is any one of the two anthropometric measures (HAZ or WAZ), *Tob* and *Alc* are the lagged household expenditure on tobacco and alcohol, *C* is a vector of child characteristics such as age, gender, birth circumstances, and current or past serious illness, *H* is a vector of household characteristics, such as household size and lagged household income, and X is a vector of environmental factors such as region and province. Lastly, *u* is the error term. A negative and statistically significant coefficient for β_1 can be interpreted as the effect of tobacco or alcohol expenditure on the relevant anthropometric measure, *ceteris paribus*.

4.4.2 Expenditure variable transformation

Household expenditure on cigarettes and alcohol often exhibit a high frequency of zero values, leading to a right-skewed variable, sometimes referred to as a zero-inflation issue in the variable. If such a variable is used as an outcome of interest, it is crucial to identify the underlying source of these observed

[18] The z-score is a measure of the deviation from the median of that sex and age in months divided by the standard deviation from the reference population for that sex and age. The z-scores for the reference population have an asymptotic standard normal distribution.

zeros, as it significantly impacts the choice of estimator. Zeros resulting from an optimisation process can be considered corner solutions, arising from a constrained utility maximisation problem. In such cases, the commonly employed estimators are the Tobit model or an extension of the two-part model (Humphreys, 2015, Wooldridge, 2010). Conversely, if the zeros stem from non-response, missing outcomes, or external factors beyond the control of the agent, they indicate a selection problem. In such situations, the more suitable estimator is the Heckman sample selection model (Humphreys, 2015, Wooldridge, 2010).

However, the issue confronted in this chapter is one of excessive zeros in an independent variable, specifically cigarette or alcohol expenditure, sometimes referred to as 'spike at zero' variables (Jenkner et al., 2016, Lorenz et al., 2017). While econometric methods do not demand specific distributional assumptions for independent variables, it is widespread practice to transform right-skewed variables such as these to mitigate their skewness. Transformation offers several advantages, including improved adherence to the assumptions of the regression model and enhanced robustness of the estimated coefficients against outliers and extreme values. Existing literature often employs the natural logarithm transformation in order to address the issue of skewed distribution associated with the independent variable. However, a notable limitation of this transformation is its ineffectiveness in handling zero and negative values.[19]

To overcome this limitation, researchers commonly employ the strategy of adding an arbitrary value to zeros before applying the transformation (MaCurdy & Pencavel, 1986). This approach also has limitations, because the choice of this arbitrary value influences the magnitude of the resulting coefficients. In this chapter, the proposed approach is the increasingly popular inverse hyperbolic sine (IHS) transformation[20] (Bahar & Rapoport, 2018, Clemens & Tiongson, 2017). The IHS transformation of a variable x is defined as $\tilde{x} = \text{arcsinh}(x) = \ln(x + \sqrt{x^2 + 1})$. The transformation offers several distinct advantages. Like the natural logarithmic transformation, it adjusts for skewness, thereby mitigating the influence of outliers and extreme values. Secondly, unlike the natural log transformation, it retains zero values, which is especially advantageous for the current analysis. Finally, when the independent variable is subjected to the IHS transformation, it can be interpreted in a similar manner to the natural logarithm. Specifically, a 1% increase in the IHS-transformed variable is associated with a unit change in the outcome variable, when the outcome variable is expressed in levels, and a percentage change when the outcome is expressed in logarithms (Friedline, Masa & Chowa, 2015).

[19] The natural log of zero or negative values cannot be taken, because it does not exist.

4.4.3 Instrumental variable approach

In the model presented above, it is important to recognise that the lagged cigarette expenditure and lagged alcohol expenditure variables are likely to be endogenous. This implies that smoking or drinking behaviour may be influenced by unobserved factors that simultaneously affect both cigarette expenditure and child health outcomes. For instance, parents who engage in smoking or drinking may possess varying attitudes towards health, experience differing levels of stress, or belong to different socioeconomic backgrounds, all of which can impact child health outcomes.

Moreover, household income is considered endogenous, due to the simultaneous determination of income and child nutritional status through the labour-time allocation decision in the utility maximisation problem. The decision regarding the allocation of time between wage-generating labour and childcare has an impact on the income variable. Additionally, it is important to note that household income poses measurement challenges in household surveys, particularly in low-and-middle-income countries (LMICs) (Deaton, 2018). In both cases, the resulting endogeneity introduces bias in the ordinary least squares (OLS) estimates, in the form of upward bias in the case of simultaneity and downward bias in the case of measurement error.

In this study, we employ several instruments to address potential endogeneity in the relationship between household expenditure on tobacco and alcohol and child health outcomes. Endogeneity concerns arise from the possibility that unobserved factors may simultaneously affect both the expenditure on these goods and child health, potentially leading to biased estimates. The instruments selected—adult-sex ratio, community drug/alcohol abuse frequency, and religious activity importance—are designed to mitigate this endogeneity. Each instrument has been chosen based on economic reasoning, and while they offer reasonable justification, their limitations must also be acknowledged.

Adult-sex ratio as an instrument for cigarette expenditure

The adult-sex ratio[21] (the ratio of adult males to females aged 15 and above in a household) is a suitable instrument for cigarette expenditure due to gender-based differences in smoking behaviour. Empirical studies consistently show that males are more likely to smoke and consume tobacco than females. A household with a higher ratio of adult males is likely to exhibit higher tobacco consumption, thereby increasing tobacco expenditure. The adult-sex

[21] Adult-sex ratio is calculated as the ratio of the number of adult males to the number of adult females in a household, all older than 15 years of age.

ratio, thus, influences cigarette expenditure in a manner that is theoretically sound and consistent with observed patterns in smoking behaviour across genders.

The primary limitation of this instrument is its potential violation of the exclusion restriction. While the adult-sex ratio may affect tobacco expenditure, it could also influence other household behaviours (such as spending on health, nutrition, or education), which could independently affect child health outcomes. Moreover, households with a greater number of adult males might also have higher incomes or different household structures, which could confound the relationship between tobacco expenditure and child health. Therefore, it is crucial to ensure that any such pathways are controlled for in the model.

Community drug or alcohol abuse frequency as an instrument for alcohol expenditure

The frequency of drug or alcohol abuse within the community[22] serves as an instrument for alcohol expenditure because social norms and peer behaviour strongly influence alcohol consumption. Communities with higher rates of substance abuse are likely to exhibit higher levels of alcohol consumption, making the frequency of abuse a reasonable instrument for household alcohol expenditure. This instrument is grounded in the idea that peer and community effects can significantly shape individual behaviour, including expenditure decisions on alcohol.

The key limitation of this instrument lies in the measurement error inherent in using a self-reported survey measure (based on a Likert scale). If respondents misreport the frequency of drug or alcohol abuse in the community, the instrument may be weak or invalid. Furthermore, while it is assumed that community-level alcohol abuse only influences child health outcomes through its effect on alcohol expenditure, community substance abuse could directly affect children in other ways (e.g., through environmental factors or family stress). This poses a risk of violating the exclusion restriction. Additionally, the relevance of this instrument may vary across different cultural or regional contexts, where alcohol consumption patterns might not follow the same social norms or peer effects.

[22] The NIDS Household questionnaire for Waves 2 to 5 include the following question: "How common is drug or alcohol abuse in your neighbourhood?" This is followed by a five-point Likert scale ranging from "Never happens" (1) to "Very Common" (5). However, Wave 1 is missing this question, and a different instrumental variable is used based on the following question: "How important are religious activities in your life?" A four-point Likert scale from 1 (not important) to 4 (very important) is used to calculate an average of these scores for all adults interviewed within a household.

Religious activity importance as an instrument for alcohol expenditure (Wave 1)

The importance of religious activities in a household can serve as a valid instrument for alcohol expenditure, as religious participation is often negatively correlated with alcohol consumption. This instrument was used for the first wave of NIDS as the community alcohol variable was not available. Households that place a high value on religious activities are typically less likely to spend on alcohol due to religious proscriptions against drinking. The strength of this instrument lies in its reflection of social and cultural norms that shape consumption patterns, making it a relevant tool for isolating the impact of alcohol expenditure on child health outcomes.

The main limitation of using religious activity importance as an instrument is the potential heterogeneity in its effect across different households. The influence of religion on alcohol consumption may differ depending on the specific religious denomination or the personal beliefs of household members, which could weaken the instrument's relevance in certain contexts. Moreover, like the community drug/alcohol abuse instrument, religious activity could influence child health outcomes through other channels, such as affecting household attitudes toward healthcare, nutrition, and education. This could violate the exclusion restriction, as religious norms may shape child health outcomes independently of alcohol expenditure.

Household asset index as an instrument for household income

The household asset index is a widely used proxy for household income in econometric analyses, particularly in contexts where direct income data are unreliable or unavailable. The asset index captures long-term wealth and is correlated with the household's ability to purchase goods, including tobacco and alcohol. As such, it is a relevant instrument for household income, which in turn affects tobacco and alcohol expenditure.

The household asset index may not perfectly capture short-term income fluctuations, which are important drivers of expenditure on non-essential goods like tobacco and alcohol. As a result, it could be a weaker instrument for household income in contexts where income variability plays a significant role in consumption decisions. Additionally, the asset index may affect child health outcomes not only through its correlation with income but also through its relationship with access to healthcare, nutrition, and other resources. This could pose a challenge to the exclusion restriction.

The instruments employed in this study are grounded in sound economic reasoning and offer a robust approach to addressing endogeneity concerns. However, their validity hinges on the assumption that they affect child health outcomes solely through their impact on tobacco and alcohol expenditure. To assess the robustness of the results, it will be crucial to perform tests for weak instruments and over-identification. Additionally, the limitations associated with these instruments, such as potential violations of the exclusion restriction and measurement errors, will be carefully considered when interpreting the results.

4.4.4 Mediation analysis with multiple endogenous regressors

Mediation analysis is used to identify and evaluate the mechanism of any effects uncovered by the preceding analysis. The goal of this type of analysis is to decompose the total treatment effect into two components: the indirect effect and the direct effect (Celli, 2021, Heckman, James J & Pinto, 2015). The indirect effect captures the share of the total effect that occurs due to one or more intermediate variables, called mediators, while the direct effect captures the remaining effect that does not operate through the mediator.

The traditional approach to mediation analysis uses OLS to estimate three equations:

$$\textit{Model 1:} \quad Y = \alpha + \beta_1 T + \mu \tag{2}$$

$$\textit{Model 2:} \quad M = \alpha + \beta_2 T + \varepsilon \tag{3}$$

$$\textit{Model 3:} \quad Y = \alpha + \beta_3 T + \gamma M + \delta \tag{4}$$

where Y represents the outcome of interest, T represents the treatment variable, M represents the mediator, $\alpha, \beta_1, \beta_2, \beta_3$ and γ are parameters to be estimated and μ, ε and δ are the respective error terms. This approach follows Baron & Kenny (1986) and makes the strong assumption that T and M are exogenous. If, however, the exogeneity assumption is violated, the usual biases associated with OLS apply to this analysis.

In the current context, the outcome of interest (Y) represents one of the child nutrition variables (HAZ or WAZ), the treatment variable (T) represents lagged expenditure in cigarettes or alcohol, and the mediator (M) represents lagged expenditure on food. As highlighted earlier, both cigarette and alcohol expenditure are likely to be endogenous in the model. In addition, household income is also considered endogenous.

Thus, following Frölich & Huber (2017), this chapter adopts separate instruments to address the endogeneity of T. In this case, as noted earlier, the adult-sex ratio is used as an instrument for cigarette expenditure. Frequency of alcohol or drug abuse in a community is used as an instrument for alcohol expenditures in Waves 2 to 5, while a household's average level of religiosity is an instrument for alcohol expenditure in Wave 1. An asset index serves as an instrument for household income in all five waves.

The resulting identification framework specifies the three models as discussed earlier. First, the model specification shown in Equation 1 is estimated, using the instrumental variables as described.

$$y = \alpha + \beta_1 Tob + \beta_2 Alc + \beta_3 C + \beta_4 H + \beta_5 X + u \quad (5)$$

The coefficient $\beta_1$and β_2 in the above model estimate the *total effect* of cigarette and alcohol expenditure.

Second, a model regressing lagged food expenditure on lagged cigarette or alcohol expenditure using the same instruments as in the first step is estimated:

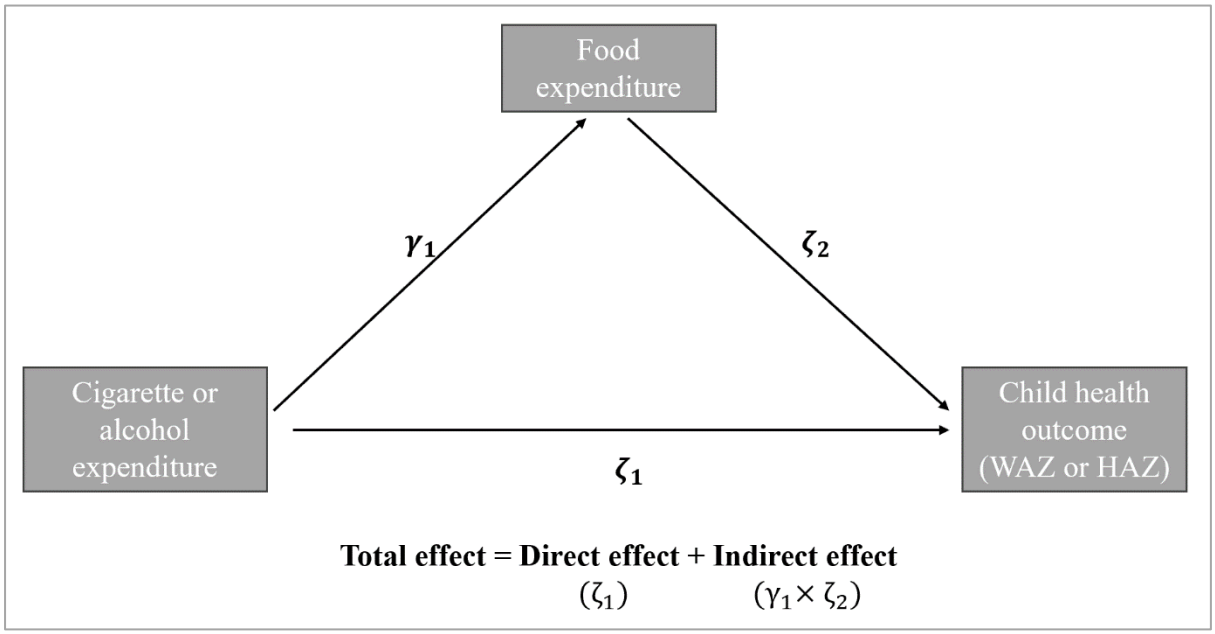
$$Foodexp = \alpha + \gamma_1 Tob(Alc)expn + \gamma_2 C + \gamma_3 H + \gamma_4 X + \delta \quad (6)$$

The last step estimates a model that augments Equation 1 by including lagged food expenditure as follows:

$$y = \alpha + \zeta_1 Tob + \zeta_2 Alc + \zeta_3 Foodexpn + \zeta_4 C + \zeta_5 H + \zeta_6 X + \delta \quad (7)$$

In the third model, the coefficients ζ_1 and ζ_2 estimate the *direct effects* of cigarette and alcohol expenditure. To derive the indirect effect, one simply subtracts the direct effect from the total effect, yielding: *Indirect Effect = Total Effect – Direct Effect.* Another method to calculate the indirect effect involves multiplying the coefficient on food in Equation 4 by the coefficient on tobacco in Equation 3, resulting in: *Indirect effect* = $\gamma_1 \times \zeta_2$ (MacKinnon, 2012, MacKinnon, Lockwood & Williams, 2004). In linear models, the two methods are equivalent. Figure 4.2 presents these concepts in a diagram representing the direct and indirect effects of cigarette or alcohol expenditure on child health outcomes.

Figure 4.2: Mediated effect of cigarette or alcohol expenditure on child health outcomes



In the absence of additional theoretical guidance for this instrumental variable (IV) framework, the chapter adopts the standard practice in the IV literature by utilising recommended first-stage F-test statistics to evaluate weak instrument identification after estimation (Andrews, Stock & Sun, 2019, Stock, James H & Yogo, 2005).

4.5 Data

Data for this chapter are drawn from South Africa's National Income Dynamics Study (NIDS) survey, which is a nationally representative panel dataset, conducted roughly every two years between 2008 and 2017. A stratified two-stage cluster sampling design was adopted for the base survey (Leibbrandt, Woolard & de Villiers, 2009). A total of 7,296 households were interviewed in the first wave and were followed up over the following four waves.

Data collection involved administering household and individual questionnaires to capture a comprehensive range of information from all adults and children included in the sample. This encompassed various household characteristics and details of household members, such as income, education levels, and age. Of particular importance is that the NIDS data aimed to measure the height and weight of all children aged six months to 14 years, forming the focal population of interest across all surveys. The sample was restricted to observations with complete information on children in the households surveyed.

It is worth noting that, while efforts were made to follow up on children from the initial 2008 sample in subsequent waves, the considerable wave-on-wave attrition rates and within-wave

missing observations rendered the utilisation of panel data methods incorporating all five waves impractical for this chapter. As a result, the chapter employs cross-sectional techniques, as previously described, while leveraging the availability of lagged expenditure on alcohol or tobacco to examine the long-term effects of such expenditure on child health outcomes.

4.6 Descriptive statistics

Table 4.1 presents summary statistics for the variables used in the regression models. The percentage of male and female children was evenly distributed across all waves. In terms of race, the sample was predominantly African, making up around 85% of the sample. Coloured children accounted for approximately 8%, while Whites made up about 5%, and Indian or Asian children approximately 2% of the sample across all waves.

NIDS examined several key indicators related to child health and household socioeconomic status over the five waves of data collection. The percentage of children experiencing major illnesses exhibited only slight fluctuations, remaining consistently below 6% across all waves. In contrast, there were marked improvements in all three measures of child malnutrition, as evidenced by declines in the prevalence of stunting (from 17.2% to 11.8%), underweight (from 9.6% to 5.5%), and wasting (from 4.7% to 3.3%) over the same period.

Meanwhile, the data also revealed significant changes in household income and expenditures. In real terms, the average monthly income increased from R7,856.50 to R10,883.0 between Wave 1 and Wave 5. In comparison, monthly adult spending on cigarettes increased between Waves 1 and 4 from R21.30 to R24.60 per adult, but declined to R19.90 per adult in Wave 5. Monthly alcohol expenditure also exhibited a significant increase from R13.10 to R31.00 per adult between Wave 1 and 3, but declined slightly to R26.60 per adult in Wave 4 and R25.20 per adult in Wave 5. Marginal increases were also observed in real monthly food expenditures per capita, which rose from R310.30 in Wave 1 to approximately R341.40 in Wave 5.

Table 4.1: Summary descriptive statistics

Variable	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5
Child Characteristics					
Male (%)	49.7	49.3	49.8	49.0	50.0
Female (%)	50.3	50.7	50.2	51.0	50.0
African (%)	85.0	84.1	84.0	85.2	85.8
Coloured (%)	8.1	8.5	8.7	7.8	8.0
Indian (%)	1.8	1.7	1.9	2.2	1.8
White (%)	5.1	5.7	5.4	4.8	4.4
No history of major illness (%)	94.3	96.8	95.9	95.0	95.5
History of major illness (%)	5.7	3.2	4.1	5.0	4.5
Stunted[#] (%)	17.2	21.5	20.5	12.5	11.8
Underweight[#] (%)	9.6	8.9	8.4	5.6	5.5
Wasted[#] (%)	4.7	3.7	5.7	3.9	3.3
Household characteristics					
Mean monthly hh income (Rands)[±]	7,856	8,686	9,782	9,826	10,883
Mean monthly tob. exp. per adult (Rands)[±]	21.3	23.6	22.0	24.6	19.9
Mean monthly alc. exp. per adult (Rands)[±]	13.1	13.3	31.0	26.6	25.2
Mean monthly food exp. per capita (Rands)[±]	310.3	310.9	297.4	317.1	341.4
Mean household size	6.3	6.6	6.3	6.3	6.0
Mean number of children (below 5 yrs)	3.0	3.1	2.9	3.0	2.9
Mean age of children (below 18 yrs)	6.7	6.8	6.7	6.7	6.6
Mean number of adults (18 yrs and above)	3.3	3.7	3.4	3.4	3.2
Toilet type: Flush or chemical	86.1	90.3	88.4	88.1	89.7
Toilet type: Pit latrine or other	2.1	2.5	3.1	4.1	4.3
Toilet type: None	11.8	7.2	8.5	7.8	6.0
Water source: Tap	46.4	59.0	55.4	53.8	55.9
Water source: Borehole or tank	45.1	37.4	40.9	43.1	41.8
Water source: Rain, spring or dam	8.5	3.7	3.7	3.1	2.4
Community characteristics					
Urban households (%)	51.1	53.2	55.1	55.8	58.2
Rural households (%)	48.9	46.8	44.9	44.2	41.8
Number of observations	8,649	6,806	7,295	7,538	7,693

± Income and expenditure amounts expressed in real 2017 Rands using consumer price indices from StatsSA. # Stunted refers to a low height-for-age z-score, underweight indicates a low weight-for-age z-score, and wasted denotes a low weight-for-height z-score

As for household demographics, the data reveal that average household size remained relatively constant at around six people, with half of those being children under the age of five. The average age of children below 18 years was just below 7 years of age. Most households had access to flush or chemical toilets (over 85% across all waves) and obtained water from tap sources (about 55% across all waves), followed by borehole or tank sources (about 42%). The remaining households relied on rain, spring, or dam water. Most households were in urban areas (55%), with the remaining 45% living in rural areas.

4.7 Econometric results

This chapter aims to estimate a model that examines the impact of lagged tobacco and alcohol expenditure on child WAZ and HAZ scores. Initially, it employs OLS models following the specification outlined in Equation 9, treating household income, lagged cigarette expenditure, and lagged alcohol expenditure as exogenous variables. However, the coefficients on the variables of interest are attenuated towards zero (refer to Tables C4 and C5 in the annex of this chapter), indicating a potential bias arising from the previously discussed endogeneity concerns. To address this issue, a two-stage least squares (2SLS) regression methods is used for this analysis.

The 2SLS regression results are presented in two distinct phases. The first phase focuses on the effects of lagged tobacco and alcohol expenditure on current child WAZ and HAZ scores, which are presented in Tables 4.2. Subsequent tables provide diagnostic results by which to evaluate the appropriateness and quality of the chosen instruments. In the second phase, mediation analysis is conducted in order to explore the potential pathways through which the observed impacts from the first phase operate.

Table 4.2 displays the key findings of the analysis, focusing on the relationship between lagged cigarette expenditure and child WAZ and HAZ scores. Recall that interpretation of the coefficients on the IHS-transformed variables follows the standard level-log regression interpretation, thus a "*1% increase in x leads to a $\beta/100$ unit change in y".*

The coefficient estimates indicate that, on average, and holding all else constant, household income consistently exhibits a positive and statistically significant association with both WAZ and HAZ across multiple waves. For instance, in Wave 2, the coefficient on the log of household income for WAZ is 0.48 ($p<0.05$), increasing to 0.60 ($p<0.10$) by Wave 5. These results align with theoretical expectations that higher household income improves access to nutrition, healthcare, and other essential goods, thereby supporting better child health

outcomes. The slight decline in the magnitude of the income coefficient in later waves may suggest diminishing marginal returns to income, which warrants further investigation, particularly with non-linear specifications or interaction terms to capture potential threshold effects.

Expenditure on cigarettes and alcohol, while of primary interest, does not exhibit statistically significant effects on child nutrition outcomes. For instance, the coefficient for lagged cigarette expenditure per adult on WAZ in Wave 2 is -0.47 ($p>0.10$), with similar non-significant results in subsequent waves. Likewise, alcohol expenditure per adult remains consistently insignificant, with coefficients such as -4.34 ($p>0.10$) in Wave 3 and -0.67 ($p>0.10$) in Wave 5. These findings suggest that any potential resource diversion from essential needs due to spending on cigarettes or alcohol is not evident in the data. Possible explanations include measurement error in self-reported expenditures, compensatory behaviours within households, or the presence of informal support systems that mitigate the negative effects of such expenditures. These issues highlight the need for further research incorporating more granular data on household resource allocation and intra-household dynamics.

Child age demonstrates a weak but statistically significant negative association with WAZ in most waves. For instance, in Wave 2, the coefficient is -0.00 ($p<0.05$), and this association remains consistent in later waves. This result may reflect growth-related declines in weight-for-age z-scores or changes in nutritional requirements as children age. The relationship between age and HAZ is less consistent, with smaller coefficients and varying significance across waves. These findings underscore the importance of accounting for age-specific differences in nutrition and growth trajectories in policy interventions.

Gender effects are modest and inconsistent across waves. The coefficient for female children is generally positive but small, such as 0.06 ($p<0.05$) for WAZ in Wave 4, with no clear pattern across waves. These results suggest minimal gender differences in child nutrition outcomes within the sample but point to the potential role of unobserved factors, such as cultural practices or resource allocation biases, which merit further investigation.

.

Table 4.2: The 2SLS Estimation of the Effect of Lagged Tobacco and Alcohol Expenditure on Child Nutrition Outcomes

	Weight-for-Height z-score				Height-for-Age z-score			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep var	Wave 2	Wave 3	Wave 4	Wave 5	Wave 2	Wave 3	Wave 4	Wave 5
Lagged cigarette exp. per adult§	-0.59	-11.89	-1.53	0.61	-0.57	-0.95	1.55	0.75
	(1.49)	(201.99)	(59.92)	(1.47)	(5.50)	(2.05)	(9.83)	(4.11)
Lagged alcohol exp. per adult§	0.50	15.04	1.59	-0.71	0.95	0.78	-2.22	-1.00
	(2.02)	(269.17)	(73.21)	(1.22)	(9.02)	(2.48)	(12.72)	(3.64)
log of hh income	0.49**	-0.47	0.29	0.66*	0.45	0.33	0.53	0.64
	(0.20)	(18.36)	(9.06)	(0.35)	(0.36)	(0.27)	(1.06)	(1.00)
Age in months	-0.00**	-0.02	-0.00	-0.00	-0.00	-0.00*	0.00	0.00
	(0.00)	(0.27)	(0.06)	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)
Female	-0.04	-0.13	0.02	-0.00	0.03	0.03	0.05	0.01
	(0.08)	(3.22)	(2.01)	(0.10)	(0.07)	(0.06)	(0.21)	(0.27)
Coloured	-0.16	8.77	0.49	-0.81	0.37	0.83	-0.67	-0.72
	(1.96)	(152.27)	(24.58)	(1.68)	(7.06)	(2.09)	(4.10)	(4.51)
Asian/Indian	-0.68	14.78	-0.65	-1.12	0.48	1.79	-0.23	-0.48
	(1.64)	(259.61)	(17.44)	(1.62)	(6.89)	(2.75)	(2.27)	(2.77)
White	0.39	-7.57	-0.51	-1.34	-0.14	0.50	1.66	-0.02
	(1.68)	(148.92)	(19.97)	(4.23)	(1.44)	(0.82)	(8.11)	(5.04)
Had major illness	-0.08	1.61	0.13	-0.14	-0.17	0.06	-0.57	-0.15
	(0.34)	(29.36)	(14.63)	(0.15)	(0.67)	(0.24)	(1.82)	(0.19)
Household size	-0.06**	-0.25	-0.03	-0.09	-0.03	-0.04	-0.05	-0.08
	(0.03)	(3.46)	(1.18)	(0.06)	(0.11)	(0.03)	(0.09)	(0.12)
Water source: Borehole or tank	0.03	2.34	-0.50	0.08	0.16	-0.10	0.33	0.03
	(0.26)	(46.54)	(17.67)	(0.13)	(0.30)	(0.53)	(2.62)	(0.11)
Water source: Rain, spring or dam	0.15	2.55	-0.18	0.08	-0.02	-0.01	-0.14	-0.15
	(0.29)	(45.27)	(4.48)	(0.15)	(0.85)	(0.39)	(0.22)	(0.10)
Toilet type: Pit latrine or other	-0.00	-1.36	0.27	0.12	0.01	-0.08	-0.40	0.03
	(0.15)	(26.38)	(8.91)	(0.12)	(0.45)	(0.32)	(2.00)	(0.26)
Toilet type: None	0.07	-1.26	0.32	0.08	0.15	-0.30	-0.69	-0.07
	(0.19)	(21.13)	(19.22)	(0.19)	(0.15)	(0.24)	(2.72)	(0.18)
Urban	-0.03	-0.28	-0.04	0.11	0.03	-0.16*	-0.06	0.07
	(0.18)	(3.86)	(1.80)	(0.14)	(0.31)	(0.09)	(0.30)	(0.41)
Province fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	-2.84***	5.51	-1.55	-5.09**	-4.22***	-2.68	-4.93	-5.52
	(1.08)	(161.49)	(93.92)	(2.60)	(0.92)	(2.06)	(9.78)	(7.79)
Observations	2,423	3,314	3,433	3,646	3,806	4,812	5,316	5,572

§ Variable was transformed using the inverse hyperbolic sine transformation. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Ethnic group variables reveal significant heterogeneity in child nutrition outcomes. For instance, children identified as "White" have a large positive coefficient for WAZ in Wave 2 (4.19, $p<0.01$), though this effect diminishes in later waves. Conversely, the coefficients for the "Coloured" group vary in magnitude and direction, with values such as 1.06 ($p>0.10$) in Wave 2 and -1.60 ($p>0.10$) in Wave 3. These results suggest that ethnic disparities in nutrition outcomes are likely driven by structural and contextual factors, including differences in socioeconomic status, healthcare access, and local infrastructure. Future analyses should aim to disentangle these factors using more detailed data and methodological approaches

Household size negatively correlates with child nutrition outcomes, with a coefficient of -0.02 ($p<0.10$) for WAZ in Wave 4. This finding is consistent with the resource dilution hypothesis, where increasing household size reduces per capita access to resources, adversely affecting nutrition outcomes. Environmental factors, such as water and sanitation access, further underscore the importance of external determinants of nutrition. Reliance on rain or dam water sources, for example, is associated with poorer outcomes, as evidenced by a coefficient of -0.26 ($p<0.01$) for WAZ in Wave 2. These findings support the view that interventions aimed at improving household-level environmental conditions could yield significant gains in child health.

Regional fixed effects remain significant across all models, capturing unobserved heterogeneity in infrastructure, healthcare availability, and economic conditions. These results emphasise the need to control for spatial and contextual variations to accurately assess the determinants of child nutrition.

Diagnostics

To assess the endogeneity of the regressors, I employed the *endogtest* option of *ivreg2* in STATA. As previously mentioned, the lagged cigarette expenditure, lagged alcohol expenditure, and household income are, from a conceptual perspective, potentially endogenous. The *endogtest* option provides formal tests for endogeneity and produces test statistics that are robust to violations of conditional homoskedasticity. This test statistic is equivalent to the Hausman test statistic.

Table 4.3: Chi-square statistic from endogeneity tests

	Lagged cigarette expenditure		Lagged alcohol expenditure	
	WAZ	HAZ	WAZ	HAZ
Wave 2	21.35***	18.07***	24.43***	18.51***
Wave 3	30.91***	17.31***	25.11***	19.01***
Wave 4	15.23***	21.18***	12.56**	17.29***
Wave 5	22.67***	33.80***	22.70***	34.10***

The Chi-square statistic presented for the different dependent variables is a variant of the C-statistic, defined as the difference of two Sargan-Hansen statistics. In Stata, it is generated by the *endogtest ()* option of *ivreg2* and can report test statistics that are robust to various violations of conditional homoskedasticity. The null hypothesis that the endogenous regressor can be treated as exogenous is rejected if it is statistically significant, indicated as *** p<0.01, ** p<0.05 All categories reject this null hypothesis.

The null hypothesis under examination posits that the regressors can be treated as exogenous variables. By testing this hypothesis, we gain insight into the validity of treating these regressors as exogenous in our analysis.

Table 4.3 presents the results of the endogeneity test, indicating a high degree of statistical significance for all test statistics. Consequently, the null hypothesis that these regressors are exogenous is rejected, justifying the use of 2SLS regression methods. Employing OLS estimates would lead to biased outcomes, as seen in the results of the OLS regressions presented in Tables C4 and C5.

Ensuring instrument validity is a critical component of the 2SLS estimation approach, necessitating the use of instruments that are relevant to the endogenous variable. To assess instrument relevance, it is customary for researchers to conduct auxiliary regressions by regressing the endogenous variable on the instrumental variable, accompanied by a set of control variables. These control variables encompass factors such as income, age, gender, race, history of major illness, household size, type of toilet, type of water sources, and province dummies. The regression results are presented in Table 4.4.

The findings presented in Table 4.4 provide evidence of a robust positive association between previous cigarette expenditure and the adult-sex ratio, indicating that the inclusion criterion necessary for instrument validity has been met. Additionally, there is a consistent and statistically significant positive relationship between household income and the asset index across all waves.

Table 4.4: Auxiliary regression coefficients

	Lagged Cigarette Expenditure			
	(1)	(2)	(3)	(4)
VARIABLES	Wave 2	Wave 3	Wave 4	Wave 5
Adult-sex ratio	0.85***	0.53***	1.13***	1.01***
	(0.11)	(0.08)	(0.10)	(0.10)
Observations	6,555	7,578	7,901	9,288
R-squared	0.17	0.15	0.16	0.15
	Lagged Alcohol Expenditure			
	(1)	(2)	(3)	(4)
VARIABLES	Wave 2¥	Wave 3	Wave 4	Wave 5
Frequency of alcohol/drug abuse in community¥	-0.09***	0.02**	0.06***	0.07***
	(0.03)	(0.01)	(0.02)	(0.02)
Observations	6,535	7,461	7,831	9,246
R-squared	0.06	0.05	0.08	0.07
	Lagged household Income			
	(1)	(2)	(3)	(4)
VARIABLES	Wave 2	Wave 3	Wave 4	Wave 5
Log of asset index	0.31***	0.26***	0.26***	0.28***
	(0.01)	(0.01)	(0.01)	(0.01)
Observations	6,346	6,917	7,890	9,101
R-squared	0.34	0.26	0.26	0.28

¥ Note that since the endogenous variables are lagged, the corresponding lagged instrumental variables are used as well. However, the categorical variable "Frequency of drug/alcohol abuse in community" is not available in Wave 1 of NIDS, thus for Wave 2 the variable "Household level of religiosity" is used instead. This is a categorical variable that measures how important religion is among adults within a household on average. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

As anticipated, the relationship between the perceived frequency of alcohol or drug abuse in a community and household alcohol expenditure is both consistently positive and statistically significant. However, there is no variable measuring the frequency of alcohol or drug abuse in a community in Wave 1, and, in this case, the household average level of religiosity was used as a proxy in Wave 2. I argue that more religious households are less likely to consume alcohol, while less religious households are more likely to consume alcohol. As expected, this variable

also exhibits statistical significance, but with a negative association to household alcohol expenditure.

Table 4.5: First-stage F-statistics for endogenous lagged cigarette expenditure, alcohol expenditure, and household income

	Lagged Cigarette Expenditure		Income	
	WAZ	HAZ	WAZ	HAZ
Wave 2	24.79	32.18	146.47	220.05
Wave 3	7.52	8.83	149.87	233.4
Wave 4	26.49	33.18	187.38	303.5
Wave 5	10.22	21.48	263.88	399.75
	Lagged Alcohol Expenditure		Income	
	WAZ	HAZ	WAZ	HAZ
Wave 2	4.02	5.16	141.43	214.85
Wave 3	7.58	7.86	148.92	169.86
Wave 4	6.87	8.92	156.95	295.27
Wave 5	5.57	11.24	272.39	410.09

First-stage F-statistic less than 10 are considered weak instruments.

The final step in the econometric analysis involves obtaining the first-stage F-statistics, which serve as a crucial measure of instrument strength. Generally, recommendations in the literature suggest that an instrument with a first-stage F-statistic below 10 is considered weak (Andrews, Stock & Sun, 2019, Stock, James H., Wright & Yogo, 2002).

Table 4.5 provides additional diagnostic test results for the previously estimated instrumental variable (IV) regression models. The first-stage F-statistics for the regression of lagged cigarette expenditure on the adult-sex ratio are all above 10, except in Wave 3. On the other hand, most of the F-statistics for the alcohol expenditure instruments are below 10, except for HAZ in Wave 5. These findings suggest that the 2SLS regression results reported earlier may suffer from bias towards the OLS estimates, particularly for lagged cigarette expenditures in Wave 3, and lagged alcohol expenditures in all waves, except for Wave 5 HAZ. The significant result observed in Table 4.3 earlier was confined to this exception.

Furthermore, the F-statistic for the regression of the income variable consistently exceeds 10 in both the cigarette- and alcohol-related regressions. This indicates that the asset index variable serves as a strong instrument for income in these models. These diagnostic test results

have significant implications for the validity of the 2SLS regression estimates, which will receive further discussion later in this chapter.

Mediation analysis results

This section presents the mediation analysis results, which focus on examining the pathway through which the effect of cigarette and alcohol expenditures impact child health outcomes, as observed in the previous results. Specifically, this study investigates whether the effects of expenditure on food mediate the relationship between cigarette or alcohol expenditure and child health outcomes. Given the general lack of statistically significant results on the effects of alcohol expenditure, this analysis primarily aims to extend the results of cigarette expenditure on child health outcomes. However, mechanisms for the effect of both cigarette and alcohol expenditure are explored below.

I adopted the method proposed by Baron & Kenny (1986) for mediation analysis in this context. This method involves estimating three regression models. The first model (Model 1 or Equation 2) regresses child health outcomes on lagged cigarette (alcohol) expenditure and a set of controls. The second model (Model 2 or Equation 3) regresses lagged food expenditure on lagged cigarette alcohol) expenditure and the same set of controls as the first model. The third model (Model 3 or Equation 4) regresses child health outcomes on lagged cigarette (alcohol) expenditure, and lagged food expenditure, and a set of controls. The same set of controls are used in each model.

The first regression (Model 1) estimates the total effect of cigarette (alcohol) expenditure on child health outcomes. The second regression (Model 2) estimates the relationship between the mediator (food expenditure) and the main independent variable (cigarette or alcohol expenditure), while the third regression (Model 3) estimates the direct effect (cigarette or alcohol expenditure) on WAZ or HAZ. Furthermore, the results of Equation 9 were presented in Table 4.2 for lagged cigarette expenditure, and in Table 4.3 for lagged alcohol expenditure.

Table 4.6 presents the relationship between lagged food expenditure and lagged cigarette or alcohol expenditure, corresponding to the second step (i.e., Equation 3). This determines whether the effects of cigarette or alcohol expenditure on child health outcomes operate through expenditure on food.

The relationship between lagged cigarette expenditure per adult and food expenditure exhibits mixed results across waves, with the coefficients varying in sign and magnitude. While none

of the coefficients are statistically significant, the negative signs in Waves 2 and 3 align with the crowding-out hypothesis, which suggests that cigarette expenditures reduce food expenditures in resource-constrained households. These findings are consistent with previous studies which document significant reductions in food expenditure due to tobacco spending (Chelwa & Van Walbeek, 2014, Efroymson et al., 2001, Nyagwachi, Chelwa & van Walbeek, 2020).

However, the positive coefficients observed in Waves 4 and 5, though also statistically insignificant, deviate from this pattern. These inconsistencies might reflect heterogeneity in household budget allocations or contextual differences, such as variations in income or household composition across survey waves.

The relationship between lagged alcohol expenditure per adult and food expenditure is similarly inconsistent. The coefficients are positive in Waves 2 and 3 but turn negative in Waves 4 and 5. As with cigarette expenditure, these results are not statistically significant, except for Wave 3, where a marginal significance level (10%) suggests a slight positive association.

The lack of consistent significance in the alcohol results may reflect the variability in alcohol consumption patterns, which might be less uniformly detrimental to food expenditure compared to cigarette spending. Additionally, alcohol expenditures in some households might coexist with sufficient income levels, mitigating their crowding-out effect on food.

Across all models, the control variables provide additional insights into the determinants of food expenditure. Household income exhibits a strong positive association with food expenditure in Wave 5 (statistically significant at 1%), underscoring the critical role of income in household consumption decisions. This finding reinforces the notion that higher income levels can offset the adverse effects of discretionary spending on cigarettes or alcohol.

Conversely, household size is consistently negatively associated with per capita food expenditure and statistically significant in Waves 4 and 5. Larger households may face resource constraints, leading to reduced per capita allocations for food.

Other control variables, such as water and sanitation access, urban residence, and demographic characteristics, display less consistent patterns, with sporadic significance levels. Notably, the negative association between urban residence and food expenditure in Wave 2 (significant at 5%) may reflect higher costs of living or alternative consumption priorities in urban areas.

Table 4.6: Results of regression of lagged food expenditure regressed on lagged cigarette and alcohol expenditure

	Wave 2	Wave 3	Wave 4	Wave 5
Variables	(1)	(2)	(3)	(4)
Lagged cigarette exp. per adult§	-0.52	-3.77	2.01	0.19
	(8.97)	(6.34)	(3.25)	(0.30)
Lagged alcohol exp. per adult§	0.02	4.53	-2.84	-0.42
	(12.25)	(8.27)	(4.22)	(0.33)
log of hh income	0.53	0.30	0.97	0.55***
	(1.56)	(0.81)	(0.68)	(0.06)
Age in months	0.00	-0.00	-0.00	-0.00
	(0.01)	(0.00)	(0.00)	(0.00)
Female	-0.00	0.02	-0.04	-0.02
	(0.15)	(0.15)	(0.12)	(0.02)
Coloured	0.88	3.89	-0.42	0.08
	(11.67)	(6.31)	(1.25)	(0.30)
Asian/Indian	1.55	4.60	-0.92	-0.10
	(14.91)	(7.21)	(2.48)	(0.38)
White	1.49	-0.54	1.20	0.47
	(1.86)	(3.25)	(1.49)	(0.34)
Had major illness	0.05	0.35	-0.26	0.01
	(1.30)	(0.75)	(0.41)	(0.06)
Household size	-0.11	-0.17	-0.09***	-0.09***
	(0.27)	(0.15)	(0.02)	(0.01)
Water source: Borehole or tank	-0.02	0.82	0.68	-0.06*
	(1.02)	(1.38)	(1.03)	(0.04)
Water source: Rain, spring or dam	-0.14	0.57	-0.05	-0.03
	(0.14)	(0.85)	(0.17)	(0.05)
Toilet type: Pit latrine or other	-0.05	-0.50	-0.44	0.02
	(0.58)	(0.98)	(0.73)	(0.03)
Toilet type: None	0.06	-0.57	-0.55	-0.05
	(0.51)	(1.31)	(0.87)	(0.06)
Urban	-0.09**	-0.08	-0.05	0.01
	(0.05)	(0.37)	(0.18)	(0.04)
Province fixed effects	Yes	Yes	Yes	Yes
Constant	2.53	3.88	-1.59	2.15***
	(6.81)	(6.26)	(5.80)	(0.42)
Observations	6,237	6,797	7,658	9,008

§ Variable was transformed using the inverse hyperbolic sine transformation. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

The mixed and largely insignificant findings in Table 4.6 suggest that while there may be some evidence of a crowding-out effect, the mediating role of food expenditure in the relationship between cigarette or alcohol expenditures and child health outcomes is not robustly established. This underscores the importance of considering other potential mediators or contextual factors, such as health care access or broader socioeconomic conditions, in future analyses.

The observed inconsistencies may also indicate that the effect of cigarette and alcohol expenditures on child health outcomes is more direct than mediated through food expenditure. Alternatively, the weak mediation results might stem from measurement errors, unobserved heterogeneity, or insufficient statistical power in certain waves.

Table 4.7: The mediated effect of cigarette and alcohol expenditure by food expenditure on child nutrition outcomes

	Wave 2	Wave 3	Wave 4	Wave 5	Wave 2	Wave 3	Wave 4	Wave 5
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variable	WAZ	WAZ	WAZ	WAZ	HAZ	HAZ	HAZ	HAZ
Total Effect: Cigarette expenditure	-0.59	-11.89	-1.53	0.61	-0.57	-0.95	1.55	0.75
	(1.49)	(201.99)	(59.92)	(1.47)	(5.50)	(2.05)	(9.83)	(4.11)
Total Effect: Alcohol expenditure	0.50	15.04	1.59	-0.71	0.95	0.78	-2.22	-1.00
	(2.02)	(269.17)	(73.21)	(1.22)	(9.02)	(2.48)	(12.72)	(3.64)
Direct Effect	-0.19	-0.67***	-0.25**	-0.28**	0.00	-0.33	-0.21**	-0.36***
	(0.17)	(0.24)	(0.11)	(0.14)	(0.14)	(0.20)	(0.09)	(0.10)
A. Lagged cigarette exp. per adult§	-0.52	-3.77	2.01	0.19	-0.52	-3.77	2.01	0.19
	(8.97)	(6.34)	(3.25)	(0.29)	(8.97)	(6.34)	(3.25)	(0.29)
B. Lagged alcohol exp. per adult§	0.02	4.53	-2.84	-0.42	0.02	4.53	-2.84	-0.42
	(12.25)	(8.27)	4.22	(0.33)	(12.25)	(8.27)	4.22	(0.33)
C. Per capita hh food exp.§	0.06	-0.06	-0.02	0.09**	-0.02	0.00	-0.01	0.04
	(0.07)	(0.06)	(0.05)	(0.04)	(0.05)	(0.03)	(0.04)	(0.04)
Bootstrap results								
Cigarette indirect effect [A x C]	-0.12	-15.57	-1.14	0.26	0.14	0.39	0.84	-0.29
	(41.79)	(387.38)	(21.84)	(18.79)	(29.44)	(29.72)	(19.50)	(146.44)
z-Statistic	-0.00	-0.04	-0.05	0.00	0.00	0.01	0.04	-0.00
P-value	0.99	0.97	0.96	0.99	0.99	0.99	0.97	0.99
Alcohol indirect effect [B x C]	0.13	20.58	1.39	-0.04	-0.19	-0.45	-1.11	0.23
	(53.92)	(511.84)	(29.01)	(17.56)	(44.94)	(39.78)	(25.17)	(116.70)
z-Statistic	0.00	0.04	0.05	-0.00	-0.00	-0.01	-0.04	0.00
P-value	0.99	0.97	0.96	0.99	0.99	0.99	0.96	0.99
Observations	2,423	3,313	3,433	3,646	3,806	4,810	5,316	5,572

§ Variable was transformed using the inverse hyperbolic sine transformation. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

The mediation results presented in Table 4.7 evaluate the indirect effect of cigarette and alcohol expenditures on child nutrition outcomes (measured by weight-for-age z-scores (WAZ) and

height-for-age z-scores (HAZ)) through their influence on household food expenditure. These findings provide insights into whether food expenditure mediates the relationship between discretionary spending on harmful products and child nutrition.

Total effects of cigarette and alcohol expenditures

The total effects of cigarette and alcohol expenditures on child nutrition outcomes display variability across waves. For cigarette expenditures, the total effects are negative in most waves, with substantial magnitudes in Waves 2 and 3 for WAZ and Wave 3 for HAZ. However, none of these total effects are statistically significant, indicating weak evidence of a direct or overall relationship between cigarette expenditures and child nutrition outcomes.

Alcohol expenditures, in contrast, show inconsistent effects. The total effect coefficients are positive in most cases but lack statistical significance, suggesting limited evidence of an association between alcohol expenditures and child nutrition. The variability in signs and magnitudes across waves might reflect household heterogeneity in spending priorities, resource allocation, or measurement error.

Direct effects of cigarette and alcohol expenditures

The direct effects, which measure the impact of cigarette and alcohol expenditures on child nutrition outcomes net of mediation by food expenditure, are more consistent and statistically significant in certain cases. For cigarette expenditures, significant negative direct effects emerge for WAZ in Waves 3, 4, and 5 and for HAZ in Waves 4 and 5, suggesting that cigarette spending directly compromises child nutrition. This finding aligns with previous literature emphasising the detrimental effects of tobacco expenditures on household welfare and child health.

Similarly, alcohol expenditures exhibit negative direct effects for HAZ in Waves 4 and 5, though these are less consistent across waves. These results indicate that alcohol expenditures may also directly impact long-term child growth indicators, likely reflecting resource constraints or competing household priorities.

Mediated (indirect) effects through food expenditure

The indirect effect is calculated as the product of the coefficient on cigarette expenditure from Equation 3 and the coefficient on food expenditure from Equation 4. This is calculated

manually in Stata, and subsequently bootstrapped using the *bootstrap* Stata command, employing 1000 replications.[23]

The mediated effects, represented by the indirect effect coefficients, measure the extent to which cigarette and alcohol expenditures affect child nutrition outcomes through changes in per capita household food expenditure. Across all waves, these coefficients are statistically insignificant, with p-values exceeding 0.95.

These findings suggest that changes in household food expenditure—while theoretically a plausible mediator—do not account for the observed relationships between cigarette or alcohol expenditures and child nutrition outcomes. There are several potential explanations for this. First, the weak or inconsistent associations between cigarette and alcohol expenditures and food expenditure, as observed in Table 4.6, likely diminish the potential for food expenditure to act as a substantive mediator. Households spending on cigarettes or alcohol may not always reduce food expenditure in ways that directly compromise nutritional outcomes. For example, some households might reduce spending on other non-essential goods, rather than food, to accommodate these discretionary expenditures.

Second, the relative stability of household food expenditure, even amidst varying levels of cigarette or alcohol spending, may reflect household resource allocation strategies. Many households prioritise food expenditure as a critical need, ensuring that it is preserved at the expense of other discretionary spending categories. This prioritization might reduce the observable impact of cigarette and alcohol expenditures on food budgets. Moreover, the responsiveness of food expenditure to discretionary spending might depend on factors such as household income levels, cultural norms, and access to supplementary food resources, such as food aid programs or community support, which are not explicitly accounted for in this analysis.

Third, child nutrition outcomes, particularly long-term indicators such as height-for-age z-scores (HAZ), are influenced by a complex array of factors beyond immediate food expenditure. While adequate food consumption is necessary for optimal nutrition, other determinants—such as the quality and diversity of diet, access to healthcare, parental

[23] In Stata, I developed a short programme that manually calculates the indirect effect in three steps: first, I estimate Equation 10 via 2SLS and store the coefficient on cigarette expenditure; second, I estimate Equation 11 and store the coefficient on food expenditure; third, I multiply the two stored coefficients to estimate the indirect effect. Bootstrapping replicated this process 1000 times to obtain the standard errors, z-statistics, and p-values in Table 9.

education, and broader environmental conditions—may play a more critical role. It is plausible that cigarette and alcohol expenditures impact child nutrition outcomes through these alternative channels, which remain unexamined in this mediation framework. For example, discretionary spending on cigarettes or alcohol might deplete household resources for healthcare or educational investments, both of which have been shown to significantly impact child development outcomes..

4.8 Discussion

The findings presented in this chapter provide an analysis of the impact of household expenditures on cigarettes and alcohol on child nutrition outcomes, specifically WAZ and HAZ. The results suggest a complex interplay between discretionary spending on harmful products, household resource allocation, and child health. While direct effects of cigarette and alcohol expenditures on child nutrition outcomes are observed in some cases, the hypothesised mediating role of food expenditure is not supported by the data.

Total effects of cigarette and alcohol expenditures

The total effects of cigarette and alcohol expenditures on child nutrition outcomes, as evaluated across waves, reveal no consistent or statistically significant associations. Although the coefficients for cigarette expenditures are predominantly negative and those for alcohol expenditures are mostly positive, their lack of statistical significance implies that these expenditures do not have a straightforward or strong overall relationship with child nutrition outcomes.

The variability in the signs and magnitudes of total effects may reflect heterogeneity in household circumstances, such as differences in income levels, spending priorities, and coping strategies. It is also possible that measurement error in expenditure data or unobserved confounders influences these relationships, masking any true effects.

Direct effects of cigarette and alcohol expenditures

The analysis of direct effects reveals significant negative associations between cigarette expenditures and child nutrition outcomes, particularly for WAZ and HAZ in later waves. These findings suggest that cigarette spending directly compromises household resources critical to supporting child health, a result consistent with literature emphasising the adverse impacts of tobacco use on household welfare.

For alcohol expenditures, the evidence of negative direct effects is less consistent but emerges for HAZ in Waves 4 and 5. These results may indicate that alcohol spending undermines long-term child growth, potentially through resource constraints or prioritization of alcohol over essential goods and services. Together, these findings support the hypothesis that discretionary spending on harmful products can adversely affect child nutrition outcomes through mechanisms other than food expenditure.

Mediated effects through food expenditure

The mediation analysis finds no statistically significant indirect effects of cigarette or alcohol expenditures on child nutrition outcomes through food expenditure. This result suggests that food expenditure does not mediate the relationship between discretionary spending on harmful products and child nutrition.

Several plausible explanations exist for the lack of significant mediation. First, as shown in Table 4.6, the weak or inconsistent associations between cigarette and alcohol expenditures and food expenditure limit the potential for food expenditure to act as a mediator. Households may adjust spending on non-essential goods, rather than food, to accommodate discretionary expenditures, thereby preserving food budgets.

Second, the stability of food expenditure across varying levels of cigarette and alcohol spending may reflect its prioritization within household budgets. Food, being a necessity, is likely shielded from resource reallocations in many households, especially those with stronger financial management or access to supplementary resources, such as food aid or community support.

Third, child nutrition outcomes are determined by a broader array of factors beyond food expenditure, including dietary quality and diversity, access to healthcare, and parental education. The current mediation framework does not account for these additional pathways, which may play a more significant role in mediating the relationship between discretionary spending and child nutrition outcomes.

The findings highlight the direct burden imposed by cigarette and alcohol expenditures on child nutrition outcomes, underscoring the need for interventions targeting these spending behaviours. Policies aimed at reducing the affordability and consumption of harmful products, such as excise tax increases, can play a critical role in protecting household resources for essential goods and services. Additionally, public health campaigns to raise awareness about

the opportunity costs of cigarette and alcohol expenditures may encourage more optimal resource allocation within households.

At the same time, targeted support programs, such as conditional cash transfers or nutrition-specific interventions, could help buffer vulnerable households from the negative impacts of discretionary spending on child health. Policymakers should also consider integrating these measures into broader strategies addressing social determinants of child health, including healthcare access, parental education, and income support.

Limitations

Despite the chapter's identification of the adverse effects of household cigarette expenditure on child health outcomes, it is important to acknowledge several limitations in the research that warrant consideration. An ongoing challenge lies in the endogeneity of cigarette expenditure, which persists even when instrumental variables are utilised in the models. Previous studies have employed various approaches, such as non-parametric techniques, or the inclusion of child-fixed effects, to address endogeneity concerns. In this chapter, an instrumental variable approach was employed to account for the endogenous nature of cigarette or alcohol expenditure and household income. Diagnostic tests indicated that the adult-sex ratio generally serves as a robust instrument for cigarette expenditure, as evidenced by first-stage F-statistics exceeding 10. In contrast, the first-stage F-statistics for the frequency of alcohol or drug abuse in a community, used as an instrument for alcohol expenditure, fell below 10. This probably contributed to the absence of statistically significant results in the analysis of alcohol expenditure.

A second limitation relates to the use of the IHS transformation on the primary independent variables. The IHS transformed variables can be sensitive to units of measurement (Aihounton & Henningsen, 2021, Bellemare & Wichman, 2019). For instance, the arbitrary choice of measuring cigarette expenditure on a monthly or yearly basis may lead to substantial disparities in policy recommendations. However, Bellemare & Wichman (2019) highlight that elasticities calculated from IHS-transformed variables exhibit greater precision for values greater than 10. In this study, the average monthly expenditures on tobacco and alcohol, as depicted in the summary statistics, both satisfy this criterion. Consequently, the scaling of the IHS-transformed variables does not represent a major concern in this specific context.

Lastly, one of the key limitations of this study is the potential misrepresentation of household expenditure in the NIDS data, particularly regarding alcohol, tobacco, and food expenditures.

The NIDS data reports lower-than-expected expenditure on these items, with alcohol and tobacco expenditure appearing too small, and food expenditure seemingly implausibly low when compared to other national datasets, such as the IES from 2010-11. For instance, the IES data indicates that tobacco and alcohol expenditure should account for approximately 1.1% and 0.9% of household budgets, respectively, while food expenditure constitutes about 25.2%. If we assume a monthly household budget of ZAR7,700, this translates to ZAR70 in tobacco, ZAR77 in alcohol, and ZAR1,940 in food. In contrast, the NIDS data reports far lower values, suggesting a potential underreporting or misclassification of household expenditure.

This discrepancy raises concerns about the validity of the NIDS data for capturing actual household expenditure patterns. Such underreporting may lead to biased estimates of the relationship between discretionary spending on harmful products (such as alcohol and tobacco) and child nutrition outcomes. The limitations of NIDS in this regard suggest that future studies should consider supplementing this data with more reliable sources or recalibrating expenditure estimates to better reflect actual spending patterns.

4.9 Conclusion

This chapter provides evidence of the adverse direct effects of cigarette and alcohol expenditures on child nutrition outcomes, specifically weight-for-age z-scores (WAZ) and height-for-age z-scores (HAZ). While the hypothesised mediating role of food expenditure was not supported, the findings underscore the significant negative impact of cigarette expenditures on child health, highlighting the broader consequences of household spending on harmful products.

Malnutrition remains a critical public health challenge globally, contributing to approximately 45% of deaths among children under the age of five. Effective policy responses require a clear understanding of the multifaceted causes of malnutrition. Although maternal tobacco and alcohol use has long been recognised as detrimental to birth outcomes and infant health, the findings in this chapter extend this understanding by demonstrating the harmful effects of cigarette expenditure on the nutritional outcomes of older children.

The absence of a direct link between reduced food expenditure and the observed effects suggests that the adverse outcomes operate through alternative pathways. Potential mechanisms may include reduced household spending on healthcare, education, or other essential services. Identifying these pathways is crucial for tailoring interventions to mitigate the negative effects of cigarette expenditures on child nutrition.

In contrast, the analysis did not find consistent or significant effects of alcohol expenditures on child nutrition outcomes. This result may be attributable to data limitations, household heterogeneity, or other unmeasured factors. Future research should revisit this issue with updated and more granular data to better understand the relationship between alcohol expenditures and child health.

The findings of this chapter have important policy implications. Strengthening tobacco-control measures to reduce household cigarette expenditure could free up resources for critical needs, mitigating the observed negative effects on child nutrition. Similarly, addressing alcohol consumption in households may have broader benefits for resource allocation and well-being, even if its direct impact on child nutrition remains inconclusive.

Addressing child malnutrition requires an interdisciplinary approach. Economists contribute insights into household resource allocation and the opportunity costs of discretionary spending, while nutritionists provide an understanding of the direct determinants of child health. Collaboration between these fields can enable the development of holistic strategies to combat malnutrition effectively.

In summary, this chapter emphasises the need for targeted policies to reduce household spending on harmful products, such as tobacco and alcohol, as part of a broader strategy to improve child nutrition outcomes. Future research should explore alternative mechanisms to inform more comprehensive policy responses, ultimately contributing to the global fight against child malnutrition.

4.10 Annex C

C1. Construction of asset index

Asset indices are often constructed using principal components analysis (PCA), a method popularised by Filmer & Pritchett (2001). They recommended the use of several binary asset ownership variables to construct the index, based on the PCA first principal component. The underlying intuition is that, for $\boldsymbol{k}$ indicators of ownership, there is a latent variable $(\tilde{\boldsymbol{a}}_1, \tilde{a}_2, \dots, \tilde{a}_k)$ that manifests in ownership of various assets $(a_1, a_2, \dots, a_k)$. This can be summarised into a single index using the formula:

$$Asset_index = w_1 a_1 + w_2 a_2 + \cdots + w_k a_k \quad (8)$$

where w_1 indicates the weights associated with each indicator. The maximum likelihood estimators of the *w*s are the eigenvectors of the covariance matrix, also known as the principal component of the dataset. It is typically the case that only the eigenvector with the highest eigenvalue is used, as it provides the most variation about the variables. The first eigenvector is the vector that minimises the squared distances from the observations to a line going through the various dimensions.

One limitation of the above method is that certain assets may be assigned negative weights, for instance livestock, which is commonly owned by rural households. This may be misleading, resulting in the ranking of rural households with livestock lower than households with no assets at all (Wittenberg & Leibbrandt, 2017). To address this and other limitations (see Wittenberg & Leibbrandt (2017) and McKenzie (2005) for detailed critiques), a second method, proposed by Wittenberg & Leibbrandt (2017) and originally developed by Banerjee (2010), is adopted for calculating the asset index, that is, the uncentered PCA (UC PCA).

The asset indices calculated in this case use 26 ownership variables available in all five waves of the NIDS dataset. The variables are all dummy variables with 1 representing ownership of the asset, and 0 non-ownership. Table C1 below provides the summary statistics of the 26 variables across all waves.

Table C 1: Descriptive statistics for household assets used to construct asset index

Household Asset	Wave 1		Wave 2		Wave 3		Wave 4		Wave 5	
	Obs	Mean	Obs	Mean	Obs	Mean	Obs	Mean	Obs	Mean
Radio	28,163	0.676	28,255	0.626	33,199	0.705	37,984	0.626	40,860	0.6
Hi-fi stereo	28,185	0.42	28,235	0.333	33,176	0.423	37,975	0.386	40,858	0.331
TV	28,188	0.663	28,238	0.714	33,171	0.79	37,998	0.801	40,883	0.83
Satellite dish	28,104	0.088	28,200	0.127	33,199	0.204	37,974	0.295	40,864	0.41
DVD player	28,155	0.316	28,240	0.362	33,195	0.444	37,963	0.351	40,860	0.328
Computer	28,129	0.099	28,200	0.083	33,190	0.124	37,967	0.152	40,818	0.18
Camera	28,075	0.126	28,213	0.063	33,176	0.089	37,971	0.084	40,857	0.087
Cell phone	28,162	0.775	28,239	0.79	33,201	0.888	38,006	0.895	40,899	0.917
Electric stove	28,156	0.601	28,240	0.665	33,196	0.753	38,003	0.815	40,902	0.842
Gas stove	28,120	0.167	28,209	0.119	33,190	0.132	37,988	0.155	40,819	0.194
Paraffin stove	28,081	0.288	28,207	0.255	33,195	0.193	37,969	0.179	40,797	0.173
Microwave	28,106	0.319	28,244	0.335	33,191	0.415	37,999	0.469	40,851	0.534
Fridge/freezer	28,175	0.584	28,252	0.631	33,203	0.715	38,006	0.763	40,872	0.818
Washing mach.	28,137	0.233	28,247	0.236	33,179	0.275	37,975	0.324	40,876	0.377
Sew/Knit mach.	28,133	0.12	28,184	0.066	33,186	0.07	37,986	0.064	40,752	0.077
Lounge suite	28,141	0.397	28,232	0.465	33,200	0.532	37,986	0.549	40,883	0.616
Private car	28,178	0.172	28,228	0.12	33,172	0.145	37,938	0.17	40,864	0.216
Motorcycle	28,157	0.019	28,248	0.01	33,180	0.011	37,971	0.012	40,886	0.018
Bicycle	28,157	0.083	28,243	0.064	33,162	0.081	37,957	0.081	40,836	0.102
Boat	28,169	0.006	28,226	0.003	33,176	0.003	37,984	0.004	40,815	0.003
Motorboat	28,146	0.004	28,230	0.005	33,188	0.003	37,955	0.005	40,840	0.004
Donkey cart	28,158	0.011	28,241	0.009	33,183	0.009	37,984	0.011	40,790	0.007
Plough	28,175	0.073	28,243	0.024	33,176	0.015	37,979	0.024	40,849	0.029
Tractor	28,153	0.017	28,250	0.012	33,170	0.009	37,986	0.012	40,833	0.01
Wheelbarrow	28,162	0.235	28,237	0.25	33,167	0.3	37,963	0.312	40,866	0.375
Grinding mill	28,131	0.012	28,220	0.012	33,172	0.009	37,904	0.015	40,795	0.02

Table C 2: Results for the Kaiser-Meyer-Olkin test of sampling adequacy

Household asset	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5
Radio	0.9098	0.9026	0.8719	0.8808	0.8672
Hi-fi stereo	0.9469	0.9494	0.9210	0.9229	0.9234
TV	0.9090	0.8942	0.8815	0.8866	0.8699
Satellite dish	0.9278	0.9241	0.9283	0.9296	0.9295
DVD player	0.9518	0.9446	0.9328	0.9370	0.9365
Computer	0.9030	0.8802	0.8891	0.8864	0.8867
Camera	0.9026	0.8605	0.8771	0.8838	0.8750
Cell phone	0.9261	0.9009	0.9158	0.8988	0.8926
Electric stove	0.9019	0.8974	0.8724	0.8610	0.8669
Gas stove	0.9069	0.8496	0.8151	0.8695	0.8975
Paraffin stove	0.7755	0.7702	0.8691	0.7754	0.8479
Microwave	0.9417	0.9314	0.9184	0.9146	0.9159
Fridge/freezer	0.9100	0.8977	0.8900	0.8826	0.8708
Washing machine	0.9299	0.9231	0.9171	0.9174	0.9190
Sewing/knitting machine	0.9393	0.9223	0.9168	0.9205	0.9130
Lounge suite	0.9487	0.9440	0.9363	0.9364	0.9367
Private car	0.9460	0.9168	0.9169	0.9052	0.9106
Motorcycle	0.8820	0.7754	0.8308	0.8497	0.8790
Bicycle	0.9340	0.8704	0.8912	0.8933	0.9043
Boat	0.6733	0.7257	0.6749	0.6709	0.6505
Motorboat	0.7139	0.6912	0.6568	0.6753	0.6741
Donkey cart	0.8070	0.6698	0.7295	0.6600	0.5218
Plough	0.7253	0.6243	0.6311	0.6174	0.5796
Tractor	0.6391	0.6538	0.6705	0.6352	0.5681
Wheelbarrow	0.7658	0.7339	0.8036	0.7574	0.6965
Grinding mill	0.6316	0.6665	0.7619	0.7228	0.6534
Overall	0.9003	0.8978	0.8869	0.8867	0.8913

The Kaiser rule-of-thumb is used to retain components in the analysis with eigenvalues over one, while those below one are dropped (Schreiber, 2021). In addition, to determine the suitability of this data for PCA, the Kaiser-Meyer-Olkin (KMO) measure of adequacy is calculated. The test measures the sampling adequacy of each variable in the model, as well as for the complete model. Values for this statistic are bounded between 0 and 1 (Schreiber, 2021, Shrestha, 2021). KMO values above 0.8 are indicative of an adequate sample, those between 0.7 and 0.79 are considered middle values, while values below 0.6 are considered inadequate. Table C2 above shows a summary of these values and indicates overall KMO values above 0.8 across all waves.

The coefficient estimates on the asset index variables are given in Table C3 below. As can be seen, several variables using PCA have negative coefficients. Specifically, these are ownership of a paraffin stove, donkey cart, and plough. In contrast all UC PCA values are positive.

Table C 3: Coefficient estimates on asset variables

	Wave 1		Wave 2		Wave 3		Wave 4		Wave 5	
Asset	PCA	UC_PCA	PCA	UC_PCA	PCA	UC_PCA	PCA	UC_PCA	PCA	UC_PCA
Radio	0.221	0.012	0.264	0.010	0.379	0.010	0.339	0.014	0.274	0.007
Hi-fi stereo	0.512	0.035	0.495	0.024	0.464	0.022	0.442	0.029	0.476	0.021
TV	0.515	0.016	0.581	0.008	0.639	0.008	0.633	0.010	0.667	0.005
Satellite dish	0.749	0.468	0.704	0.118	0.644	0.074	0.609	0.055	0.604	0.015
DVD player	0.592	0.060	0.497	0.016	0.453	0.017	0.454	0.033	0.459	0.014
Computer	0.789	0.450	0.781	0.233	0.709	0.202	0.650	0.190	0.659	0.082
Camera	0.637	0.298	0.810	0.414	0.757	0.421	0.737	0.466	0.779	0.265
Cellphone	0.359	0.012	0.380	0.007	0.428	0.006	0.443	0.008	0.400	0.005
Electric stove	0.474	0.019	0.524	0.010	0.551	0.008	0.522	0.010	0.498	0.005
Gas stove	0.288	0.109	0.300	0.115	0.353	0.157	0.359	0.111	0.395	0.058
Paraffin stove	-0.161	0.008	-0.273	0.022	-0.288	0.021	-0.166	0.046	-0.246	0.023
Microwave	0.618	0.059	0.623	0.035	0.601	0.024	0.612	0.027	0.606	0.011
Fridge/freezer	0.553	0.018	0.585	0.010	0.628	0.009	0.645	0.011	0.698	0.006
Washing machine	0.682	0.106	0.653	0.064	0.650	0.053	0.631	0.051	0.648	0.023
Sewing/knitting mach	0.542	0.236	0.560	0.355	0.585	0.215	0.654	0.816	0.671	0.201
Lounge suite	0.538	0.038	0.536	0.018	0.519	0.015	0.521	0.018	0.549	0.009
Private car	0.704	0.192	0.742	0.171	0.686	0.167	0.696	0.169	0.684	0.068
Motorcycle	0.857	11.820	0.791	4.586	0.934	20.432	0.868	13.010	0.775	4.288
Bicycle	0.583	0.451	0.546	0.595	0.567	0.316	0.626	0.429	0.612	0.181
Boat	0.545	57.750	0.617	284.835	0.580	182.905	0.818	64.370	0.779	314.297
Motorboat	1.050	201.135	0.777	65.185	0.822	240.349	0.814	200.678	0.730	73.024
Donkey cart	0.073	3.596	-0.189	0.228	-0.015	6.087	0.036	2.543	-0.104	3.281
Plough	-0.112	0.157	0.035	1.563	0.050	2.343	0.129	1.029	-0.065	0.175
Tractor	0.179	1.943	0.200	1.391	0.257	8.512	0.492	1.656	0.100	4.796
Wheelbarrow	0.207	0.077	0.125	0.026	0.235	0.038	0.200	0.029	0.181	0.015
Grinding mill	0.284	2.707	0.110	0.442	0.514	6.777	0.491	1.275	0.294	1.313
Cons_	-3.204	0.000	-3.491	0.000	-4.144	0.000	-4.247	0.000	-4.412	0.000

Table C 4: OLS regression results on the impact of lagged cigarette and alcohol expenditure on child health outcomes

	Wave 2	Wave 3	Wave 4	Wave 5	Wave 2	Wave 3	Wave 4	Wave 5
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep var	WAZ	WAZ	WAZ	WAZ	HAZ	HAZ	HAZ	HAZ
Lagged cigarette exp. per adult§	0.02	0.01	-0.01	-0.02*	0.03	-0.01	-0.01	-0.01
	(0.02)	(0.016)	(0.013)	(0.011)	(0.02)	(0.014)	(0.009)	(0.008)
Lagged alcohol exp. per adult§	-0.01				-0.01			
	(0.03)				(0.03)			
Age in months	-0.01***	-0.00**	-0.00	-0.00	-0.00	-0.00**	0.00***	0.00***
	(0.00)	(0.001)	(0.001)	(0.001)	(0.00)	(0.000)	(0.000)	(0.000)
Female	0.07*	0.11**	0.06	0.03	0.17***	0.03	0.08***	0.08***
	(0.04)	(0.045)	(0.038)	(0.037)	(0.03)	(0.038)	(0.029)	(0.029)
Coloured	-0.36***	-0.47***	-0.34***	-0.16*	-0.25***	-0.10	-0.15**	-0.08
	(0.11)	(0.098)	(0.082)	(0.085)	(0.09)	(0.079)	(0.062)	(0.060)
Asian/Indian	-0.93***	-0.07	0.02	-0.13	-0.23	0.77***	0.33*	0.23
	(0.23)	(0.245)	(0.217)	(0.271)	(0.19)	(0.164)	(0.188)	(0.173)
White	0.25	0.02	0.13	0.66**	0.40**	0.53**	0.42**	0.77***
	(0.18)	(0.228)	(0.249)	(0.274)	(0.16)	(0.210)	(0.187)	(0.254)
Had major illness	-0.22**	-0.13	-0.16	-0.10	-0.14*	-0.09	-0.22***	-0.18**
	(0.09)	(0.118)	(0.103)	(0.092)	(0.08)	(0.100)	(0.078)	(0.080)
log of hh income	0.12***	0.14***	0.18***	0.18***	0.10***	0.08***	0.10***	0.16***
	(0.03)	(0.026)	(0.025)	(0.026)	(0.03)	(0.022)	(0.018)	(0.019)
Household size	-0.06***	-0.03***	-0.04***	-0.05***	-0.03***	-0.02***	-0.03***	-0.04***
	(0.01)	(0.007)	(0.005)	(0.005)	(0.01)	(0.006)	(0.004)	(0.004)
Water source: Borehole or tank	0.29**	-0.28**	-0.13	0.11	-0.09	-0.26***	-0.13*	0.04
	(0.12)	(0.110)	(0.081)	(0.076)	(0.10)	(0.099)	(0.065)	(0.063)
Water source: Rain, spring or dam	-0.20***	-0.05	-0.07	0.12*	-0.16***	-0.19***	-0.17***	-0.13**
	(0.06)	(0.078)	(0.062)	(0.066)	(0.05)	(0.071)	(0.051)	(0.056)
Toilet type: Pit latrine or other	-0.10	0.06	-0.04	-0.01	-0.02	-0.06	-0.14***	-0.12**
	(0.07)	(0.066)	(0.059)	(0.059)	(0.06)	(0.055)	(0.049)	(0.046)
Toilet type: None	-0.32***	-0.17	-0.25**	-0.18*	-0.22***	-0.29***	-0.34***	-0.28***
	(0.09)	(0.112)	(0.104)	(0.101)	(0.08)	(0.095)	(0.074)	(0.086)
Urban	-0.10	-0.02	-0.02	0.03	-0.13**	-0.13**	-0.10**	-0.08*
	(0.07)	(0.072)	(0.060)	(0.058)	(0.06)	(0.060)	(0.047)	(0.046)
Province fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	-0.71***	-0.83***	-1.27***	-1.55***	-1.54***	-1.08***	-1.42***	-2.01***
	(0.26)	(0.243)	(0.236)	(0.243)	(0.22)	(0.201)	(0.171)	(0.183)

Observations	4,591	3,701	3,555	3,787	6,586	5,378	5,487	5,771
R-squared	0.09	0.049	0.075	0.069	0.03	0.021	0.038	0.050

§ Variable was transformed using the inverse hyperbolic sine transformation. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table C 5: OLS regression results on the impact of lagged alcohol expenditure on child health outcomes

	Wave 2	Wave 3	Wave 4	Wave 5	Wave 2	Wave 3	Wave 4	Wave 5
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep var	WAZ	WAZ	WAZ	WAZ	HAZ	HAZ	HAZ	HAZ
Lagged alcohol exp. per adult§	-0.01	-0.00	-0.03***	-0.03***	0.01	-0.03*	-0.03***	-0.02***
	(0.021)	(0.017)	(0.011)	(0.010)	(0.017)	(0.015)	(0.008)	(0.007)
Age in months	-0.00***	-0.00**	-0.00	-0.00	-0.00***	-0.00**	0.00***	0.00***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)
Female	-0.04	0.11**	0.06	0.03	0.02	0.03	0.08***	0.08***
	(0.056)	(0.045)	(0.038)	(0.037)	(0.045)	(0.038)	(0.029)	(0.029)
Coloured	-0.84***	-0.46***	-0.33***	-0.17**	-0.26***	-0.12	-0.15**	-0.08
	(0.143)	(0.094)	(0.080)	(0.083)	(0.095)	(0.076)	(0.061)	(0.059)
Asian/Indian	-0.84***	-0.06	-0.05	-0.15	0.17	0.76***	0.30*	0.23
	(0.290)	(0.243)	(0.204)	(0.270)	(0.196)	(0.164)	(0.172)	(0.172)
White	0.32	0.04	0.17	0.64**	0.34	0.54**	0.46**	0.79***
	(0.311)	(0.230)	(0.250)	(0.268)	(0.222)	(0.209)	(0.187)	(0.251)
Had major illness	-0.18	-0.13	-0.19*	-0.11	-0.06	-0.09	-0.24***	-0.18**
	(0.185)	(0.118)	(0.107)	(0.092)	(0.116)	(0.100)	(0.080)	(0.079)
log of hh income	0.11***	0.14***	0.19***	0.19***	0.15***	0.08***	0.11***	0.16***
	(0.037)	(0.026)	(0.025)	(0.026)	(0.030)	(0.022)	(0.018)	(0.020)
Household size	-0.03***	-0.03***	-0.04***	-0.05***	-0.02***	-0.02***	-0.03***	-0.04***
	(0.008)	(0.007)	(0.005)	(0.005)	(0.007)	(0.006)	(0.004)	(0.004)
Water source: Borehole or tank	0.08	-0.28**	-0.11	0.11	0.13	-0.26***	-0.12*	0.04
	(0.128)	(0.110)	(0.082)	(0.076)	(0.103)	(0.099)	(0.065)	(0.063)
Water source: Rain, spring or dam	0.03	-0.05	-0.07	0.12*	-0.18**	-0.19***	-0.17***	-0.14**
	(0.094)	(0.079)	(0.062)	(0.066)	(0.082)	(0.071)	(0.051)	(0.055)
Toilet type: Pit latrine or other	-0.02	0.05	-0.04	-0.01	0.01	-0.06	-0.15***	-0.11**
	(0.072)	(0.066)	(0.059)	(0.058)	(0.062)	(0.055)	(0.049)	(0.046)
Toilet type: None	-0.13	-0.17	-0.27***	-0.17*	0.05	-0.30***	-0.37***	-0.28***
	(0.150)	(0.112)	(0.105)	(0.101)	(0.123)	(0.094)	(0.074)	(0.085)
Urban	-0.01	-0.02	-0.01	0.04	0.07	-0.13**	-0.10**	-0.08*
	(0.087)	(0.072)	(0.061)	(0.058)	(0.068)	(0.060)	(0.047)	(0.046)
Province fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	-2.09***	-0.04	-0.82***	-1.32***	-1.59***	-1.75***	-1.09***	-1.46***
	(0.771)	(0.346)	(0.243)	(0.237)	(0.242)	(0.255)	(0.201)	(0.170)
Observations	2,690	2,531	3,700	3,551	3,789	3,975	5,377	5,483
R-squared	0.007	0.064	0.049	0.078	0.071	0.020	0.021	0.040

§ Variable was transformed using the inverse hyperbolic sine transformation. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table C 6: 2SLS regression results on the impact of current cigarette expenditure on child health outcomes

VARIABLES	Wave 2 (1) WAZ	Wave 3 (2) WAZ	Wave 4 (3) WAZ	Wave 5 (4) WAZ	Wave 2 (5) HAZ	Wave 3 (6) HAZ	Wave 4 (7) HAZ	Wave 5 (8) HAZ
Current cigarette exp. per adult§	-0.01	-0.26**	-0.05	-0.24	-0.18	-0.10	-0.22**	-0.24*
	(0.209)	(0.105)	(0.119)	(0.176)	(0.171)	(0.107)	(0.086)	(0.138)
log of hh income	0.41***	0.46***	0.53***	0.48***	0.47***	0.36***	0.45***	0.37***
	(0.097)	(0.081)	(0.061)	(0.052)	(0.084)	(0.068)	(0.053)	(0.044)
Age in months	-0.00***	-0.00***	-0.00***	-0.00***	-0.00	-0.00	0.00***	0.00***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)
Female	-0.06	0.14***	0.05	0.06	0.03	0.07*	0.09***	0.08***
	(0.060)	(0.043)	(0.035)	(0.036)	(0.049)	(0.038)	(0.031)	(0.031)
Coloured	-0.78**	0.00	-0.27	0.03	-0.06	-0.07	0.13	0.20
	(0.320)	(0.197)	(0.199)	(0.304)	(0.253)	(0.183)	(0.153)	(0.245)
Asian/Indian	-1.24**	-0.29	-0.66***	-0.29	0.26	0.55**	0.22	0.21
	(0.535)	(0.230)	(0.254)	(0.265)	(0.503)	(0.249)	(0.237)	(0.190)
White	-0.26	-0.05	-0.27	0.29	-0.01	0.30	0.31	0.76**
	(0.635)	(0.300)	(0.354)	(0.468)	(0.412)	(0.266)	(0.284)	(0.381)
Had major illness	-0.12	-0.20*	-0.15	-0.12	-0.06	-0.09	-0.25***	-0.16*
	(0.186)	(0.114)	(0.109)	(0.100)	(0.126)	(0.094)	(0.080)	(0.089)
Household size	-0.06***	-0.06***	-0.08***	-0.07***	-0.05***	-0.05***	-0.06***	-0.05***
	(0.010)	(0.009)	(0.008)	(0.008)	(0.009)	(0.007)	(0.007)	(0.007)
Water source: Borehole or tank	0.03	-0.19*	-0.14*	0.15*	0.15	-0.32***	-0.15**	0.10
	(0.132)	(0.104)	(0.076)	(0.078)	(0.112)	(0.102)	(0.069)	(0.069)
Water source: Rain, spring or dam	0.02	0.00	0.10	0.14**	-0.11	-0.18**	-0.07	-0.04
	(0.097)	(0.075)	(0.063)	(0.069)	(0.087)	(0.072)	(0.055)	(0.060)
Toilet type: Pit latrine or other	0.09	0.13*	0.11**	0.01	0.07	-0.02	-0.00	-0.11**
	(0.076)	(0.066)	(0.055)	(0.059)	(0.068)	(0.059)	(0.053)	(0.050)
Toilet type: None	0.00	-0.05	0.09	-0.02	0.09	-0.18*	0.11	-0.10
	(0.164)	(0.114)	(0.114)	(0.114)	(0.136)	(0.102)	(0.098)	(0.096)
Urban	-0.07	-0.03	0.03	0.00	0.00	-0.17***	-0.06	-0.09*
	(0.095)	(0.069)	(0.060)	(0.062)	(0.081)	(0.060)	(0.051)	(0.053)
Province fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	-2.19***	-3.10***	-3.96***	-3.49***	-4.21***	-3.23***	-4.17***	-3.60***
	(0.734)	(0.615)	(0.521)	(0.436)	(0.637)	(0.507)	(0.438)	(0.379)
Observations	2,726	4,598	4,866	4,821	4,077	6,303	6,805	6,840

§ Variable was transformed using the inverse hyperbolic sine transformation. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table C 7: 2SLS regression results on the impact of log of lagged cigarette expenditure on child health outcomes

	Wave 2	Wave 3	Wave 4	Wave 5	Wave 2	Wave 3	Wave 4	Wave 5
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep var	WAZ	WAZ	WAZ	WAZ	HAZ	HAZ	HAZ	HAZ
Log of lagged cigarette exp. per adult§	-15.26	-1.97	1.32	0.96	3.47	-1.46	4.02	2.28
	(72.417)	(1.391)	(4.009)	(1.116)	(10.123)	(1.072)	(11.556)	(1.975)
log of hh income	6.42	2.30**	-0.73	0.33	-0.88	1.34**	-2.99	-0.28
	(27.433)	(1.113)	(3.223)	(0.418)	(4.542)	(0.677)	(9.065)	(0.655)
Age in months	0.03	-0.00	-0.00	-0.00	-0.01	0.00	-0.00	0.00
	(0.166)	(0.004)	(0.006)	(0.002)	(0.010)	(0.002)	(0.014)	(0.002)
Female	0.27	-0.16	-0.10	0.06	0.12	0.08	-0.08	0.27
	(1.772)	(0.363)	(0.276)	(0.106)	(0.294)	(0.173)	(0.519)	(0.181)
Coloured	6.16	1.29	-0.61	-0.22	-1.63	1.05	-0.99	-0.83
	(32.482)	(1.224)	(0.897)	(0.263)	(4.650)	(0.926)	(2.658)	(0.682)
Asian/Indian	5.30	-0.48	-0.06	-0.34	-2.50	0.95	-0.73	-0.13
	(33.553)	(1.331)	(0.754)	(0.601)	(5.185)	(0.734)	(2.450)	(0.944)
White	27.04	1.53	-0.99	-1.27	-6.43	1.57	-2.53	-1.86
	(132.408)	(2.147)	(3.433)	(1.878)	(16.168)	(1.698)	(9.010)	(2.454)
Had major illness	2.31	0.01	-0.07	-0.16	0.08	0.31	-0.05	-0.15
	(11.511)	(0.653)	(0.517)	(0.240)	(0.709)	(0.471)	(1.032)	(0.466)
Household size	-1.74	-0.30	0.06	-0.01	0.29	-0.20	0.33	0.12
	(7.942)	(0.193)	(0.314)	(0.077)	(1.066)	(0.145)	(1.027)	(0.148)
Water source: Borehole or tank	1.58	-0.63	0.05	-0.30	-0.39	-0.38	0.15	-0.68
	(9.042)	(0.800)	(0.430)	(0.324)	(0.916)	(0.531)	(0.821)	(0.580)
Water source: Rain, spring or dam	-0.89	0.83	-0.16	-0.37	0.44	0.48	-0.11	-1.38
	(5.534)	(0.861)	(0.636)	(0.696)	(2.063)	(0.558)	(0.834)	(1.098)
Toilet type: Pit latrine or other	-2.28	0.03	-0.45	-0.18	0.67	-0.10	-1.24	-0.34
	(11.163)	(0.393)	(1.377)	(0.151)	(2.611)	(0.272)	(3.334)	(0.315)
Toilet type: None	-2.14	-0.38	-0.56	0.30	0.43	-0.42	-1.53	0.41
	(10.033)	(0.871)	(1.683)	(0.411)	(1.603)	(0.587)	(3.883)	(0.602)
Urban	0.85	0.17	-0.14	0.00	-0.27	0.15	0.09	0.12
	(4.007)	(0.385)	(0.648)	(0.141)	(0.491)	(0.237)	(0.408)	(0.265)
Province fixed effects	0.61	0.09	0.57	0.38*	-0.27	-0.02	1.66	0.00
	(4.405)	(0.524)	(1.286)	(0.222)	(2.366)	(0.322)	(4.189)	(0.312)
Province fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	-71.32	-20.80**	7.46	-2.29	11.20	-13.50**	27.25	3.13
	(316.889)	(10.435)	(30.715)	(4.122)	(49.213)	(6.472)	(84.710)	(6.699)
Observations	471	590	821	818	723	839	1,257	1,262

§ Variable was transformed using the natural log transformation. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table C 8: 2SLS regression results on the impact of current alcohol expenditure on child health outcomes

VARIABLES	Wave 2 (1) WAZ	Wave 3 (2) WAZ	Wave 4 (3) WAZ	Wave 5 (4) WAZ	Wave 2 (5) HAZ	Wave 3 (6) HAZ	Wave 4 (7) HAZ	Wave 5 (8) HAZ
Current alcohol exp. per adult[§]	-0.26	-0.26	-0.02	0.15	0.39	-0.18	-0.34**	1.16
	(0.817)	(0.236)	(0.174)	(0.685)	(0.798)	(0.177)	(0.170)	(1.640)
log of hh income	0.42***	0.50***	0.55***	0.45**	0.45***	0.39***	0.51***	0.03
	(0.117)	(0.090)	(0.060)	(0.186)	(0.103)	(0.078)	(0.058)	(0.489)
Age in months	-0.00***	-0.00***	-0.00***	-0.00***	-0.00	-0.00	0.00***	0.00**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)
Female	-0.10	0.15***	0.05	0.06	0.06	0.07*	0.07*	0.12
	(0.111)	(0.044)	(0.037)	(0.045)	(0.059)	(0.039)	(0.036)	(0.076)
Coloured	-0.77***	-0.22	-0.35***	-0.40*	-0.35**	-0.08	-0.00	-0.52
	(0.204)	(0.224)	(0.127)	(0.211)	(0.172)	(0.163)	(0.130)	(0.473)
Asian/Indian	-0.99	-0.37	-0.73***	-0.63***	-0.74	0.49**	0.07	-0.13
	(0.889)	(0.226)	(0.222)	(0.206)	(1.137)	(0.242)	(0.206)	(0.339)
White	0.16	0.04	-0.36	-0.63	-0.89	0.41	0.56	-2.01
	(1.430)	(0.586)	(0.427)	(1.200)	(1.203)	(0.400)	(0.418)	(3.091)
Had major illness	-0.13	-0.28**	-0.15	-0.11	-0.07	-0.11	-0.27***	-0.29
	(0.198)	(0.117)	(0.107)	(0.121)	(0.141)	(0.097)	(0.086)	(0.261)
Household size	-0.06***	-0.07***	-0.08***	-0.08***	-0.06***	-0.05***	-0.06***	-0.03
	(0.011)	(0.009)	(0.007)	(0.022)	(0.018)	(0.007)	(0.007)	(0.051)
Water source: Borehole or tank	0.09	-0.21*	-0.15*	0.16	0.07	-0.31***	-0.21***	0.24
	(0.223)	(0.107)	(0.082)	(0.110)	(0.163)	(0.100)	(0.074)	(0.227)
Water source: Rain, spring or dam	0.03	-0.00	0.11	0.16**	-0.14	-0.17**	-0.10*	-0.05
	(0.111)	(0.077)	(0.068)	(0.068)	(0.090)	(0.073)	(0.061)	(0.103)
Toilet type: Pit latrine or other	0.06	0.09	0.11**	0.01	0.08	-0.03	-0.03	-0.06
	(0.126)	(0.071)	(0.056)	(0.066)	(0.088)	(0.060)	(0.057)	(0.127)
Toilet type: None	-0.04	-0.02	0.08	-0.07	0.16	-0.15	0.07	-0.11
	(0.225)	(0.114)	(0.105)	(0.177)	(0.175)	(0.104)	(0.095)	(0.184)
Urban	-0.08	-0.01	0.03	-0.03	-0.02	-0.14**	-0.03	-0.15
	(0.098)	(0.075)	(0.070)	(0.084)	(0.077)	(0.064)	(0.062)	(0.107)

Province fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	-2.09***	-3.49***	-4.11***	-3.56***	-4.38***	-3.44***	-4.40***	-2.87**
	(0.771)	(0.662)	(0.508)	(0.548)	(0.687)	(0.565)	(0.475)	(1.449)
Observations	2,690	4,555	4,836	4,763	4,029	6,247	6,767	6,764

§ Variable was transformed using the inverse hyperbolic sine transformation. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table C 9: 2SLS regression results on the impact of log of lagged alcohol expenditure on child health outcomes

	Wave 2	Wave 3	Wave 4	Wave 5	Wave 2	Wave 3	Wave 4	Wave 5
VARIABLES	WAZ	WAZ	WAZ	WAZ	HAZ	HAZ	HAZ	HAZ
Log lagged alcohol exp. per adult§	4.13	0.12	-0.04	-1.72	0.99	-0.02	-0.15	-0.04
	(2.761)	(0.605)	(1.880)	(3.604)	(1.114)	(0.816)	(0.543)	(1.892)
log of hh income	0.77	0.74**	0.27	1.19	0.91**	1.01***	0.23	0.33
	(1.174)	(0.293)	(0.863)	(1.346)	(0.370)	(0.273)	(0.246)	(0.755)
Age in months	-0.01	-0.00*	0.00	-0.00	-0.00	0.00	0.00**	0.00**
	(0.012)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.001)	(0.001)
Female	0.27	0.06	-0.15	-0.09	0.22	0.12	0.07	0.16
	(0.555)	(0.148)	(0.126)	(0.264)	(0.190)	(0.138)	(0.084)	(0.131)
Coloured	-0.33	-0.67***	-0.52**	-0.10	-0.45	-0.44	-0.25*	-0.09
	(1.764)	(0.253)	(0.245)	(0.298)	(0.437)	(0.266)	(0.139)	(0.108)
Asian/Indian	-0.41	-0.67	-1.13	0.07	-0.79	0.20	-0.31	-0.22
	(2.316)	(0.410)	(1.777)	(1.100)	(0.871)	(0.526)	(0.433)	(0.494)
White	-2.42	-1.16*	0.03	0.23	-2.46**	-1.36**	0.36	0.45
	(3.084)	(0.661)	(0.780)	(0.819)	(1.078)	(0.582)	(0.309)	(0.620)
Had major illness	-0.78	0.11	-0.42	0.23	0.33	0.17	-0.37	0.07
	(1.078)	(0.443)	(0.885)	(0.780)	(0.381)	(0.362)	(0.349)	(0.237)
Household size	0.35	-0.04*	-0.07	-0.13	0.02	-0.03	-0.07***	-0.05
	(0.316)	(0.021)	(0.083)	(0.151)	(0.120)	(0.026)	(0.022)	(0.092)
Water source: Borehole or tank	-3.24	-0.20	0.22	-0.35	-0.37	-0.03	0.01	0.12
	(2.109)	(0.633)	(0.320)	(0.517)	(0.646)	(0.721)	(0.169)	(0.166)
Water source: Rain, spring or dam	0.36	0.08	-0.29	0.53	0.61	-0.71	-0.19	-0.35
	(1.299)	(0.299)	(0.254)	(1.445)	(0.538)	(0.492)	(0.202)	(0.679)
Toilet type: Pit latrine or other	0.67	0.05	-0.11	-0.30	0.02	-0.10	-0.12	-0.22
	(0.687)	(0.325)	(0.421)	(0.398)	(0.220)	(0.344)	(0.160)	(0.138)

Toilet type: None	1.00	-0.40	0.12	-0.45	-0.59	0.33	-0.32	-0.42
	(1.570)	(0.526)	(1.140)	(0.712)	(0.623)	(0.487)	(0.236)	(0.423)
Urban	-0.25	-0.26	0.10	-0.86	-0.22	-0.51**	-0.03	-0.30
	(0.705)	(0.238)	(0.228)	(1.444)	(0.253)	(0.228)	(0.129)	(0.582)
Province fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	-22.85*	-5.93**	-1.90	-1.10	-11.54***	-8.66***	-1.71	-3.17
	(11.681)	(2.515)	(2.611)	(8.229)	(3.823)	(2.999)	(1.240)	(3.583)
Observations	255	351	604	657	373	494	898	991

Table C 10: The Mediated effect of Alcohol Expenditure by Food Expenditure on Child Nutrition Outcomes

	Wave 2	Wave 3	Wave 4	Wave 5	Wave 2	Wave 3	Wave 4	Wave 5
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variable	WAZ	WAZ	WAZ	WAZ	HAZ	HAZ	HAZ	HAZ
Total Effect	-1.07	-0.46	-0.28	-0.12	-0.13	-0.17	-0.15	-0.37**
	-2.14	-0.54	-0.21	-0.2	-0.96	-0.38	-0.16	-0.15
Direct Effect	-0.77	-0.46	-0.28	-0.11	-0.16	-0.16	-0.16	-0.37**
	-1.34	-0.62	-0.21	-0.19	-0.83	-0.42	-0.16	-0.15
C. Lagged alcohol exp. per adult§	-0.05	0.50*	0.10	-0.13	-0.05	0.50*	0.10	-0.13
	-0.05	-0.26	-0.10	-0.08	-0.05	-0.26	-0.10	-0.08
D. Per capita hh food exp.§	0.00	0.01	0.02	0.06	0.00	0.02	0.02	0.07
	0.00	-0.15	-0.05	-0.05	0.00	-0.10	-0.04	-0.05
Bootstrap results								
Indirect effect [C x D]	-0.32	0.01	0.00	-0.01	0.04	0.01	0.00	-0.01
Bootstrapped std. errors	-1.66	-53.06	-0.03	-0.06	-48.1	-0.49	-0.01	-0.02
z-Statistic	-0.19	0.00	0.09	-0.16	0.00	0.02	0.31	-0.46
P-value	0.85	1.00	0.93	0.88	1.00	0.99	0.75	0.65
Observations	2,425	3,316	3,445	3,652	3,808	4,813	5,329	5,580

§ Variable was transformed using the inverse hyperbolic sine transformation. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

First, the Stata user-written program *ivhettest* was used to test for heteroskedasticity, following the Pagan & Hall (1983) method. This is an essential test, as the consequences of heteroskedasticity in instrumental variable regression are akin to those observed in OLS regression. That is, the instrumental variable coefficient estimates remain consistent, but the standard errors of these estimates become inconsistent (Baum, Schaffer & Stillman, 2003). The presence of heteroskedasticity has implications for the choice of diagnostic tests, as most tests assume homoskedasticity.

Table C11: Pagan-Hall statistic for heteroskedasticity

	Lagged cigarette expenditure		Lagged alcohol expenditure	
	WAZ	HAZ	WAZ	HAZ
Wave 2	31.12	49.08**	18.80	48.90**
Wave 3	34.89*	82.02***	60.84***	95.22***
Wave 4	26.67	33.27*	27.26	37.17**
Wave 5	24.33	38.32**	31.34	33.098*

The Pagan-Hall general statistic is presented for the different dependent variables in each wave. The null hypothesis of homoskedasticity is rejected if it is statistically significant indicated as *** p<0.01, ** p<0.05, * p<0.1.

The test results in Table 4.4 provide insight on the presence of heteroskedasticity in the models, using WAZ and HAZ as dependent variables and estimating the impact of lagged cigarette and alcohol expenditure.

For the models using WAZ, the results indicate that heteroskedasticity is not a significant concern, except for the Wave 3 data. In other words, the assumption of homoskedasticity holds reasonably well for most waves when examining the relationship between lagged cigarette or alcohol expenditure and WAZ scores.

In contrast, the models using HAZ reveal a different pattern. All HAZ models reject the null hypothesis of homoskedasticity. Thus, the assumption of equal variance across observations is violated when examining the impact of lagged cigarette or alcohol expenditure on HAZ scores.

These findings highlight the need to consider and address heteroskedasticity, particularly in the HAZ models, in order to ensure the validity of the statistical inferences drawn from the analysis.

5. CHAPTER 5: CONCLUSION

This thesis primarily investigates the repercussions of tobacco expenditure on economic and health outcomes, while exploring the role of alcohol expenditure in specific contexts, such as its interplay with tobacco expenditure (Chapter 3) and its potential effects on child health outcomes (Chapter 4). The analysis rested heavily upon the crowding-out framework, a well-established concept within the literature pertaining to household-level cigarette expenditure. Framed against the backdrop of escalating cigarette prices, Chapters 2 and 3 used data from four Income and Expenditure Surveys (IES) datasets to identify the specific categories of expenditure influenced by cigarette expenditure and to examine the evolution of budget allocations over time. Building upon this framework, Chapter 4 delved into the potential repercussions of such expenditure on children's health outcomes. In the following section, I synthesise the findings, highlight their implications, discuss limitations, and offer guidance on potential avenues for future research.

5.1 Main findings and contributions of this thesis

Chapter 2 contributes to the literature that identifies categories of expenditure displaced by cigarette expenditure. Prior SSA studies in this vein were conducted in countries such as Zambia (Chelwa & Van Walbeek, 2014), South Africa (Chelwa & Koch, 2019, Koch & Tshiswaka-Kashalala, 2008), Ghana (Masa-Ud, Chelwa & Van Walbeek, 2020), and Kenya (Nyagwachi, Chelwa & van Walbeek, 2020). This study adds to the small number of studies from the region on this topic and reinforces the findings of others on the effect of cigarette expenditure on household spending patterns.

In contrast to prior research that utilised IES data from South Africa (van Walbeek, 2005), this study is the first to address various data quality issues comprehensively before conducting the crowding-out analysis. These issues encompassed changes in survey instruments, modifications in sampling procedures, and the handling of outliers and influential data points. Resolving these concerns was crucial to enabling meaningful comparisons over multiple years, although data quality remains an ongoing area of research and refinement.

The findings demonstrated that cigarette spending displaced expenditures across various categories, including food, education, housing, clothing, transportation, communication, healthcare, household goods, and miscellaneous items, when a simple difference in budget shares was employed. 3SLS estimation techniques confirmed that education, transport, housing, and household goods expenditures had indeed been crowded-out. The potential

developmental ramifications are evident, particularly in terms of reduced household-level investment in education, potentially resulting in diminished human capital development.

The third chapter extends our understanding of crowding-out by analysing how this displacement phenomenon has evolved over time, especially in the presence of substantial cigarette-price increases. This analysis builds upon the DiD approach used by Nyagwachi, Chelwa & van Walbeek (2020), introducing a novel DiD method that integrates propensity scores within a direct weighting scheme (Abadie, 2005). This innovative approach sidesteps the potential pitfalls associated with propensity-score matching DiD methods and approaches used in earlier studies.

Beyond scholarly interest, dynamic analysis of the impact of tobacco-tax increases has several important contributions to make to policymaking. First, from the literature, we understand that when tobacco consumption drops because of tobacco-price increases, it is because of reduction on the extensive margin (reduction in the number of smokers) and reduction on the intensive margin (reduction in the number of cigarettes smoked per smoker). This dynamic analysis shows how these factors interplay with changes in fiscal policy, affecting household-level budgeting decisions. It thus helps explain the effectiveness of cigarette tax policies over time, highlighting their continued relevance. Secondly, it reveals possible unintended consequences among households that continue purchasing cigarettes after substantial price increases, such as deeper expenditure displacements, which policymakers must consider. Lastly, policymakers gain valuable insights into the long-term effectiveness of policies, and can refine existing strategies, identify areas for improvement, and shape future interventions.

While some evidence suggests possible gains, for instance, reductions in crowding-out on education among low-income households between 1995 and 2000, there are negative effects as well, such as the crowding-in effect seen for alcohol expenditure. Enhancement of other tobacco-control interventions, such as cessation programs, may mitigate these negative effects.

The chapter also underscores a clear relationship between tobacco and alcohol consumption. It consistently reveals that alcohol expenditure is crowded-in, though the level of crowding-in is reduced during the latter part of the review period. This underscores the potential impact of tobacco-control policies on alcohol consumption, highlighting the need for policymakers to consider these effects to understand public health implications comprehensively. The evidence from this chapter suggests that reductions in tobacco use can be associated with a reduction in alcohol use, thus tobacco control could be used as a tool to complement alcohol control policy.

The fourth chapter advances our understanding of child health. Specifically, it focuses on the impact of parental alcohol and tobacco consumption on children aged 6 months to 5 years, a developmental phase often overlooked in existing literature. The chapter adopts a holistic approach by implicitly considering the influence of both parents' health behaviours, as indicated by household expenditure on cigarettes and alcohol, rather than exclusively focusing on maternal behaviour, which has more often been the focus in the literature.

Regarding household-level expenditures, this chapter has conceptual similarities to Block & Webb (2009a) and Mellor (2011) in that it aims to establish a connection between household tobacco expenditure and child health. In the first place, it introduces two methodological enhancements. It employs various instruments to address endogeneity in cigarette and alcohol expenditure, together with household income. Secondly, it utilises the inverse hyperbolic sine (IHS) transformation for expenditure variables, mitigating potential issues related to their distributions. These methodological refinements enhance the credibility of the findings and provide a more accurate understanding of the relationships involved.

Moreover, the chapter introduces mediation analysis to explore the mechanisms through which cigarette or alcohol expenditure influences child health outcomes. This offers a more nuanced understanding of the relationship between household expenditures and child health outcomes.

5.2 Main limitations of the thesis

In the second chapter, several technical constraints emerged. Specifically, the analysis was incapable of accommodating price clustering within distinct geographic regions. The 3SLS cross-sectional analysis employed omitted regional fixed effects, a customary inclusion (Deaton, 2018). The resulting loss in regional specificity means that some of the regional nuances of the results could not be presented. The second limitation pertained to the endogeneity inherent in the tobacco expenditure and adjusted total household expenditure variables. Although partially mitigated through the use of the adult-sex ratio and income as instruments, these remedies remained imperfect. Consequently, the chapter used Nevo & Rosen (2012) method to estimate confidence intervals within which the true point estimates are likely to lie. Furthermore, this chapter, mainly founded upon different cross-sectional analyses, lacked provisions that could account for changes in the economic landscape that might influence household spending patterns.

In the third chapter, the models adopted also failed to incorporate fixed effects for clustering, consistent with the framework suggested by Bertrand, Duflo & Mullainathan (2004). This

omission, as previously noted, stemmed from inconsistencies within the dataset variables, culminating in their exclusion from the final model. As well, while the diagnostic procedures supported the effective deployment of the semi-parametric difference-in-difference method, it rested on the premise that only the observable elements employed in the methodology impacted household budget allocations. This overlooked latent unobservable heterogeneity, presenting a data limitation that could be addressed using panel data.

In the fourth chapter, the challenge of endogeneity within cigarette expenditure was confronted again. Prior studies employed diverse techniques, such as non-parametric methodologies or the inclusion of child-fixed effects, to counter this issue. In this chapter, an instrumental variable approach was employed to account for the endogenous nature of both cigarette and alcohol expenditure and household income. Diagnostic tests indicated that the adult-sex ratio was generally a robust instrument for cigarette expenditure. In contrast, the instrument for alcohol expenditure was not strong. This is likely to have contributed to the absence of statistically significant results in the analysis of alcohol expenditure.

5.3 Ideas for future research

Collectively, the findings presented in this thesis contribute to a comprehensive understanding of the multifaceted impact of tobacco and alcohol usage on various dimensions of the well-being of individuals and households. These results underscore the intricate interconnections between these effects, emphasising the importance of adopting holistic approaches in policymaking.

Several potential avenues for future research emerge from the limitations discussed above and the scope of this thesis:

- *Extending analysis on child health impact*: Another avenue to explore would be to broaden the scope of research on the impact of alcohol and tobacco use on child health outcomes in two significant directions. First, one could explore the intergenerational consequences by assessing the effects of parental tobacco and alcohol consumption on children's educational outcomes, employment prospects, and income-earning abilities in later life. Leveraging panel data methods could enhance the depth of this exploration. Secondly, one could revisit the mediation analysis to scrutinise whether the adverse effects on child health outcomes can be attributed to reduced food expenditure. Although the present study did not establish a strong connection, alternative techniques

might yield more robust insights. Investigating other potential pathways that could elucidate the observed decline in child health outcomes would be a valuable effort.

- *Gender-specific effects*: A comprehensive analysis of the gender-specific dimensions of tobacco and alcohol consumption could also be worthwhile. This could be done in the context of an analysis of the intra-household decision-making processes. A thorough analysis of how gender influences patterns of alcohol and tobacco use may unveil previously unexplored relationships. This, in turn, has the potential to inform and shape more nuanced and effective policy interventions for these products. The intersections between gender dynamics and substance use could offer fresh perspectives that contribute to a comprehensive understanding of the issue.
- *Mental health, social networks, and coping mechanisms*: A possible area of investigation is the interplay between stress, mental health, social networks, and coping mechanisms in shaping the relationship between expenditure on tobacco or alcohol and health outcomes. Individually, each of these factors has an influence on whether individuals spend on tobacco and alcohol, and how much they spend. However, these factors are also likely to influence each other, either reinforcing or entrenching certain patterns of behaviour that ultimately have some impact on tobacco and alcohol expenditure. Understanding these multifaceted relationships could provide valuable insights to help refine relevant policies affecting these behaviours.

These potential future research areas offer opportunities to build upon the foundations established in this thesis, contributing to a more comprehensive and nuanced understanding of the economic and health implications of tobacco and alcohol consumption. Such endeavours hold the promise of informing more effective policymaking in economics and public health.

In summary, this thesis emphasises the necessity for policymakers to consider more than just the immediate effects of tobacco-control policies on other household expenditures, and especially to consider their potential influences on alcohol consumption patterns. Furthermore, it underscores the importance of safeguarding the well-being of children in households affected by these behaviours. By adopting such comprehensive strategies, policymakers can better address the multifaceted implications of tobacco and alcohol use on both individuals and society.

6. REFERENCES

Abadie, A. 2005. Semiparametric difference-in-difference estimators. *Review of Economic Studies.* 72:1-19.

Abadie, A. & Imbens, G.W. 2006. Large sample properties of matching estimators for average treatment effects. *Econometrica.* 74(1):235-267.

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.

Abouk, R., Courtemanche, C.J., Dave, D.M., Feng, B., Friedman, A.S., Maclean, J.C., Pesko, M.F., Sabia, J.J. et al. 2021. *Intended and unintended effects of e-cigarette taxes on youth tobacco use.*

Abrevaya, J. & Dahl, C.M. 2008. The effects of birth inputs on birthweight: evidence from quantile estimation on panel data. *Journal of Business & Economic Statistics.* 26(4):379-397.

Adda, J. & Cornaglia, F. 2010. The effect of bans and taxes on passive smoking. *American Economic Journal: Applied Economics.* 2(1):1-32.

Agee, M.D., Atkinson, S.E. & Crocker, T.D. 2009. Multiple-Output Child Health Production Functions: The Impact of Time-Varying and Time-Invariant Inputs. *Southern Economic Journal.* 75(3):909-927.

Aihounton, G.B.D. & Henningsen, A. 2021. Units of measurement and the inverse hyperbolic sine transformation. *The Econometrics Journal.* 24(2):334-351. DOI:10.1093/ectj/utaa032.

Akombi, B.J., Agho, K.E., Hall, J.J., Wali, N., Renzaho, A.M.N. & Merom, D. 2017. Stunting, Wasting and Underweight in Sub-Saharan Africa: A Systematic Review. *Int J Environ Res Public Health.* 14(8). DOI:10.3390/ijerph14080863.

Almond, D. & Currie, J. 2011. Killing Me Softly: The Fetal Origins Hypothesis. *J Econ Perspect.* 25(3):153-172. DOI:10.1257/jep.25.3.153.

Andrews, I., Stock, J.H. & Sun, L. 2019. Weak Instruments in Instrumental Variables Regression: Theory and Practice. *Annual Review of Economics.* 11(1):727-753. DOI:10.1146/annurev-economics-080218-025643.

Angrist, J.D. & Pischke, J.-S. 2009. *Mostly harmless econometrics: An empiricist's companion.* Princeton university press.

Ardington, C., Lam, D., Leibbrandt, M. & Welch, M. 2006. The sensitivity to key data imputations of recent estimates of income poverty and inequality in South Africa. *Economic Modelling.* 23(5):822-835. DOI:10.1016/j.econmod.2005.10.009.

Bahar, D. & Rapoport, H. 2018. Migration, Knowledge Diffusion and the Comparative Advantage of Nations. *The Economic Journal.* 128(612):F273-F305. DOI:10.1111/ecoj.12450.

Baker, A., Larcker, D.F. & Wang, C.C. 2021. How Much Should We Trust Staggered Difference-In-Differences Estimates? *Available at SSRN 3794018.*

Banerjee, A.K. 2010. A multidimensional Gini index. *Mathematical Social Sciences.* 60(2):87-93. DOI:10.1016/j.mathsocsci.2010.06.001.

Banks, J., Blundell, R. & Lewbel, A. 1997. Quadratic Engel Curves and Consumer Demand. *The Review of Economics and Statistics.* 79(4):527-539.

Barker, D.J. 1990. The fetal and infant origins of adult disease. *BMJ: British Medical Journal.* 301(6761):1111.

Baron, R.M. & Kenny, D.A. 1986. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology.* 51(6):1173.

Baum, C.F., Schaffer, M.E. & Stillman, S. 2003. Instrumental variables and GMM: Estimation and testing. *The Stata Journal.* 3(1):1-31.

Behrman, J.R. & Deolalikar, A.B. 1988. Health and nutrition. *Handbook of development economics.* 1:631-711.

Bellemare, M.F. & Wichman, C.J. 2019. Elasticities and the Inverse Hyperbolic Sine Transformation. *Oxford Bulletin of Economics and statistics.* 82(1):50-61. DOI:10.1111/obes.12325.

Ben Lakhdar, C., Vaillant, N.G. & Wolff, F.C. 2016. Does smoke cross the border? Cigarette tax avoidance in France. *Eur J Health Econ.* 17(9):1073-1089. DOI:10.1007/s10198-015-0746-1.

Bertrand, M., Duflo, E. & Mullainathan, S. 2004. *How much should we trust difference in difference estimates?*

Bharadwaj, P., Johnsen, J.V. & Løken, K.V. 2014. Smoking bans, maternal smoking and birth outcomes. *Journal of Public Economics.* 115:72-93. DOI:10.1016/j.jpubeco.2014.04.008.

Bishai, D. 2001. Lifecycle changes in the rate of time preference: testing the theory of endogenous preferences and its relevance to adolescent substance abuse. *Economic Analysis of Substance Abuse.*155-177.

Block, S. & Webb, P. 2009a. Up in smoke: tobacco use, expenditure on food, and child malnutrition in developing countries. *Economic Development and Cultural Change.* 58(1):1-23.

Block, S. & Webb, P. 2009b. Up in smoke_tobacco use, expenditure on food and child malnutrition in developing countries. *Economic Development and Cultural Change.* 58(1):1-23.

Blundell, R. & Macurdy, T. 1999. Labor Supply: A Review of Alternative Approaches. 1559-1695. DOI:10.1016/s1573-4463(99)03008-4.

Blundell, R. & Costa Dias, M. 2009. Alternative Approaches to Evaluation in Empirical Microeconomics. *Journal of Human Resources.* 44(3):565-640.

Boachie, M.K., Rossouw, L. & Ross, H. 2021. The Economic Cost of Smoking in South Africa, 2016. *Nicotine & Tobacco Research.* 23(2):286-293. DOI:10.1093/ntr/ntaa162.

Browning, M. & Meghir, C. 1991. The effects of male and female labor supply on commodity demands. *Econometrica: Journal of the econometric society.*925-951.

Burger, R. & Yu, D. 2007. Wage Trends in Post-Apartheid South Africa: Constructing an Earnings Series from Household Survey Data. *SSRN Electronic Journal.* 10.2139/ssrn.966118. DOI:10.2139/ssrn.966118.

Busch, S.H., Jofre-Bonet, M., Falba, T.A. & Sindelar, J.L. 2004. Burning A Hole In The Budget. *Appl Health Econ Health Policy.* 3(4):263-272.

Bütikofer, A., Mølland, E. & Salvanes, K.G. 2018. Childhood nutrition and labor market outcomes: Evidence from a school breakfast program. *Journal of Public Economics.* 168:62-80. DOI:10.1016/j.jpubeco.2018.08.008.

Calitz, E. & Fourie, J. 2016. The historically high cost of tertiary education in South Africa. *Politikon.* 43(1):149-154.

Callinan, S., Room, R. & Dietze, P. 2015. *Alcohol price policies as an instrument of health equity: Differential effects of tax and minimum price measures.* Oxford University Press.

Callison, K. & Kaestner, R. 2014. Do Higher Tobacco Taxes Reduce Adult Smoking? New Evidence of the Effect of Recent Cigarette Tax Increases on Adult Smoking. *Economic Inquiry.* 52(1):155-172. DOI:10.1111/ecin.12027.

Cameron, A.C. & Trivedi, P.K. 2005. *Microeconometrics: methods and applications.* Cambridge university press.

Cameron, A.C. & Miller, D.L. 2015. A practitioner's guide to cluster-robust inference. *Journal of Human Resources.* 50(2):317-372.

Caro, J.C., Valizadeh, P., Correa, A., Silva, A. & Ng, S.W. 2020. Combined fiscal policies to promote healthier diets: Effects on purchases and consumer welfare. *PLoS One.* 15(1):e0226731.

Catalano, M.A. & Gilleskie, D.B. 2021. Impacts of local public smoking bans on smoking behaviors and tobacco smoke exposure. *Health Economics.* 30(8):1719-1744.

Celli, V. 2021. Causal mediation analysis in economics: Objectives, assumptions, models. *Journal of Economic Surveys.* 36(1):214-234. DOI:10.1111/joes.12452.

Chang, N.C. 2020. Double/debiased machine learning for difference-in-differences models. *The Econometrics Journal.* 23(2):177-191. DOI:10.1093/ectj/utaa001.

Charmarbagwala, R., Ranger, M., Waddington, H. & White, H. 2004. The determinants of child health and nutrition: a meta-analysis.

Chelwa, G. & Van Walbeek, C. 2014. *Assessing the causal impact of tobacco expenditure on household spending patterns in Zambia*.

Chelwa, G. & Koch, S.F. 2019. The effect of tobacco expenditure on expenditure shares in South African households: A genetic matching approach. *PLoS One.* 14(9):e0222000. DOI:10.1371/journal.pone.0222000.

Chelwa, G., van Walbeek, C. & Blecher, E. 2016. Evaluating South Africa's tobacco control policy using a synthetic control method. *Tob Control.* 26(5):509-517. DOI:10.1136/tobaccocontrol-2016-053011.

Clemens, M.A. & Tiongson, E.R. 2017. Split Decisions: Household Finance When a Policy Discontinuity Allocates Overseas Work. *The Review of Economics and Statistics.* 99(3):531-543. DOI:10.1162/REST_a_00657.

Cohen, N. 1981. Smoking, health, and survival: prospects in Bangladesh. *The Lancet.* 317(8229):1090-1093.

Coneus, K. & Spiess, C.K. 2012. Pollution exposure and child health: evidence for infants and toddlers in Germany. *J Health Econ.* 31(1):180-196. DOI:10.1016/j.jhealeco.2011.09.006.

Cooper, M.T. & Pesko, M.F. 2017. The effect of e-cigarette indoor vaping restrictions on adult prenatal smoking and birth outcomes. *J Health Econ.* 56:178-190. DOI:10.1016/j.jhealeco.2017.10.002.

Croson, R. & Gneezy, U. 2009. Gender differences in preferences. *Journal of Economic Literature.* 47(2):448-474.

Currie, J. & Almond, D. 2011. Human capital development before age five. Elsevier. 1315-1486. DOI:10.1016/s0169-7218(11)02413-0.

Dasgupta, K., Ghimire, K.M. & Pacheco, G. 2022. How heavy is the price of smoking? Estimating the effects of prenatal smoking on child weight outcomes. *Applied Economics.* 54(44):5159-5175. DOI:10.1080/00036846.2022.2041178.

De Chaisemartin, C. & D'Haultfœuille, X. 2020. Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects. *American Economic Review.* 110(9):2964-2996. DOI:10.1257/aer.20181169.

Deaton, A. 2018. *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy. Reissue Edition with a New Preface*. Washington, DC.

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

Do, Y.K. & Bautista, M.A. 2015. Tobacco use and household expenditures on food, education, and healthcare in low- and middle-income countries: a multilevel analysis. *BMC Public Health.* 15:1098. DOI:10.1186/s12889-015-2423-9.

Dube, S.R., Anda, R.F., Felitti, V.J., Croft, J.B., Edwards, V.J. & Giles, W.H. 2001. Growing up with parental alcohol abuse: exposure to childhood abuse, neglect, and household dysfunction. *Child abuse & neglect.* 25(12):1627-1640.

Dünnbier, M. 2020. *Alcohol obstacle to development: How alcohol affects the sustainable development goals.* Movendi International. https://movendi. ngo/wp-content/uploads/2020/02 ….

Efroymson, D., Ahmed, S., Townsend, J., Alam, S.M., Dey, A.R., Saha, R., Dhar, B., Sujon, I.A. et al. 2001. Hungry for tobacco: an analysis of the economic impact of tobacco consumption on the poor in Bangladesh. *Tob Control.* 10:212-217.

Evans, W.N. & Farrelly, M.C. 1998. The Compensating Behavior of Smokers: Taxes, Tar, and Nicotine. *The RAND Journal of Economics.* 29(3):578. DOI:10.2307/2556105.

Farrelly, M.C., Davis, K.C., Haviland, M.L., Messeri, P. & Healton, C.G. 2005. Evidence of a Dose—Response Relationship Between "truth" Antismoking Ads and Youth Smoking Prevalence. *American Journal of Public Health.* 95(3):425-431. DOI:10.2105/ajph.2004.049692.

Filmer, D. & Pritchett, L.H. 2001. Estimating wealth effects without expenditure data—or tears: an application to educational enrollments in states of India. *Demography.* 38(1):115-132.

Foltran, F., Gregori, D., Franchin, L., Verduci, E. & Giovannini, M. 2011. Effect of alcohol consumption in prenatal life, childhood, and adolescence on child development. *Nutrition reviews.* 69(11):642-659.

Fontes Marx, M., London, L., Harker, N. & Ataguba, J.E. 2019. Equity in household spending on alcoholic beverages in South Africa: assessing changes between 1995 and 2011. *International journal for equity in health.* 18:1-10.

Francis, D. & Webster, E. 2019. Poverty and inequality in South Africa: critical reflections. *Development Southern Africa.* 36(6):788-802.

Friedline, T., Masa, R.D. & Chowa, G.A. 2015. Transforming wealth: using the inverse hyperbolic sine (IHS) and splines to predict youth's math achievement. *Soc Sci Res.* 49:264-287. DOI:10.1016/j.ssresearch.2014.08.018.

Frölich, M. & Huber, M. 2017. Direct and indirect treatment effects–causal chains and mediation analysis with instrumental variables. *Journal of the Royal Statistical Society. Series B (Statistical Methodology).*1645-1666.

Gallego, J.M., Paraje, G. & Rodríguez-Lesmes, P. 2024. Inequality of the crowding-out effect of tobacco expenditure in Colombia. *PLoS One.* 19(5):e0303328.

Gallego, J.M., Llorente, B., Maldonado, N., Otalvaro-Ramirez, S. & Rodriguez-Lesmes, P. 2020. Tobacco taxes and illicit cigarette trade in Colombia. *Econ Hum Biol.* 39:100902. DOI:10.1016/j.ehb.2020.100902.

Gallet, C. 2007. A comparative analysis of the demand for higher education: Results from a meta-analysis of elasticities. *Economics Bulletin.* 9(7):1-14.

Gao, J. & Baughman, R.A. 2016. Do Smoking Bans Improve Infant Health? Evidence from U.S. Births: 1995–2009. *Eastern Economic Journal.* 43(3):472-495. DOI:10.1057/s41302-016-0010-0.

Giannelli, G.C., Mangiavacchi, L. & Piccoli, L. 2013. Do parents drink their children's welfare? Intra-household allocation of time between market labour, domestic work and child care in Russia. *IZA Journal of Labor & Development.* 2:1-23.

Gibbs, N., Angus, C., Dixon, S., Parry, C.D., Meier, P.S., Boachie, M.K. & Verguet, S. 2022. Equity impact of minimum unit pricing of alcohol on household health and finances among rich and poor drinkers in South Africa. *BMJ global health.* 7(1):e007824.

Goodchild, M., Nargis, N. & d'Espaignet, E.T. 2018. Global economic cost of smoking-attributable diseases. *Tobacco Control.* 27(1):58-64.

Greene, W.H. 2000. Econometric analysis 4th edition. *International edition, New Jersey: Prentice Hall.*201-215.

Grossman, M. 1972. On the Concept of Health Capital and the Demand for Health. *Journal of political Economy.* 80(2):223-255. Available: http://www.jstor.org/stable/1830580 [2022/12/15/].

Hajdu, T. & Hajdu, G. 2018. Smoking ban and health at birth: Evidence from Hungary. *Econ Hum Biol.* 30:37-47. DOI:10.1016/j.ehb.2018.05.003.

Hamilton, B.H. 2001. Estimating treatment effects in randomized clinical trials with non-compliance: the impact of maternal smoking on birthweight. *Health Econ.* 10(5):399-410. DOI:10.1002/hec.629.

Hampel, F.R. 1974. The influence curve and its role in robust estimation. *Journal of the American statistical Association.* 69(346):383-393.

Hansen, B.E. & Lee, S. 2020. *Inference for Iterated GMM Under Misspecication.*

Hawkins, S.S. & Baum, C.F. 2014. Impact of state cigarette taxes on disparities in maternal smoking during pregnancy. *Am J Public Health.* 104(8):1464-1470. DOI:10.2105/AJPH.2014.301955.

Hawkins, S.S., Bach, N. & Baum, C.F. 2018. Impact of tobacco control policies on adolescent smokeless tobacco and cigar use: a difference-in-differences approach. *BMC Public Health.* 18(1):154. DOI:10.1186/s12889-018-5063-z.

Hawkins, S.S., Kull, M. & Baum, C.F. 2019. US state cigarette tax increases and smoke-free legislation in relation to cigarette expenditure across household socio-economic circumstances: a quasi-experimental study. *Addiction.* 114(4):721-729. DOI:10.1111/add.14505.

Heckman, J., Ichimura, H. & Todd, P. 1997. Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The Review of Economic Studies.* 64(4):605-654.

Heckman, J., Ichimura, H., Smith, J. & Todd, P. 1998. *Characterizing selection bias using experimental data.* National bureau of economic research.

Heckman, J.J. & Pinto, R. 2015. Econometric mediation analyses: Identifying the sources of treatment effects from experimentally estimated production technologies with unmeasured and mismeasured inputs. *Econometric reviews.* 34(1-2):6-31.

Hoddinott, J., Alderman, H., Behrman, J.R., Haddad, L. & Horton, S. 2013. The economic rationale for investing in stunting reduction. *Matern Child Nutr.* 9 Suppl 2(Suppl 2):69-82. DOI:10.1111/mcn.12080.

Holland, P.W. 1986. Statistics and Causal Inference. *Journal of the American statistical Association.* 81(396):945-960.

Horvitz, D.G. & Thompson, D.J. 1952. A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association.* 47(260):663-685.

Houngbedji, K. 2016. Abadie's semiparametric difference-in-differences estimator. *The Stata Journal.* 16(2):482–490.

Huber, P.J. 2004. *Robust statistics.* John Wiley & Sons.

Humphreys, B. 2015. *Dealing with Zeros in Economic Data: Theory and Application to Participation in Leisure Time Physical Activity*.

Husain, M.J., Datta, B.K., Virk-Baker, M.K., Parascandola, M. & Khondker, B.H. 2018. The crowding-out effect of tobacco expenditure on household spending patterns in Bangladesh. *PLoS One.* 13(10):e0205120. DOI:10.1371/journal.pone.0205120.

Imai, K., Keele, L. & Yamamoto, T. 2010. Identification, Inference and Sensitivity Analysis for Causal Mediation Effects. *Statistical Science.* 25(1). DOI:10.1214/10-sts321.

Imai, K., Keele, L., Tingley, D. & Yamamoto, T. 2011. Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies. *American Political Science Review.* 105(4):765-789. DOI:10.1017/s0003055411000414.

Imbens, G.W. & Wooldridge, J.M. 2009. Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature.* 47(1):5-86. DOI:10.1257/jel.47.1.5.

Imbens, G.W. & Rubin, D.B. 2015. *Causal inference in statistics, social, and biomedical sciences.* Cambridge University Press.

Ismail, G. & Suffla, S. 2013. Child Safety. *Peace and Health Promotion Child Malnutrition.*

Jenkner, C., Lorenz, E., Becher, H. & Sauerbrei, W. 2016. Modeling continuous covariates with a "spike" at zero: Bivariate approaches. *Biom J.* 58(4):783-796. DOI:10.1002/bimj.201400112.

Jha, P. 2009. Avoidable global cancer deaths and total deaths from smoking. *Nature Reviews Cancer.* 9(9):655-664.

Jin, H.J. & Cho, S.M. 2021. Effects of cigarette price increase on fresh food expenditures of low-income South Korean households that spend relatively more on cigarettes. *Health Policy.* 125(1):75-82. DOI:10.1016/j.healthpol.2020.08.004.

John, R. 2008. Crowding out effect of tobacco expenditure and its implications on household resource allocation in India. *Soc Sci Med.* 66(6):1356-1367. DOI:10.1016/j.socscimed.2007.11.020.

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

John, R., Chelwa, G., Vulovic, V. & Chaloupka, F. 2019. *Using Household Expenditure Surveys for Research in the Economics of Tobacco Control. A Tobacconomics Toolkit.* Chicago, IL. Available: www.tobacconomics.org.

Jumrani, J. & Birthal, P.S. 2017. Does consumption of tobacco and alcohol affect household food security? Evidence from rural India. *Food Security.* 9(2):255-279. DOI:10.1007/s12571-017-0660-8.

Keino, S., Plasqui, G., Ettyang, G. & van den Borne, B. 2014. Determinants of stunting and overweight among young children and adolescents in sub-Saharan Africa. *Food and nutrition bulletin.* 35(2):167-178.

Kennedy, E. & Peters, P. 1992. Household food security and child nutrition: the interaction of income and gender of household head. *World Development.* 20(8):1077-1085. DOI:10.1016/0305-750x(92)90001-c.

Kerr, A. & Wittenberg, M. 2015. Sampling methodology and fieldwork changes in the October Household Surveys and Labour Force Surveys. *Development Southern Africa.* 32(5):603-612. DOI:10.1080/0376835X.2015.1044079.

King, C., Markowitz, S. & Ross, H. 2015. Tobacco Control Policies and Sudden Infant Death Syndrome in Developed Nations. *Health Econ.* 24(8):1042-1048. DOI:10.1002/hec.3090.

King, G. & Nielsen, R. 2019. Why Propensity Scores Should Not Be Used for Matching. *Political Analysis.* 27(4):435-454. DOI:10.1017/pan.2019.11.

Koch, S.F. 2018. Quasi-experimental evidence on tobacco tax regressivity. *Social Science & Medicine.* 196:19-28. DOI:10.1016/j.socscimed.2017.11.004.

Koch, S.F. & Tshiswaka-Kashalala, G. 2008. *Tobacco substitution and the poor*.

Kostakis, I. 2020. Does tobacco spending crowd-out the household budget? Preliminary results using nationwide survey data.

Lechner, M. 2011. The Estimation of Causal Effects by Difference-in-Difference Methods Estimation of Spatial Panels. *Foundations and Trends in Econometrics.* 4(3):165-224. DOI:10.1561/0800000014.

Lee, N. 2009. Measurement error and its impact on estimates of income and consumption dynamics. *Available at SSRN 1299330.*

Leibbrandt, M. & Levinsohn, J. 2014. Fifteen years on: Household incomes in South Africa. In *African Successes, Volume I: Government and Institutions.* University of Chicago Press. 333-355.

Leibbrandt, M., Woolard, I. & de Villiers, L. 2009. Methodology: Report on NIDS wave 1. *Technical paper.* 1.

Leibbrandt, M., Finn, A. & Woolard, I. 2012. Describing and decomposing post-apartheid income inequality in South Africa. *Development Southern Africa.* 29(1):19-34. DOI:10.1080/0376835x.2012.645639.

Leicester, A. & Levell, P. 2016. Anti-smoking policies and smoker well-being: evidence from Britain. *Fiscal Studies.* 37(2):224-257.

Leys, C., Ley, C., Klein, O., Bernard, P. & Licata, L. 2013. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology.* 49(4):764-766. DOI:10.1016/j.jesp.2013.03.013.

Lien, D.S. & Evans, W.N. 2005. Estimating the impact of large cigarette tax hikes the case of maternal smoking and infant birth weight. *Journal of Human Resources.* 40(2):373-392.

Linegar, D.J. & van Walbeek, C. 2018. The effect of excise tax increases on cigarette prices in South Africa. *Tob Control.* 27(1):65-71. DOI:10.1136/tobaccocontrol-2016-053340.

Lopez-Nicolas, A., Badillo-Amador, L. & Cobacho-Tornel, M.B. 2013. Will the European Union's new tobacco tax legislation lead to reductions in smoking prevalence? Evidence from a quasi-experiment in Spain. *Nicotine Tob Res.* 15(12):1963-1970. DOI:10.1093/ntr/ntt038.

Lorenz, E., Jenkner, C., Sauerbrei, W. & Becher, H. 2017. Modeling Variables With a Spike at Zero: Examples and Practical Recommendations. *Am J Epidemiol.* 185(8):650-660. DOI:10.1093/aje/kww122.

Machemedze, T., Kerr, A. & Wittenberg, M. 2014. Recalibrating the OHSs to adjust for sampling changes. *Cape Town: University of Cape Town, DataFirst.(Technical paper 28).*

MacKinnon, D.P. 2012. *Introduction to statistical mediation analysis.* Routledge.

MacKinnon, D.P., Lockwood, C.M. & Williams, J. 2004. Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate behavioral research.* 39(1):99-128.

MaCurdy, T.E. & Pencavel, J.H. 1986. Testing between competing models of wage and employment determination in unionized markets. *Journal of political Economy.* 94(3, Part 2):S3-S39.

Manivong, P., Harper, S. & Strumpf, E. 2017. The contribution of excise cigarette taxes on the decline in youth smoking in Canada during the time of the Federal Tobacco Control Strategy (2002-2012). *Can J Public Health.* 108(2):e117-e123. DOI:10.17269/cjph.108.5705.

Manthey, J., Shield, K.D., Rylett, M., Hasan, O.S., Probst, C. & Rehm, J. 2019. Global alcohol exposure between 1990 and 2017 and forecasts until 2030: a modelling study. *The Lancet.* 393(10190):2493-2502.

Manthey, J., Hassan, S.A., Carr, S., Kilian, C., Kuitunen-Paul, S. & Rehm, J. 2021. What are the economic costs to society attributable to alcohol use? A systematic review and modelling study. *Pharmacoeconomics.* 39(7):809-822.

Markowitz, S. 2008. The effectiveness of cigarette regulations in reducing cases of Sudden Infant Death Syndrome. *J Health Econ.* 27(1):106-133. DOI:10.1016/j.jhealeco.2007.03.006.

Markowitz, S., Adams, E.K., Dietz, P.M., Tong, V.T. & Kannan, V. 2013. Tobacco control policies, birth outcomes, and maternal human capital. *Journal of human capital.* 7(2):130-160.

Masa-Ud, A.G., Chelwa, G. & Van Walbeek, C. 2020. Does tobacco expenditure influence household spending patterns in Ghana?: Evidence from the Ghana 2012/2013 Living Standards Survey. *Tobacco Induced Diseases.* 18(June). DOI:10.18332/tid/120936.

Mayosi, B.M. & Benatar, S.R. 2014. Health and Health Care in South Africa — 20 Years after Mandela. *New England Journal of Medicine.* 371(14):1344-1353. DOI:10.1056/nejmsr1405012.

McGeary, K.A., Dave, D., Lipton, B. & Roeper, T. 2020. Impact of Comprehensive Smoking Bans on the Health of Infants and Children. *American Journal of Health Economics.* 6(1):1-38. DOI:10.1086/706622.

McKenzie, D.J. 2005. Measuring inequality with asset indicators. *Journal of Population Economics.* 18(2):229-260. DOI:10.1007/s00148-005-0224-7.

Mellor, J.M. 2011. Do cigarette taxes affect children's body mass index? The effect of household environment on health. *Health Econ.* 20(4):417-431. DOI:10.1002/hec.1598.

Mensch, B.S., Chuang, E.K., Melnikas, A.J. & Psaki, S.R. 2019. Evidence for causal links between education and maternal and child health: systematic review. *Trop Med Int Health.* 24(5):504-522. DOI:10.1111/tmi.13218.

Mosley, W.H. & Chen, L.C. 1984. An analytical framework for the study of child survival in developing countries. *Population and development review.* 10:25-45.

National Department of Health & ICF. 2019. *South Africa Demographic and Health Survey 2016*. Pretoria. Available: http://dhsprogram.com/pubs/pdf/FR337/FR337.pdf.

Neter, J. & Waksberg, J. 1964. A study of response errors in expenditures data from household interviews. *Journal of the American statistical Association.* 59(305):18-55.

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

Nguyen, N.-M. & Nguyen, A. 2020. Crowding-out effect of tobacco expenditure in Vietnam. *Tobacco Control.* 29(Suppl 5):s326-s330. DOI:10.1136/tobaccocontrol-2019-055307.

Nonnemaker, J. & Sur, M. 2007. Tobacco expenditures and child health and nutritional outcomes in rural Bangladesh. *Social Science & Medicine.* 65(12):2517-2526.

Nyagwachi, A.O. 2019. Essays on the economics of tobacco and alcohol control policy in Kenya.

Nyagwachi, A.O., Chelwa, G. & van Walbeek, C. 2020. The effect of tobacco- and alcohol-control policies on household spending patterns in Kenya: An approach using matched difference in differences. *Soc Sci Med.* 256:113029. DOI:10.1016/j.socscimed.2020.113029.

Öberg, M., Jaakkola, M.S., Woodward, A., Peruga, A. & Prüss-Ustün, A. 2011. Worldwide burden of disease from exposure to second-hand smoke: a retrospective analysis of data from 192 countries. *The Lancet.* 377(9760):139-146.

Pagan, A.R. & Hall, A.D. 1983. Diagnostic tests as residual analysis. *Econometric reviews.* 2(2):159-218.

Paraje, G. & Araya, D. 2018. Relationship between smoking and health and education spending in Chile. *Tob Control.* 27(5):560-567. DOI:10.1136/tobaccocontrol-2017-053857.

Parry, C.D. 2010. Alcohol policy in South Africa: a review of policy development processes between 1994 and 2009. *Addiction.* 105(8):1340-1345.

Pesko, M.F., Courtemanche, C.J. & Maclean, J.C. 2020. The effects of traditional cigarette and e-cigarette tax rates on adult tobacco product use. *Journal of risk and uncertainty.* 60(3):229-258.

Pichon-Riviere, A., Alcaraz, A., Palacios, A., Rodríguez, B., Reynales-Shigematsu, L.M., Pinto, M., Castillo-Riquelme, M., Torres, E.P. et al. 2020. The health and economic burden of smoking in 12 Latin American countries and the potential effect of increasing tobacco taxes: an economic modelling study. *The Lancet Global Health.* 8(10):e1282-e1294.

Pollak, R.A. 1969. Conditional demand functions and consumption theory. *The Quarterly Journal of Economics.* 83(1):60-78.

Probst, C., Parry, C.D.H., Wittchen, H.-U. & Rehm, J. 2018. The socioeconomic profile of alcohol-attributable mortality in South Africa: a modelling study. *BMC Medicine.* 16(1). DOI:10.1186/s12916-018-1080-0.

Pu, C.Y., Lan, V., Chou, Y.J. & Lan, C.F. 2008. The crowding-out effects of tobacco and alcohol where expenditure shares are low: analyzing expenditure data for Taiwan. *Soc Sci Med.* 66(9):1979-1989. DOI:10.1016/j.socscimed.2008.01.007.

Rambachan, A. & Roth, J. 2019. *An honest approach to parallel trends.*

Reddy, P., Meyer-Weitz, A. & Yach, D. 1996. Smoking status, knowledge of health effects and attitudes towards tobacco control in South Africa. *South African Medical Journal.* 86(11).

Reddy, P., Zuma, K., Shisana, O., Kim, J. & Sewpaul, R. 2015. Prevalence of tobacco use among adults in South Africa: Results from the first South African National Health and Nutrition Examination Survey. *South African Medical Journal.* 105(8):648. DOI:10.7196/samjnew.7932.

Reichman, N.E., Corman, H., Noonan, K. & Dave, D. 2009. Infant health production functions: what a difference the data make. *Health Econ.* 18(7):761-782. DOI:10.1002/hec.1402.

Reitsma, M.B., Kendrick, P.J., Ababneh, E., Abbafati, C., Abbasi-Kangevari, M., Abdoli, A., Abedi, A., Abhilash, E. et al. 2021. Spatial, temporal, and demographic patterns in

prevalence of smoking tobacco use and attributable disease burden in 204 countries and territories, 1990–2019: a systematic analysis from the Global Burden of Disease Study 2019. *The Lancet.* 397(10292):2337-2360.

Rozema, K. & Ziebarth, N.R. 2017. Taxing consumption and the take-up of public assistance: The case of cigarette taxes and food stamps. *The Journal of Law and Economics.* 60(1):1-27.

Rubin, D.B. 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology.* 66(5):688.

Ryan, A.M., Kontopantelis, E., Linden, A. & Burgess, J.F., Jr. 2019. Now trending: Coping with non-parallel trends in difference-in-differences analysis. *Stat Methods Med Res.* 28(12):3697-3711. DOI:10.1177/0962280218814570.

Sabia, J.J. 2008. Every breath you take: The effect of postpartum maternal smoking on childhood asthma. *Southern Economic Journal.* 75(1):128-158.

Sacks, G., Kwon, J. & Backholer, K. 2021. Do taxes on unhealthy foods and beverages influence food purchases? *Current nutrition reports.* 10(3):179-187.

Saffer, H., Dench, D., Grossman, M. & Dave, D. 2020. E-cigarettes and adult smoking: evidence from Minnesota. *Journal of risk and uncertainty.* 60:207-228.

Said-Mohamed, R., Micklesfield, L.K., Pettifor, J.M. & Norris, S.A. 2015. Has the prevalence of stunting in South African children changed in 40 years? A systematic review. *BMC Public Health.* 15(1):1-10.

San, S. & Chaloupka, F.J. 2016. The impact of tobacco expenditures on spending within Turkish households. *Tob Control.* 25(5):558-563. DOI:10.1136/tobaccocontrol-2014-052000.

Sánchez, A.M. & Gómez, A.G. 2024. Crowding out and impoverishing effect of tobacco in Mexico. *Tobacco Control.* 33(Suppl 2):s68-s74.

Sarkola, T., Kahila, H., Gissler, M. & Halmesmäki, E. 2007. Risk factors for out-of-home custody child care among families with alcohol and substance abuse problems. *Acta Paediatrica.* 96(11):1571-1576.

Sassi, F., Belloni, A., Mirelman, A.J., Suhrcke, M., Thomas, A., Salti, N., Vellakkal, S., Visaruthvong, C. et al. 2018. Equity impacts of price policies to promote healthy behaviours. *The Lancet.* 391(10134):2059-2070. DOI:10.1016/s0140-6736(18)30531-2.

Schreiber, J.B. 2021. Issues and recommendations for exploratory factor analysis and principal component analysis. *Res Social Adm Pharm.* 17(5):1004-1011. DOI:10.1016/j.sapharm.2020.07.027.

Sen, A. & Piérard, E. 2011. Estimating the effects of cigarette taxes on birth outcomes. *Canadian Public Policy.* 37(2):257-276.

Shea, J. 1997. Instrument Relevance in Multivariate Linear Models: A Simple Measure. *Review of Economics and statistics.* 79(2):348-352. DOI:10.1162/rest.1997.79.2.348.

Shrestha, N. 2021. Factor Analysis as a Tool for Survey Analysis. *American Journal of Applied Mathematics and Statistics.* 9(1):4-11. DOI:10.12691/ajams-9-1-2.

Simkins, C. 2004. *What happened to the distribution of income in South Africa between 1995 and 2001?*

Simon, D. 2016. Does Early Life Exposure to Cigarette Smoke Permanently Harm Childhood Welfare? Evidence from Cigarette Tax Hikes. *American Economic Journal: Applied Economics.* 8(4):128-159. DOI:10.1257/app.20150476.

Snow Jones, A., Miller, D.J. & Salkever, D.S. 1999. Parental use of alcohol and children's behavioural health: a household production analysis. *Health Economics.* 8(8):661-683. DOI:10.1002/(sici)1099-1050(199912)8:8<661::Aid-hec481>3.0.Co;2-o.

South African Medical Research Council. 2022. *Global Adult Tobacco Survey: South Africa, 2021. Factsheet*

South African Revenue Services. 2023. *Schedules to the Customs and Excise Act, 1964.*

StataCorp, L. Stata Statistical Software: Release 17. 2021. *College Station, Texas, United States of America.*

Statistics South Africa. 2001. *Consumer Price Index (CPI). Statistical Release P0141.4.*

Statistics South Africa. 2017. *Living conditions of households in South Africa: an analysis of household expenditure and income data using the LCS 2014/2015.* Statistics South Africa Pretoria.

Statistics South Africa. 2021. *Consumer Price Index (CPI) History.*

Stock, J.H. & Yogo, M. 2005. Testing for weak instruments in linear IV regression. *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg.* 80(4.2):1.

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 & Economic Statistics.* 20(4):518-529. DOI:10.1198/073500102288618658.

The World Bank Group. 2023. *World Development Indicators.*

Thomson, G.W., Wilson, N.A., O'Dea, D., Reid, P.J. & Howden-Chapman, P. 2002. Tobacco spending and children in low income households. *Tobacco Control.* 11:372-375.

Tominey, E. 2007. Maternal smoking during pregnancy and early child outcomes.

U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion & Office on Smoking and Health. 2014. *The Health Consequences of Smoking —50 Years of Progress: A Report of the Surgeon General.* Atlanta, G.A.

UNICEF. 1990. *Strategy for improved nutrition of children and women in developing countries.* UNICEF.

UNICEF. 2020. UNICEF Conceptual Framework on Maternal and Child Nutrition.

United Nations. 2003. *United Nations Treaty Collection: WHO Framework Convention on Tobacco Control*. Available: https://treaties.un.org/pages/ViewDetails.aspx?src=TREATY&mtdsg_no=IX-4&chapter=9&clang=_en [2023, 7 September].

United Nations. 2018. *Classification of Individual Consumption According to Purpose (COICOP) 2018*. New York.

Van Der Zee, K., Van Walbeek, C. & Magadla, S. 2020. Illicit/cheap cigarettes in South Africa. *Trends in Organized Crime.* 23(3):242-262. DOI:10.1007/s12117-019-09372-9.

van Walbeek, C. 2004. Tobacco control in South Africa in the 1990s: A mix of advocacy, academic research and policy. *South African Journal of Economic History.* 19(1-2):100-131. DOI:10.1080/10113430409511172.

van Walbeek, C. 2005. The economics of tobacco control in South Africa.

Van Walbeek, C. & Blecher, E. 2014. The economics of alcohol use, misuse and policy in South Africa. *Cape Town.*

Van Walbeek, C. & Chelwa, G. 2021. The case for minimum unit prices on alcohol in South Africa. *South African Medical Journal.* 111(7):680-684.

Vellios, N. 2022. *How big is the illicit cigarette market in South Africa?* Available: https://www.econ3x3.org/article/how-big-illicit-cigarette-market-south-africa [2023, 10 September].

Vellios, N. & Van Walbeek, C. 2016. Determinants of regular smoking onset in South Africa using duration analysis. *BMJ Open.* 6(7):e011076. DOI:10.1136/bmjopen-2016-011076.

Vellios, N., van Walbeek, C. & Ross, H. 2020. Illicit cigarette trade in South Africa: 2002–2017. *Tobacco Control.* 29(Suppl 4):s234-s242.

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

Victora, C.G., Adair, L., Fall, C., Hallal, P.C., Martorell, R., Richter, L., Sachdev, H.S., Maternal et al. 2008. Maternal and child undernutrition: consequences for adult health and human capital. *Lancet.* 371(9609):340-357. DOI:10.1016/S0140-6736(07)61692-4.

Vladisavljevic, M., Zubović, J., Jovanovic, O. & Đukić, M. 2024. Crowding-out effect of tobacco consumption in Serbia. *Tobacco Control.* 33(Suppl 2):s88-s94.

Walker, M.B., Tekin, E. & Wallace, S. 2009. Teen smoking and birth outcomes. *Southern Economic Journal.* 75(3):892-907.

Wang, H., Sindelar, J.L. & Busch, S.H. 2006. The impact of tobacco expenditure on household consumption patterns in rural China. *Soc Sci Med.* 62(6):1414-1426. DOI:10.1016/j.socscimed.2005.07.032.

Wehby, G.L., Prater, K., McCarthy, A.M., Castilla, E.E. & Murray, J.C. 2011. The impact of maternal smoking during pregnancy on early child neurodevelopment. *Journal of human capital.* 5(2):207-254.

WHO. 2018. *WHO global report on trends in prevalence of tobacco smoking 2000-2025.* World Health Organization.

WHO Framework Convention on Tobacco Control Secretariat. 2017. The who framework convention on tobacco control: an accelerator for sustainable development.

Williamson, J.M. 2010. Environmental Tobacco Smoke and Children's Health. *Eastern Economic Journal.* 36(4):423-449. DOI:10.1057/eej.2009.27.

Wittenberg, M. 2008. *Income in the October Household Survey 1994*. Cape Town.

Wittenberg, M. 2017a. Wages and Wage Inequality in South Africa 1994–2011: Part 1–Wage Measurement and Trends. *South African Journal of Economics.* 85(2):279-297. DOI:10.1111/saje.12148.

Wittenberg, M. 2017b. Wages and Wage Inequality in South Africa 1994-2011: Part 2 - Inequality Measurement and Trends. *South African Journal of Economics.* 85(2):298-318. DOI:10.1111/saje.12147.

Wittenberg, M. & Leibbrandt, M. 2017. Measuring Inequality by Asset Indices: A General Approach with Application to South Africa. *Review of Income and Wealth.* 63(4):706-730. DOI:10.1111/roiw.12286.

Wittenberg, M., Collinson, M. & Harris, T. 2017. Decomposing changes in household measures: Household size and services in South Africa, 1994–2012. *Demographic research.* 37:1297-1326.

Wooldridge, J.M. 2010. *Econometric analysis of cross section and panel data.* MIT press.

World Health Organization. 1995. *Physical status: The use of and interpretation of anthropometry, Report of a WHO Expert Committee.* World Health Organization.

World Health Organization. 2003. *WHO framework convention on tobacco control.* World Health Organization.

World Health Organization. 2010. *Global strategy to reduce the harmful use of alcohol.* World Health Organization.

World Health Organization. 2019a. *Global strategy to accelerate tobacco control: advancing sustainable development through the implementation of the WHO FCTC 2019-2025*. Geneva.

World Health Organization. 2019b. *Global status report on alcohol and health 2018.* World Health Organization.

World Health Organization. 2022. *Global alcohol action plan 2022–2030 to strengthen implementation of the Global Strategy to Reduce the Harmful Use of Alcohol: First draft. 2021*.

Xu, X., Pesko, M.F., Tynan, M.A., Gerzoff, R.B., Malarcher, A.M. & Pechacek, T.F. 2013. Cigarette price-minimization strategies by US smokers. *American journal of preventive medicine.* 44(5):472-476.

Yan, J. 2013. Prenatal smoking cessation and infant health: Evidence from sibling births. *Southern Economic Journal.* 80(2):299-323.

Yan, J. 2020. Healthy babies: does prenatal care really matter? *American Journal of Health Economics.* 6(2):199-215. DOI:10.1086/707831.

Yu, D. 2008. *The comparability of Income and Expenditure Surveys 1995, 2000 and 2005/2006*.

Yu, D. 2016. Factors influencing the comparability of poverty estimates across household surveys and censuses. *Development Southern Africa.* 33(2):145-165.

Yu, H., Engberg, J. & Scharf, D. 2019. One size fits all? Disentangling the effects of tobacco taxes, laws, and control spending on adult subgroups in the United States. *Subst Abus.* 40(1):87-94. DOI:10.1080/08897077.2018.1449050.

Zhang, Z. & Zheng, R. 2020. The Impact of Cigarette Excise Tax Increases on Regular Drinking Behavior: Evidence from China. *Int J Environ Res Public Health.* 17(9). DOI:10.3390/ijerph17093327.