An Economic Perspective on the Influence of Social Interaction on Risky Behaviours

Thesis Presented for the Degree of
Doctor of Philosophy

In the
School of Economics
University of Cape Town

By
Byela Tibesigwa

Supervisor
Associate Professor Martine Visser

August, 2012

Table of Contents

ABSTRACT IV

DEDICATION VI

ACKNOWLEDGEMENTS VII

CHAPTER 1: BACKGROUND, INTRODUCTION AND MOTIVATION TO RESEARCH PROBLEM 1

1.1 INTRODUCTION TO RESEARCH PROBLEM 1
1.2 RESEARCH OBJECTIVE, HYPOTHESIS AND CONTRIBUTION 2
1.3 AREA OF STUDY AND SOURCES OF DATA 3
1.4 ORGANISATION OF THE THESIS AND EXPECTED PAPERS 4
1.5 REFERENCES 5

CHAPTER 2: THE EFFECTS OF SOCIAL NORMS AND SOCIAL PRESSURE ON MULTIPLE PARTNERSHIP 8

ABSTRACT 8
2.1 INTRODUCTION 9
2.2 CONTRIBUTION AND ORGANISATION OF THE STUDY 10
2.3 A SYNOPSIS OF HIV/AIDS AND MULTIPLE PARTNERSHIP 11
2.4 THE SOCIAL NORM AND SOCIAL PRESSURE NEXUS 13
2.5 THEORETICAL FRAMEWORK 15
2.6 EMPIRICAL FRAMEWORK 19
2.6.1 Area of Study and Sources of Data 19
2.6.2 Variable Specification 19
2.6.2.1 Multiple Partnerships 19
2.6.2.2 Social Norms 20
2.6.2.3 Social Pressure 21
2.6.2.4 Social Groups 21
2.6.2.4.1 Prior Knowledge on Social Group Composition 22
2.6.2.4.2 Supporting Literature to our Prior Knowledge: Social Space and Similar Others 23
2.6.2.4.3 Social Group Composition 25
2.6.2.5 Covariates 28
2.6.3 Econometric Modelling 28
2.6.3.1 Econometric Model Identification 29
2.6.3.2 Final Models Used For Estimation 31
2.6.3.3 Instrumental Variable Strategy 32
2.6.3.4 Including the Population that is not Sexually Active 34
2.7 ESTIMATION RESULTS 36
2.7.1 Regression Results 36
2.7.2 Social Group Falsification Test 41
2.8 CONCLUSION 45
2.9 REFERENCE 47

Appendix 2.A: Characteristics Of Our Sample 57

Appendix 2.B: Distribution Of Social Groups 58

Appendix 2.C: Comparison Between Sexually Active And Sexually Inactive Population 59

Appendix 2.D: Results Of The First Stage CTSM 60

Appendix 2.E: Social Norm And Social Pressure Using Adjacent Social Group Age Bracket IV 61

Appendix 2.F: Social Norm And Social Pressure Results By Gender 63

Appendix 2.H: Social Norm And Social Pressure Using Placebo Groups 65

CHAPTER 3: 'KNOW YOUR EPIDEMIC': THE EFFECTS OF EXPECTED HEALTH AND CONTEXTUAL EFFECTS OF HEALTH UNCERTAINTY ON RISKY SEX 66

ABSTRACT 66
3.1 INTRODUCTION AND STRUCTURE OF THE STUDY 67
3.2 BACKGROUND, MOTIVATION AND CONTRIBUTION OF THE STUDY 69
3.3 THEORETICAL FRAMEWORK 72
3.4 EMPIRICAL FRAMEWORK 73
3.4.1 Area of Study 74
3.4.2 Variable Specification 74
3.4.2.1 Risky Sex Index 74
3.4.2.2 Expected Future Health 76
3.4.2.3 Uncertainty of Future Health 78
3.4.2.4 Expected Income 79
3.4.2.5 Covariates 81
3.4.3 Estimation and Identification Strategy 81
3.5 ESTIMATION RESULTS 86
3.5.1 Regression Results 86
3.5.2 Social Group Falsification Test 87
3.6 CONCLUSION 91
3.7 REFERENCES 94

Appendix 3.A: Characteristics Of The DHS Data 104
Appendix 3.B: Characteristics Of The Social Groups 105
Appendix 3.C: Construction Of The Risky Sex Index Using CFA 106
Appendix 3.D: Constructing The Expected Future Health Proxy 109
Appendix 3.E: Validating The Health Uncertainty Proxy 114
Appendix 3.F: Imputing Household Income Using Regression Methods 115
Appendix 3.G: Imputing Household Income By Wealth Index 122
Appendix 3.H: Results Of The First Stage CTSM 124
Appendix 3.I: Results Using Adjacent Social Group Age Bracket IV 125
Appendix 3.J: Placebo Social Group Using Adjacent Social Group Age Bracket IV 127
Appendix 3.K: Regression Output Using An Alternative Expected Income Proxy 128
Appendix 3.L: Regression Output For Eastern Africa And Southern Africa 129
Appendix 3.M: Results Using An Alternative Outcome Variable 130

CHAPTER 4: THE ROLE OF COMMUNITY SOCIAL DISORGANISATION IN CONTACT AND CONTACT RELATED CRIME 132

ABSTRACT 132
4.1 INTRODUCTION 133
4.2 REVIEW OF PREVIOUS EMPIRICAL RESEARCH 135
4.3 MOTIVATION FOR RESEARCH PROBLEM IN THE SOUTH AFRICAN CONTEXT 137
4.4 THEORETICAL FRAMEWORK 141
4.5 EMPIRICAL FRAMEWORK 144
4.5.1 Area of Study and Sources of Data 144
4.5.2 Variable Specification 144
4.5.2.1 Contact and Contact Related Crime 145
4.5.2.2 Family Disruption 145
4.5.2.3 Community Disadvantage 146
4.5.2.4 Residential Instability 148
4.5.2.5 Lack of Basic Services 148
4.5.2.6 Unemployment 148
4.5.3 Empirical Model Specification 148
4.6 ESTIMATION RESULTS 153

4.6.1 Descriptive Statistics 153
4.6.2 Fixed Effects Model Results 155
4.7 CONCLUSION 160
4.8 REFERENCE 162

Appendix 4.A: Distribution Of Crime Rates 168

Appendix 4.B: Family Disruption Statistics In South Africa 169

Appendix 4.C: Correlation Matrix Of The Regressors 170

Appendix 4.D: Types Of Crime And Likely Estimation Challenge 171

Appendix 4.E: Probit Results Of Household Crime Reporting 172

Appendix 4.F: Results Based On 2 Waves: 2002 And 2005 173

Appendix 4.G: Results Based On Individual Basic Services 174

CHAPTER 5: CONCLUSION: SYNOPSIS OF FINDINGS, POLICY IMPLICATIONS, AND AREAS FOR FUTURE STUDY 177

5.1 SYNOPSIS OF THE FINDINGS 177
5.2 POLICY IMPLICATION 178
5.3 AREAS FOR FUTURE RESEARCH 179

Abstract

This thesis presents the theme of social interaction and risk taking. There is a growing interest in incorporating the influence of social interaction in the economic modelling of human choices. Although the current economic literature is awash with the topic of social interaction a disproportionate bias is towards theoretical as opposed to empirical work. Of the existing empirical literature, a small proportion deals with sub-Saharan African (SSA) countries as case studies. This is quite surprising given that the cultural characteristic of close-knit social kinship and ties in this region makes it more likely that human choices will be influenced by social interaction. Further to this the existing empirical literature is abounds with criticism regarding the lack of sound methods to curb identification problems which include the effects of unobservables, selection bias, simultaneity and the related *reflection problem*.

Against this backdrop the thesis makes an empirical investigation into the role of social interaction on risk taking behaviour. The thesis consists of three self-contained papers, each with the aim of unearthing the influence of social interaction on risk taking. Two types of risky behaviours apparent in SSA are investigated. The first is risky sexual behaviour associated with the HIV infection. This risk is covered in two papers that constitute chapters 2 and 3. The second type of risky behaviour relates to crimes of social behaviour, specifically *contact crimes* and *contact related crimes*, and this is covered in one paper which forms chapter 4.

The thesis uses three main data sources. The first is the Cape Town Panel Survey (CAPS) dataset which concentrates on young adults living in the Cape Town Metropolitan area of South Africa. The second dataset is the Demographic and Health Survey (DHS) which includes adults from developing countries. Six DHS datasets, covering Kenya, Tanzania and Uganda from the Eastern African region and the datasets on Zimbabwe, Lesotho and Swaziland in the Southern African region are used. The final data source is Cape Town crime data which consists of crime statistics from the South African police service covering the Cape Town Metropolitan area.

The first paper titled "*The Effects of Social norms and Social Pressure on Multiple Partnerships*" investigates the influence of *social norms* and *social pressure* in influencing multiple partnerships. This paper is motivated by the persistence of multiple partnerships in SSA, a pivotal ingredient of heterosexual transmission of HIV in SSA. Furthermore recent qualitative research suggests that the aforementioned factors are responsible for the persistence in multiple partnerships. To the best of our knowledge this is the first study to measure this relationship. The main challenge with our estimation model is the existence of the identification problem which includes selection bias, unobservables, simultaneity and the related *reflection problem*. To overcome these challenges different estimation strategies are applied to CAPS, using both cross sectional and panel methods. In these different estimation strategies similar results were obtained: *social norms* and *social pressures* that occur as a result of social interaction are indeed pivotal in influencing multiple partnerships.

The second paper titled *'Know your Epidemic': The Effects of Expected Health and Contextual Health Uncertainty on Risky Sex*, examines the effects of *expected income*, *expected health* and the contextual effects of *health uncertainty* on risky sexual behaviour using a selection of DHS data. The motivation for the second paper is UNAIDS's "*know your epidemic, know your response"* campaign calling for countries to identify and understand the underlying contextual factors driving their local HIV/AIDS epidemic. The study is closely aligned to empirical studies of Oster (2009, 2012) and Bezabih *et al*., (2010). The novel feature of this study lies in the application of an alternative estimation strategy to existing studies to overcome the identification problem arising from selection bias, unobservables and simultaneity. The results from the selected DHS surveys generally reveal a significant negative relationship between *expected health* and risky sex and a significant positive relationship between *health uncertainty* and risky sex, supporting Mannberg's (2010) theoretical framework.

An equally important risky behaviour is violent criminal behaviour, especially in South Africa. Extending the theme of contextual effects, the CAPS dataset and the crime data are used to investigate how community social disorganisation, *family disruption*, *residential instability*, *community disadvantage*, *lack of basic services* and *unemployment*, can induce individuals in Cape Town Metropolitan area of South Africa to resort to *crimes of social behaviour*. This is the background of the third and final paper titled '*The Role of Community Social Disorganisation in Contact Crime and Contact Related Crime'*. The motivation of this paper is the fact that crimes especially *contact crimes* continues unabated in South Africa. In addition South Africa has complex and unique socially disorganised communities, an aftermath of the apartheid past and the current high HIV/AIDS prevalence. In this study the community is the unit of analysis. In order to establish the relationship the study needed to first circumvent four estimation challenges namely, excessive collinearity among the regressors, measurement error as a result of reporting bias, omitted variable bias, and the influence of community unobservables. In general the results reveal that the social factors are positive and significant in predicting both contact and contact related crime, but more so among the majority Black communities.

Keywords: Social interaction, Contextual factors, Risk Taking, HIV/AIDS, Sexual behaviour, Social norms, Social Pressure, Multiple partnerships, Health Uncertainty, Community social disorganisation, Crimes of social behaviour, Contact crime, Contact related crime

Dedication

This thesis is dedicated to my dear family.

Acknowledgements

This thesis has benefitted from the encouragement, support and constructive criticisms from numerous people to whom I would like to convey my sincere appreciation. First and foremost my heartfelt appreciation is extended to my supervisor Associate Professor Martine Visser for her encouragement, dedication, support and expertise that I needed to complete this thesis. For accepting my unconventional ideas and arming me with the necessary tools required to turn these unconventional ideas to what the thesis is today. She has indeed been a wonderful mentor to me throughout my thesis, pushing me and steering me to become a better researcher, while constantly believing in my capabilities and believing that I can do more. She really is the kind of a supervisor any PhD candidate ought to have.

I would like to thank the Environmental Policy and Research Unit (EPRU) at the University of Cape Town and the Environment for Development (EfD) for their support. I am a junior research fellow in EPRU thanks to my supervisor who is a senior research fellow in the unit. Being part of EPRU I had access to various resources and networks. On this note I would like to acknowledge the senior research fellows in EPRU, Professors Anthony Black, Associate Professor Edwin Muchaponda, Associated Professors Tony Leiman, Dr. Jane Turpie and Dr Mare Sarr for the capacity building especially during the EPRU research meetings.

I also had the pleasure and opportunity to meet Professor Thomas Sterner from University of Gothenburg during the sixth EfD Conference in Arusha, Tanzania in 2011. His thoughts and expertise during our discussions assisted me greatly. I am also indebted to Professor Peter Berck from the University of California who visited EPRU in 2012 and I had the privileged to engage with him in numerous discussions. Needless to say his comments were indeed valuable. Special thanks go to other EfD network, especially Dr. Andrea Mannberg for her theoretical frameworks that were adopted in this thesis and for her comments on the empirical application of her theories. A special thank you also goes to Dr. Mintweb Bezabih for her helpful comments in the final stages of this thesis.

From the University of Cape Town I am grateful to Professor Nicoli Nattrass's orientation and guidance at the beginning of this thesis. I also acknowledge the discussions I had with Dr. Obert Pimhidzai and Dr. Richard Musa, I benefited greatly from their comments and suggestions. I am also thankful to Lynn Woolfrey and the rest of the DataFirst team for all their assistance with data issues throughout the countless times I sat at DataFirst trying to make sense of the available data and the contents of the datasets. I also thank Lynn Woolfrey for editing this thesis. I would like to thank the School of Economics and the staff, especially Paula Bassingthwaighte, for all their support during my PhD journey. My thanks also goes to the support and encouragement I received from my fellow EPRU junior research fellows Sunday Adewara, Johane Dikgang, Kerry Bricks and Coretha Kombo. My gratitude is also extended to my fellow PhD Candidates Seedwell Hove, Theresa Mtenga, Gladys Gamare and George Mutasa.

I gratefully acknowledge the financial support I received from the World Bank's Margaret McNamara Memorial Fund. The fund also provided me with access to a wonderful network of women from various countries with similar ambitions to myself. I express my sincere gratitude for EPRU's financial assistance and all other financial support that was made available to me by my supervisor. I am also very thankful for the Carnegie Scholarship from the School of Economics and the Centre for Social Science Research. Lastly and most importantly I thank God Almighty.

Chapter 1
Background, Introduction and Motivation to Research Problem

"Economists have long been ambivalent about what social interactions constitute proper domain of the discipline. The narrower view has been that economics is primarily the study of markets, a circumscribed class of institutions in which persons interact through anonymous process of pricing. The broader view has been that economics is defined by its concern with allocation of resources and its emphasis on idea that people respond to incentives. In this view, economists may properly study how incentives shape all social interactions that affect the allocations of resources (Manski, 2000: p. 115). *'Within economics, the study of social interactions has expanded the domain of inquiry to incorporate many ideas that are traditionally associated with sociology. Social interactions analysis also extends the methodological individualism of economics in new directions through its focus on the feedback between individual behaviours and aggregate outcomes* (Durlauf *et al.*, 2010: p. 452).

1.1 Introduction to Research Problem

The human behavioural consequence of *social interactions*[1] has long been central to sociological, psychological and anthropological literature (Manski 1993; Bernheim 1994; Brocks and Durlauf 2004; Fletcher 2006; Lindbeck and Persson 2009). The argument put forth is that human behaviour is complex and multifaceted, hence to understand human behavioural outcomes it becomes paramount to go over and beyond atomistic research approaches[2] (UNAIDS[3] 1999; Macintyre *et al.* 2002; Grodner and Kniesner 2006; Durlauf *et al.* 2010).

Despite mounting evidence that support this hypothesis most economic research on human behaviour has in the past been methodologically individualistic[4]. This is usually based on the assumption that we as individuals act rationally as isolated agents (Morgan and Niraula 1995; Logan *et al.* 2002; Coates *et al.* 2008; Benefo 2010; Durlauf *et al.* 2010), a basic foundation of microeconomics. As noted by Duesenberry (1960) and famously cited: 'economics is all about how people make decisions, sociology is about how they have no decisions to make' (Duesenberry, 1960 cited in Morgan and Niraula, 1995 p. 542). Only relatively recent and some

[1] Empirical research on social interactions can be grouped into three hypotheses, *endogenous interactions/simultaneity* where an individual's behaviour varies with the behaviour of the group, *contextual/exogenous interactions* where an individual's behaviour varies with the exogenous characterises of the group members, and lastly *correlated effects* where individuals in a given group behave the same as a result of the similar environment (Manski 1993). The endogenous and exogenous interaction captures social influence, whereas the correlated effects explain a non-social interaction (Manski 2000; Moffit 2001).

[2] Research dating back to as early as the twentieth century has consistently demonstrated this argument. For example, Goldberger in 1916 found the risk of pellagra to be associated with the availability of fruits and vegetables in the village, independent of individual incomes, illustrating a link between individual and village level factors. Another example is in infectious diseases, where it is hypothesised that the risk of contracting an infectious disease is dependent on not only an individuals' immune system but on the community of residence (Diez-Roux 1998; Macintyre *et al.* 2002).

[3] United Nations Programme on HIV and AIDS (UNAIDS).

[4] …theories guiding most interventions are essentially cognitive and individualistic, and assume people have motivation and freedom to adopt protective actions (Coates *et al.* 2008: p. 676).

would say reluctantly, has the influence of social interaction on human behaviour gained an exponential interest in economic research. However, much attention has been devoted to theoretical research, while empirical research still falls behind[5].

The theme of this thesis therefore involves filling the current literature gap by making an empirical investigation into the role of social interaction on risk taking. In doing so, three areas which are underexplored in economic literature are investigated, while taking into consideration the identification problem that plagues models of social interactions. The first area explores the role of social norms in multiple partnerships, to assess whether they could explain the persistence of multiple partnerships in sub-Saharan Africa. As suggested by among others Gausset (2001); Takyi (2003); Akwara *et al.* (2003); Airhihenbuwa and Webster (2004); Lamptey *et al.* (2006); Wellings *et al.* (2006); UNAIDS (2007); UNAIDS (2008); Wilson and Halperin (2008); WHO[6] (2008); and Mah and Halperin (2010a).

The second area relates to the role of contextual factors in driving sexual behaviours associated with the risk of HIV/AIDS in sub-Saharan Africa, as argued by current literature, see for example Adimora and Schoenbach (2005); Peterman *et al.* (2005); Wellings *et al.* (2006); Gillespie *et al.* (2007); Wilson and Halperin (2008); Gabrysch *et al.* (2008); Nattrass (2009); Mah and Halperin (2010a); and Mannberg (2010). The third area looks at the role of community social disorganisation in influencing crimes of social behaviour. The focus here is whether community social disorganisation matter for crimes of social behaviour in South Africa as stated by current literature, see Emmett (2003); Ward (2007); SAIRR[7] (2010); and Daniels and Adams (2010).

1.2 Research Objective, Hypothesis and Contribution

This thesis consists of three self-contained papers, with each of the papers consisting of a research problem and an objective that falls within the overall theme of the thesis.

The first objective looks at the influence of *social norms* and *social pressure* on multiple partnerships. The hypothesis is that these will have a positive effect on the number of sexual partners, as a result of the endogenous social interaction. The main contribution of this chapter lays in the novel quantitative investigation into the relationship between *social norms*, *social pressure* and multiple partnerships. To the best of our knowledge this is the first such study to make such an empirical investigation.

[5] The dominant reason behind this trend is the difficulty in econometric modelling as a result of the identification problem inherent in models of social interactions (Brocks and Durlauf 2004; Fletcher 2006; Grodner and Kniesner 2006; Lindbeck and Persson 2009; Durlauf *et al.* 2010).

[6] World Health Organisation (WHO).

[7] South African Institute of Race Relation (SAIRR).

The second objective looks at the role of *expected income, expected health* and the contextual effect of *health uncertainty* on *risky sex*. The hypothesis here is that *expected health* will have a negative effect on risky sex while *health uncertainty* will have a positive effect. The novel feature of this study and the departure from current empirical studies lays in the application of an alternative estimation strategy to overcome the identification problem arising from selection bias, unobservables and simultaneity in the empirical model. In addition we also use new datasets (selected Demographic and Health Survey) in the empirical investigation.

The third objective looks at the influence of social disorganisation, specifically, *family disruption*, *residential instability*, *community disadvantage*, *lack of basic services* and *unemployment* on *contact crime* and *contact related crime*. In this objective the hypothesis is that communities that are socially disorganised are more likely to be dominated by these crimes. This study differs from current studies in that measurement errors, omitted variable bias and unobservables are control in the empirical model. In addition new datasets (Cape Area Panel Study and South African Police Service crime data) and a new study area (Metropolitan Cape Town area of South Africa) are used.

1.3 Area of Study and Sources of Data

As earlier indicated the thesis is based on an empirical investigation. Three data sources are utilised to aid this investigation: The Demographic and Health Survey (DHS), the Cape Area Panel Study (CAPS) and crime data. DHS[8] are cross sectional surveys conducted in developing countries with an interval of approximately five years which capture nationally representative data. The surveys focus on population, nutrition and health of women and men aged between 15-49 years. In addition information on household characteristics of respondents is also gathered. The recent surveys contain information on HIV/AIDS including knowledge around the epidemic and sexual relationships and for some countries the survey project even undertakes HIV testing and includes these results in the data. The thesis uses six DHS datasets, three from the Eastern Africa region which includes Tanzania, Kenya and Uganda, while the remaining three cover selected Southern Africa countries namely Swaziland, Lesotho and Zimbabwe[9].

CAPS[10] is a longitudinal study of young adults in the Metropolitan Cape Town area of South Africa. CAPS has a total of five waves, the first wave was conducted in 2002, while the most recent wave occurred in 2006. The survey captures information on young adults who were aged

[8] http://www.measuredhs.com

[9] We utilise the 2004 Tanzanian DHS, the 2003 Kenyan survey, from Uganda we employ the 2006 DHS, the 2006 survey of Swaziland is utilised, from Lesotho we use the 2004 DHS, and in Zimbabwe we utilise the 2005 DHS. The selection of the countries in each of the region is based on the availability of the variables we use.

[10] http://www.caps.uct.ac.za or http://www.datafirst.uct.ac.za

between 14-22 years in 2002 and follow up information gathered on these young adults since then. The CAPS datasets captures information on schooling, work, health and sexual matters of the young adults. The survey also includes information on young adults' household composition, with the most recent waves capturing information on the young adults' parents and community.

The crime dataset contains crime statistics from the South African Police Service (SAPS) collected at community level in the Metropolitan Cape Town area of South Africa between 2001/2002 and 2005/2006. The dataset records 33 types of crime, which can be grouped into the five main SAPS categories which are *contact crimes*, *contact related crimes*, *property related crimes*, *crimes detected by police* and *other serious crimes*. By utilising the different data sources both cross sectional and panel data methods of analysis are possible. In addition this approach permits us to observe the findings of the research while taking into consideration variation in geographic and demographic composition within and between data sources.

1.4 Organisation of the Thesis and Expected Papers

This thesis consists of five chapters. The current chapter began with an introduction and details of the main research areas of the thesis. The next three chapters consist of the three self-contained papers dealing with the aforementioned objectives and falling into the overall theme of the thesis. The last chapter provides a conclusion. Specifically, chapter 2 begins with the first objective and first paper, which is based on a quantitative investigation of the role of *social norms* and *social pressure* in influencing *multiple sexual partnerships* using both cross sectional and panel data methods on Cape Area Panel Study (CAPS) datasets. The chapter outlines the background, motivation, theoretical and empirical framework of the paper. Additionally, because each chapter constitutes a stand-alone paper, the chapter also includes a list of references applicable to the paper. Chapter 3 introduces the second paper answering the second research question. The chapter highlights the motivation for the study and the research methods used to investigate the relationship between *expected income*, *expected health* and the contextual effects of *health uncertainty* on *risky sex* using cross sectional methods and this is because the data (Demographic Household Survey - DHS dataset) does not permit us to make a panel data analysis.. The third paper is presented in chapter 4 where the role of community social disorganisation in influencing *contact crimes* and *contact related crimes* is explored, thus answering the third and final research question. Finally, similar to chapter 2, this paper uses panel data methods on CAPS datasets. The last chapter, chapter 5, synthesises the thesis findings from the three self-contained papers and provides concluding remarks.

1.5 References

Adimora, A. and Schoenbach, V. (2005). Social context, sexual networks, and racial disparities in rates of sexually transmitted infections, *The Journal of Infectious Diseases*, 191: 115-122.

Airhihenbuwa, C. and Webster, J. (2004). Culture and African contexts of HIV/AIDS prevention, care and support, *Journal of Social Aspects of HIV/AIDS Research Alliance*, 1 (1): 4-13.

Akerlof, G. and Kranton, R. (2005). Identity and the economics of organisations, *The Journal of Economic Perspective*, 19 (1): 9-32.

Akwara, P., Madise, N., and Hinde, A. (2003). Perception of risk of HIV/AIDS and sexual behaviour in Kenya, *Journal of Biosocial Science,* 35: 385-411.

Anderson, K., and Beutel, A. (2007). HIV/AIDS Prevention knowledge among youth in Cape Town, South Africa, *Journal of Social Sciences*, 3 (3): 143-151.

Benefo, K.. (2010). Determinants of condom use in Zambia: A multilevel analysis, *Studies in Family Planning*, 41(1): 19-30.

Bernheim, B. (1994). A theory of conformity, *Journal of Political Economy*, 102 (5): 841-877.

Blackmore, M. (2003). A panel data analysis of crime in South Africa, *SAJEMS* , 6 (3): 439-458.

Breetzke, G. (2010). A socio-structural analysis in the City of Tshwane, South Africa, *South African Journal of Science*, 106 (11/12): 1-7.

Brock, W., and Durlauf, S. (2004). Chapter 54 interaction-based models, *Handbook of Econometrics*, 5: 3297-3380.

Coates, T., Richter, L., and Caceres, C. (2008). Behavioural strategies to reduce HIV transmission: How to make them work better, *Lancet*, 372: 669-684.

Coleman, E. (2002). Promoting sexual health and responsible sexual behaviour: An introduction, *The Journal of Sex Research*, 39 (1): 3-6.

Daniels, D., and Adams, Q. (2010). Breaking with township gangsterism: The struggle for place and voice, *African Studies Quarterly*, 11 (4): 45-57.

Durlauf, S., Brock, W. Blume, L., and Ionnides, Y. (2010). Identification of social interactions, Handbook of Social Economics.

Elbers, C., Lanjouw, J. and Lanjouw, P. (2003). Micro-level estimation of poverty and inequality, *Econometrica*, 71 (1): 355-364.

Elster, J. (1989). Social norms and economic theory, *The Journal of Economic Perceptive*, 3 (4): 99-117.

Emmett, T. (2003). Social disorganisation, social capital and violence prevention in South Africa, *African Safety Promotion*, 1: 4-18.

Fenton, K., Johnson, A., and McManus, S. (2001). Measuring sexual behaviour: Methodological challenges in survey research, *Sexually Transmitted Infections,* 77: 84-92.

Fletcher, J. (2006). Social multipliers in sexual initiation decision among U.S high school students, *Demography*, 44(2): 373-388.

Gabrysch, S., Edwards, T. and Glynn, J. (2008). The role of context: Neighbourhood characteristics strongly influence HIV risk in young women in Ndola, Zambia, *Tropical Medicine and International Health*, 2 (13): 162-170.

Gausset, Q. (2001). AIDS and cultural practices in Africa: The case of the Tonga (Zambia), *Social Science and Medicine*, 52: 509-518.

Gillespie, G., Greener, R., Whitworth, J. and Whiteside, A. (2007). Poverty, HIV and AIDS: Vulnerability and impact in Southern Africa, *AIDS*, 21 (7): S1-S104.

Grodner, A. and Kniesner, T. (2006). Social interactions in labour supply, *Journal of European Economic Association*, 4 (6): 1226-1248.

Kawachi, I., Kennedy, B., and Glass, R.. (1999). Social capital and self-rated health: A contextual analysis, *American Journal of Public Health*, 89 (8): 1187-1193.

Kawachi, I., Kim, D., Coutts, A., and Subramanian, S. (2004). Commentary: Reconciling the three accounts of social capital, *International Journal of Epidemiology*, 33: 682-690.

Lamptey, P., Johnson, J., and Khan, M. (2006). The global challenge of HIV and AIDS, *Population Bulletin*, 61 (1): 1-24.

Landau, L., and Misago, J. (2009). Who to blame and what's to gain? Reflections on space, state and violence in Kenya and South Africa, *Africa Spectrum*, 44 (1): 99-110.

Lau U., Seedat M., Shahnaaz S., (2010), Research and Theory, Collective Violence in South Africa: Explaining the Explanations, *African Safety Promotion Journal*, 8 (2), 1-19.

Lindbeck, A., and Persson, M. (2009). Social interaction and sickness absence, CESIFO Working Paper, 2215.

Lurie, M., and Rosenthal, S. (2010a). Concurrent partnerships as a driver of the HIV epidemic in Sub-Saharan Africa?, The Evidence is Limited, *AIDS Behaviour*, 14: 17-24.

Logan, T., Cole, J., and Leukefeld, C. (2002). Women, sex, and HIV: Social and contextual factors, meta-analysis of interventions, and implications for practice and research, *Psychological Bulletin*, 128 (6): 851-885.

Louw, A. (2007). Crime and perceptions after a decade of democracy, *Social Indicators Research*, 81: 235-255.

Macintyre, S., Ellaway, A., and Cummins, S. (2002). Place effects on health: How can we conceptualise, operationalise and measure them? *Social Science and Medicine*, 55: 125-139.

Mah, T. and Halperin, D. (2010a). Concurrent sexual partnerships and the HIV epidemics in Africa: Evidence to move forward, *AIDS Behaviour*, 14: 11-16.

Mah, T. (2010b). Prevalence and correlates of concurrent sexual partnerships among young people in South Africa, *Sexually Transmitted Diseases,* 37 (2): 105-108.

Mah, T., and Halperin, D. (2010c). The Evidence for the role of concurrent partnerships in Africa's HIV Epidemics: A response to Lurie and Rosenthal, *AIDS Behaviour*, 14: 25-28.

Mannberg, A. (2010). Risky sex in a risky world: Sexual behaviour in a HIV/AIDS environment, *Scandinavian Journal of Economics*, 114(2): 296-322.

Manski, C. (1993). Identification of endogenous social effects: The reflection problem, *The Review of Economic Studies*, 60 (3): 531-542.

Manski, C. (2000). Economic analysis of social interactions, *The Journal of Economic Perspectives*, 14 (3): 115-136.

Manski, C. (2003). Identification problems in the social sciences and everyday life, *Southern Economic Journal*, 70 (1): 11-21.

Moffitt, R. (2003). Policy interventions, low-level equilibria and social interactions, *Journal of Applied Econometrics,* 18 (5): 45-82.

Morgan, P., and Niraula B. (1995). Gender inequality and fertility in two Nepali villages, *Population and Development Review*, 21 (3): 541-561.

Morris, M. (2010). Barking up the wrong evidence tree: Comment on Lurie and Rosenthal, "Concurrent partnerships as a driver of the HIV epidemic in Sub-Saharan Africa? The evidence is limited", *AIDS Behaviour*, 14: 31-33.

Nattrass, N. (2009). Poverty, sex and HIV, *AIDS Behaviour*, 13: 833-840.

Obermeyer, C. (2005). Refraining research on sexual behaviour and HIV, *Studies in Family Planning*, 36 (1): 1-12.

Peterman, T., Lindsey C. and Selik R. (2005). This place is killing me: A comparison of countries where the incidence of AIDS increased the most and the latest, *Journal of Infectious Diseases*, 191 (1): S123-S126.

South African Institute of Race Relation (SAIRR). (2011). First Steps to Healing the South African Family, SAIRR and Donaldson Trust.

Sampson, R. and Groves, W. (1989). Community structure and crime: Testing social disorganisation theory, *The American Journal of Sociology*, 94 (4): 774-802.

Sampson, R., Morenoff, J., and Earls, F. (1999). Beyond social capital: Spatial dynamics of collective efficacy for children, *American Sociological Review*, 64 (5): 633-660.

Takyi, B. (2003). Religion and women's health in Ghana: Insights into HIV/AIDS preventive and protective behaviour, *Social Science and Medicine*, 56: 1221-1234.

UNAIDS, (1999). Sexual behavioural change for HIV: Where have theories taken us?, UNAIDS Best Practice Collection, Geneva, Switzerland.

UNAIDS, (2007). UNAIDS expert consultation on behaviour change, 25-26 September 2006, Joint United Nations Programme on HIV/AIDS, International Conference Centre, Geneva, Switzerland.

UNAIDS, (2008). 2008 Report on global aids epidemic - preventing new HIV infections: The key to reversing the Epidemic, Geneva, Switzerland.

UNAIDS, (2009). AIDS epidemic update, Joint UNAIDS and WHO, Geneva, Switzerland.

UNAIDS (2010). Joint action for results: UNAIDS outcome framework, 2009-2011, Joint United Nations Programme on HIV/AIDS Geneva, Switzerland.

Ward, C. (2007). Neighbourhood indicators: Monitoring Child rights and well-being at small area level in monitoring child wellbeing concept and context, http://www.hsrcpress.ac.za.

Ward, C., and Bakhuis K. (2010). Intervening in children's involvement in gangs: Views of Cape Town's young people, *Children and Society*, 24: (50-62).

Weir S., Tate J., Zhusupov B., and Boerma J. (2004). Where the action is: Monitoring local trends in sexual behaviour, *Sexually Transmitted Infections,* 80: ii63-ii68.

Wellings, K., Collumbien, M., Slaymaker, E., Singh, S., Hodges, Z., Patel, D., and Bajos, N. (2006). Sexual behaviour in context: A global perspective, *Lancet,* 368: 1706-1728.

WHO (2002). World report on violence and health; Summary, Geneva, Switzerland.

WHO (2008). Manual for estimating the economic costs of injuries due to interpersonal and self-directed violence, WHO and Department of Health and Human Services Centres for Disease Control and Prevention.

Whyte, E. (2010). *Aluta Continua*: The struggle continues in South Africa - Against violent crime, *Dialogue*, 7 (1): 1-30.

Wilson, D. and Halperin, D. (2008). "Know your epidemic, know your response": A useful approach, if we get it right, *Lancet*, 372: 423-42.

Chapter 2
The Effects of Social Norms and Social Pressure on Multiple Partnership

Abstract

Even though antiretroviral treatment is becoming more efficient and available to susceptible populations, new HIV infections and HIV/AIDS related death still occur. This is particularly the case in sub-Saharan Africa. It is therefore problematic that sexual risk taking, in terms of multiple sex partners persist in spite of increasing levels of HIV awareness and knowledge. Sexual transmission of HIV is still the main mode of transmission in sub-Saharan Africa, and multiple sex partners have been shown to be especially important for the spread of the disease. Although a substantial body of research highlights the importance of *social norms* and *social pressure* in terms of influencing the number of sexual partners, to our knowledge no one has conducted a quantitative analysis of the relationship. In this vein, this study examines the role of *social norms* and *social pressure* on *multiple partnerships*. We derive a simple theoretical model of sexual behaviour in the presence of HIV/AIDS to derive testable hypotheses, and test these on the Cape Area Panel Study (CAPS) data. In order to deal with identification problems such as unobservables, selection bias and the reflection problem in our empirical analysis, we employ an instrumental variable approach with both cross-sectional and panel data methods on the linear-in-means and non-linear in means model. Overall our results show that *social norms*, proxied by the average number of sex partners in the social reference group, have a positive and significant influence on the individuals' choice of number of sex partners, and that this effect is stronger, the stronger is the *social pressure*.

Keywords: Social interaction, HIV/AIDS, Social norms, Social Pressure, Multiple partnerships

"I will take a lot of pressure from the boys. They will tease and make funny jokes and tell me that having one girlfriend is the same as having no one at all." "Other people will think that you do not have a game (if not having multiple girlfriends). You do not know how to treat the girls" (Ragnarsson *et al.*, 2010, p. 3). *"There are those who discourage you when you have one girlfriend, because they say if one leaves you; you will be 'uzakusokola esisihumane' (struggling bachelor), you will struggle since you do not have a girlfriend; such names"* (Selikow *et al.* 2009, p.109).

2.1 Introduction

Recent qualitative research suggests that *social norms* and the related *social pressure* are responsible for the persistence in multiple partnerships[11] in sub-Saharan Africa, for example Ragnarsson *et al.* (2009); Selikow *et al.* (2009) and Ragnarsson *et al.* (2010). The current consensus is that multiple partnerships is the main sexual behaviour driving heterosexual transmission of HIV in sub-Saharan Africa (Horton *et al.* 2008; Wilson and Halperin 2008; WHO 2008, UNAIDS 2009; Mah and Halperin 2010a). This implies that *social norms* and *social pressure* that support and maintain multiple partnerships greatly hinder the achievement of the '*Zero new infection*' vision promoted by UNAIDS[12]. As such it becomes worthwhile to undertake an empirical investigation into the effects of social norms and pressure on the practice of *multiple partnerships*.

In the current study the empirical investigation is based on a simplified theoretical framework of Mannberg and Sjögren (2010). In our estimation model *multiple partnerships*[13] is our outcome variable. Our main regressors are the *social norm* and *social pressure*. To capture the former the mean number of sex partners within the social group[14] is used as a proxy following the current literature, while the absolute deviation from the mean is used as a proxy for *social pressure*. The main challenge with our estimation model is the existence of the identification problem usually present in models of social interactions (Manski 1993; Blume *et al.* 2010). These include the effects of selection bias, unobservables, simultaneity and the related *reflection problem* caused by the social interaction regressors (Manski 1993; Blume *et al.* 2010; Giorgi *et al.* 2010). To avoid the identification problem alternative estimation strategies are employed following the work of Manski (1993, 2000, 2003);

[11] Although multiple sexual partners appears to be a dominant trend among men in SSA in general (Wellings *et al.* 2006; Selikow *et al.* 2009; Ragnarsson *et al.* 2010), this trend is however gaining popularity among women (Selikow *et al.* 2009; Mah 2010b).

[12] It is increasingly recognised that for the UNAIDS vision '*Zero New HIV Infection*' to be realised there needs to be a revolution in sexual behaviour (UNAIDS 2011). In the case of SSA this implies a decrease in multiple partnerships, since it is the dominant sexual behaviour associated with the risk of HIV/AIDS.

[13] This is defined as the *number of sex partners in the last 12 months*. Because this information does not allow us to distinguish concurrent partnership from serial partnership, we refer to this as *multiple partnerships*.

[14] Our social groups comprise of individuals with the same age, race, gender and income. We do not included physical geography so as to avoid geography related unobservables. In addition, the increases in residential mobility and technological advancement in communication have to some extent weakened the importance of physical geography in social interaction. For this literature refer to the works of Subrahmanyann *et al.* (2008); Wellman *et al.* (1996); Csanyi and Szendroi (2004); Hu and Wang (2011); Young and Rice (2011); Thelwall (2008) and Singla and Richardson (2008).

Moffit (2001); Grodner and Kniesner (2005); Brock and Durlauf (2004); Durlauf and Ioannides (2006); Munshi and Myaux (2006); Etile (2007); Blume *et al.* (2010) and Giorgi *et al.* (2010). These estimation strategies include using cross-sectional and panel data methods, different model specifications (*linear-in-means* and *non-linear-in-means* models) and a menu of instrumental variables. The first instrument is the *trend setter of the norms IV,* the second is the *adjacent social group age bracket IV* and the last instrument is the conventional instruments based on *one year la*g and *two year lag* instruments.

These different estimation strategies are applied to the Cape Area Panel Study (CAPS) datasets consisting of young adults and within different populations (a *sexually active population* and a *sexually inactive population*). Falsification tests following Lavy and Schlosser (2007) were performed to test our social group composition. In addition, after providing a background to the validity of our instruments the Anderson canon correlation statistics of underidentification and Sargan-Hansen and Amemiya-Lee Newey statistics test of overidentification are employed. These different estimation strategies showed similar results: social norms and social pressures that occur as a result of social interaction are indeed pivotal in influencing *multiple partnerships*.

2.2 Contribution and Organisation of the Study

In support of the qualitative literature, a substantial body of literature has highlighted the importance of social norms in influencing sexual behaviour in sub-Saharan Africa[15]. For example see Gausset (2001); Takyi (2003); Akwara *et al.* (2003); Airhihenbuwa and Webster (2004); Latkin *et al.* (2005); Lamptey *et al.* (2006); Wellings *et al.* (2006); Wilson and Halperin (2008); WHO (2008); UNAIDS (2008) and Mah and Halperin (2010a). Yet, despite this strong consensus, our literature review did not identify any quantitative study that has investigated this relationship. This study therefore addresses the current literature gap, by seeking to determine whether the aforementioned factors can help explain *multiple partnerships* in SSA. Therefore the main contribution of this study is in its novel quantitative investigation of the relationship between *social norms*, *social pressure* and *multiple partnerships*. To the best of our knowledge this is the first study of its kind. Secondly, we make a methodological contribution to the empirical analysis of social interaction, especially given that social interaction[16] models in general lack consensus on how to best examine interdependence

[15] Researchers agree that African AIDS is sustained through a complex interaction of social and cultural processes (Takyi 2003: p. 1223). …in SSA, sexual activity appears to be driven by socio-cultural beliefs and practices…the ability of individuals to sustain safer sexual behaviours may largely depend upon societal sexual norms and practices (Akwara *et al.* 2003: p. 386). Culture plays a vital role …particularly in Africa, where values of extended family and community significantly influence individual behaviour (Airhihenbuwa and Webster 2004: p. 4).

[16] Empirical research on social interactions can be grouped into three hypotheses, *endogenous interactions/simultaneity* where an individual's behaviour varies with the behaviour of the group, *contextual/exogenous interactions* where an individual's behaviour varies with the exogenous characterises of the group members, and lastly *correlated effects* where individuals in a given group behave the same as a result of the similar environment (Manski 1993). The endogenous and exogenous interaction captures social influence, whereas the correlated effects explain a non-social interaction (Manski 2000; Moffit 2001).

between individuals given the existing identification problem (Manski 2000; Moffit 2001; Grodner and Kniesner 2006). Hence current empirical studies are often criticised for the lack of sound econometric methods[17].

The remainder of this paper is structured as follows. The subsequent section highlights the relationship between *multiple partnership* and HIV/AIDS infections. Section 2.4 provides a brief literature review on *social norms* and *social pressure*. After that, section 2.5 outlines the theoretical model that guides our econometric modelling. Section 2.6 begins by defining our outcome variable and the main regressors. Thereafter the empirical model and our estimation strategy to overcome the identification challenge are described. This is followed by detailing the estimation results in section 2.7. Finally, section 2.8 offers the concluding remarks.

2.3 A Synopsis of HIV/AIDS and Multiple Partnership

While it is a given that the HIV/AIDS epidemic remains a worldwide emergency, the epidemic has been evolving differently in different regions of the world (UNAIDS 2009). In sub-Saharan Africa heterosexual sex has become the main mode of HIV transmission (Dunkle *et al.* 2008; Wilson and Halperin 2008; Horton and Pam 2008; UNAIDS 2009, 2010a; Lurie and Rosenthal 2010a). One of the main arguments for this evolution besides medical advancement is the persistent lack of change in sexual behaviour[18] (Akwara *et al.*, 2003; Wellings *et al.*, 2006; Kongnyuy and Wiysonge 2007; Morris 2010). Recent HIV/AIDS statistics show that 80% of new infections occurred through sexual intercourse (UNAIDS 2010a).

As such there is general consensus on the fundamental role played by sexual behaviour in driving this heterosexual transmission (UNAIDS 2007, 2009; Horton and Pam 2008; Wilson and Halperin 2008; WHO 2008; Reniers and Watkins 2009; Mah and Halperin 2010a). There is however an on-going debate regarding the role played by multiple concurrent partnerships in HIV transmission, which has most notably been captured in the articles of Mah and Halperin (2010a, 2010c); Lurie and Rosenthal (2010a, 2010b); Morris (2010) and Epstein (2010). This debate centres on the perceived lack of sufficient evidence to support the role of concurrency in HIV transmission. Nevertheless, the current research is motivated by the role played by *social norms* and the associated *social pressure* in influencing *multiple partnerships*. Furthermore, the pivotal role of sexual behaviour in driving HIV has been highlighted by the UNAIDS (2007):

> A quarter of a century of global experience has produced some promising innovations….Despite advances in prevention, treatment, care and support…., fundamental role of behaviour in spread of HIV is increasingly clear. Remarkable regional differences in intensity and scope of pandemic serve as powerful reminders that social and

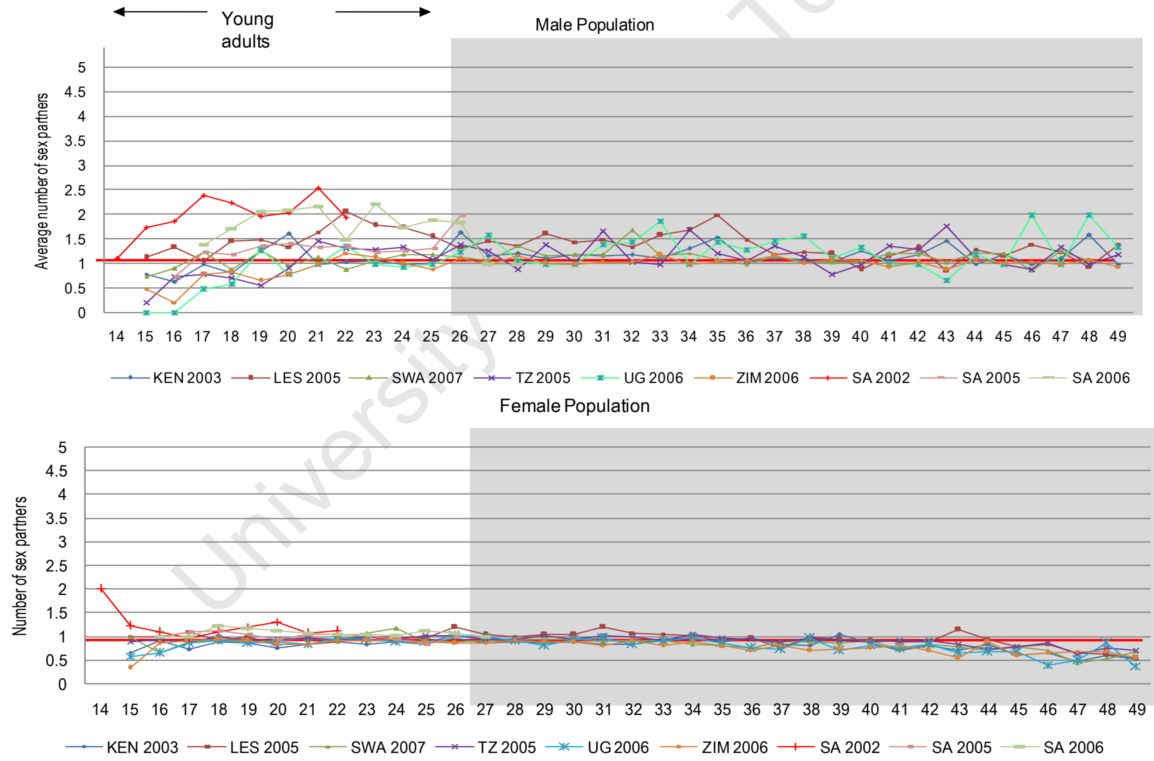
[17] See Manski (1993, 2000); Moffit (2001); and Grodner and Kniesner (2006) for some of the critiques.

[18] Sexual behaviour is probably responsible for much of differences in heterosexual epidemics among countries as well as for equally large differences among regions and demographic groups… (Akwara *et al.* 2003: p. 9).

> cultural factors ultimately shape the impact of HIV….Although we continue to place great hope in opportunities to employ male circumcision, microbicides, vaccines and other approaches, we recognise that failure to sustain a focus on behaviour change threatens to undermine benefits of such advances. Unfortunately, fostering behaviour change is not an easy task. It demands a persistent commitment to meeting the diverse and changing needs of individuals, and to addressing the characteristics of their social, cultural and physical environments. It is a collaborative process and an urgent imperative (UNAIDS 2007: p. i).

Figure 2.1 uses CAPS (wave 1 (2002), wave 3 (2005) and wave 4 (2006)) and Demographic and Health Survey (DHS) data to plot the trend in *multiple partnerships* in the selected countries, where the top panel is based on the male population while the bottom graph represents data on women. The CAPS dataset consists of data on young adults residing in the Metropolitan area of Cape Town. In order to make comparison between these datasets plausible comparisons with DHS data is limited to young adults residing in urban areas. This is represented by the un-shaded region in both graphs, where on average men have more than one sex partner at a time, while women have on average one sex partner[19].

Figure 2.1: Number of Sex partners in the last 12 months by Gender and Country[20]



[19] When we dig deeper we find that multiple partnerships is more prevalent among the population with the following relationship status: *have a partners but not living together with their partner*, and some evidence among the *divorced*, *have a partners but living together* and *never married*, and less prevalent among the *married* population in the DHS datasets.

[20] KEN represents Kenya, LES is Lesotho, SWA is Swaziland, UG is Uganda, ZIM is Zimbabwe, SA 2002 is based on CAPS wave 1, SA2005 is based on CAPS wave 3 and SA2006 is based on CAPS wave 4.

2.4 The Social Norm and Social Pressure Nexus

As earlier mentioned, the recent qualitative research of Ragnarsson *et al.* (2009); Selikow *et al.* (2009) and Ragnarsson *et al.* (2010) suggests that *social norms* and the related *social pressure* are responsible for the persistence in *multiple partnerships*. A pivotal contribution in these studies is the '*real man*' identity arising from the influence of social groups. Specifically, a '*real man*' is viewed as one who has several sexual partners, including one official partner, secondary partners and casual weekend partners. This behaviour is known and supported by the social group, and if a man deviates from this norm he is emasculated and seen as a '*lesser man*' and even faces being ostracised from the social group. This highlights the powerful and indeed disturbing influence of *social norms* and *social pressure* in supporting and maintaining *multiple partnerships*.

The literature on norms can be traced back to the social psychological disciplines of social interaction and human behaviour as early as 1900s (Cialdini and Trost 1998). At its core the social psychological literature offers the following description of a norm:

> A construct that has widespread usage because it helps describe and explain human behaviour… Norms have been conceptualised in a variety of ways. Summer (1906) wrote of 'folkways' -habitual customs exhibited by a group because they were originally expedient in meeting basic needs. Sherrif (1936) described norms as jointly negotiated rules for social behaviour, the 'customs, traditions, standards, rules, values, fashions, and all other criteria of conduct which are standardised as a consequence of the contact of individuals (Cialdini and Trost 1998: p. 152).

Social psychologists believe that social norms constitute the main source of social influence, and it is social interaction that causes people to conform to set norms (Cialdini and Trost 1998). Interestingly conformity is likely to occur even if the set social norms clash with personal norms, contradicting the expected *rational* behaviour of individuals. The main reason for this *irrational* behaviour is the fear of sanction resulting from non-conformity (Fisher 1988; Bernheim 1994; Akerlof 1997; Cialdini and Trost 1998; Munshi and Myaux 2006; Fletschner and Carter 2008; Blume *et al.* 2010). That is, individuals respond to *social pressure* to conform to *social norms* that will allow them to avoid sanction. It is this *social pressure* that will lead an individual to revise his/her behaviour in line with the prevailing *social norm* and adopt seemingly *irrational* behaviour. For example Cialdini and Trost (1998) and Bernheim (1994) state that:

> Thus we conform to others when perceived or real pressure from them causes us to act differently from how we would act if alone (Cialdini and Trost 1998: p. 152).

> Most social scientists agree that individual behaviour is motivated in large part by social factors such as the desire for prestige, esteem, popularity, or acceptance. A large body of sociology, psychological, and anthropological research supports the view that these factors are widespread and that they tend to produce conformism, social groups often penalise individuals who deviate from accepted norms (Bernheim 1994: p. 842).

The qualitative literature on *multiple partnerships* and *social pressure* that inspired this study also points to the need for conformity as a result of fear of sanctions (Selikow *et al.* 2009; Ragnarsson *et al.* 2009; Ragnarsson *et al.* 2010). For instance Ragnarsson *et al.* (2010) in highlighting the role of social pressure makes the following conclusion:

> Within peer groups, the pressure to live up to set norms further reinforced the meaning and status of the player. If a man adopted an alternative form of masculine ideal, he would risk being emasculated and thought not to have what it takes to be a real man according to prevalent norms in this specific context and group of men… Strong social pressure within male core groups to pursue and maintain these concurrent sexual relationships and temporary sexual encounters existed and helped legitimise specific behaviours that the player represented (Ragnarsson *et al.* 2010: p. 4).

To date a rapidly growing body of economic literature has used *social norms* to analyse various aspects of human behaviour, although the literature is skewed in favour of theory over empirical application. Some examples of economic disciplines that have analysed the effects of *social norms* include labour economics (Moffit 2001; Vendrik 2001; Grodner and Kniesner 2005; Akerlof and Kranton 2005; Grodner and Kniesner 2006; Mannberg and Sjögren 2010), psychology (Bamberg and Moser 2007), education (Giorgi *et al.* 2010), environment (Elster 1989), juvenile behaviour (Evans *et al.* 1992; Gaviria and Raphael 2001; Lundborg 2006; Fletcher 2009), entrepreneurship (Meek *et al.* 2009), and health (Etile 2007; Manski 2000; Munshi and Myaux 2006; Trogdon *et al.* 2008).

Related to the current study Fisher (1988) gives a literature review on how *social norms* affect risky sexual behaviour. The author makes reference to interviews that were conducted among heterosexual college students about sexual intercourse and condom use. These interviews revealed that students feared rejection (sanctions) that would result from non-conformity to group expectations. This fear meant it was easier for the students to have unprotected intercourse rather than discuss prevention methods. Another example given by the author is in culturally driven minority groups, where it is difficult for partners to negotiate safe sexual intercourse since the norms governing these minority groups determine this as inappropriate, and deviating from such norms could lead to sanctions.

As previously indicated, while our study is novel in its quantitative modelling of the effects of *social norms and social pressure* on multiple partnerships, the research question is not a new phenomenon. A large body of qualitative research has found *social norms* and *social pressure* to be in conflict with and undermine HIV/AIDS prevention methods (see for instance Latkin (2005), Ragnarsson *et al.* (2009), Selikow *et al.* (2009) and Ragnarsson *et al.* (2010). In addition, a recent quantitative study by Mah (2010b) suggested that *social norms* could be driving the variations in multiple concurrent partnerships in Cape Town (South Africa) and called for probing of this area. Further to this there is a general consensus that the African HIV/AIDS epidemic in general is driven by the underlying norms guiding sexual behaviour (Gausset 2001; Takyi 2003; Akwara *et al.* 2003; Airhihenbuwa and Webster 2004; Lamptey *et al.* 2006; Wellings *et al.* 2006; UNAIDS 2007, 2008; Wilson and Halperin 2008; WHO 2008; Mah and Halperin 2010a), providing the motivation for exploring this phenomenon in this study.

2.5 Theoretical Framework

The theoretical framework presented in this section guides our empirical analysis by outlining a theory that highlights the underlying social interaction channels that influence *multiple partnerships*. The framework is loosely based on a theoretical model developed by Mannberg and Sjögren (2010)[21]. Mannberg and Sjögren (2010), extends previous theoretical work by Vendrik (2003) and Akerlof and Kranton (2005) in terms of including both social norms, and *social pressure* into a dynamic model for behavioural choice and investments in identity, and show that, if the social pressure is strong, social norms that are in conflict with personal preferences may both give rise to conformity and to a change in preferences. The theoretical model developed in the current study is a much simpler representation of behaviour. However, in spite of the admittedly simple framework, it allows us to derive testable hypotheses on how social norms and the strength of these norms (social pressure) affect behaviour in terms of the choice to have many sexual partners. The choice to engage in sexual activity is naturally not the same as decisions related to consuming other risky goods such as cigarettes or alcohol. First and foremost, it does take two to tango, and sexual activities are therefore perhaps best represented in a bargaining model. However, we will assume that the individual can choose sex freely. This assumption will allow us to derive analytically tractable results and is perhaps motivated by the fact that men in sub-Saharan Africa still hold the power to choose when it comes to sex. Secondly, sexual decisions are in general made in the heat of the moment, and thus often plagued by myopia. However, using a hyperbolic utility function would not change our results significantly.[22] We therefore treat the individual as having time consistent time preferences.

Let us start by assuming that all individuals belong to a predetermined, and thus exogenously given, social group (g). To simplify the analysis, the individual is assumed to only derive utility from sexual activity, health and from a feeling of "*belonging*" to the social group, where the last component is assumed to depend on the individual's sexual behaviour. Hence, the individual is assumed to

[21] This is the only theory that our literature review identified that addresses social pressure and social norms. Other theories looked only at social norms, hence making this theory most compatible with our study.The theory founded by Mannberg and Sjögren (2010) was originally developed for labour economics to explain the increase in female labour supply. The intuitive reasoning for the increase in female labour supply is changes in gender norms. The theory is founded on psychological literature that emphasises the importance of personal identities and social interaction in human behaviour.

[22] If the individual has hyperbolic time preferences and the social cost/benefit of sexual consumption occurs in the future while the private benefit occurs in the present, a naive individual will conform relatively less to the socially prescribed behavior, than an individual without present biased preferences.

maximize utility in terms of choosing the optimal number of sexual partners (y). We assume that there is an offer on sex in a pool of exogenous supply of attractive and willing sexual partners. In this pool of available sexual partners each individual chooses the number of sex partners[23]. For simplicity, we also assume that nothing else contributes to an individuals' utility[24]. The utility function is assumed to be additively separable and made up of both a private utility and a social utility component. The former is referred to as the private utility as it captures only individual choices, while the latter comprises of the influence from choices of members in the social group through social interaction. Private utility is represented by a standard utility function:

$$U = u_1(y) + \beta u_2(y) \tag{2.1}$$

where, $u_1'(y) > 0, u_1''(y) < 0, u_2'(y) < 0$ and $u_2''(y) < 0$. In other words, we assume that the number of sex partners increases utility in time period 1 but reduces it in time period 2 in terms a reduction in health.[25] Hence, the individual maximizes utility by equating the marginal utility of sexual consumption in time period 1, $u_1'(y)$, to the discounted marginal health cost in time period 2, $\beta u_2'(y)$.

The social utility component captures the utility derived from interacting with other members in the social group. Recall that social norms are rules governing social behaviour within a group as a consequence of social interaction. Through this social interaction the individual's decision is affected by the prevailing social norms in the social group. In this model the social norm is captured by the average number of sex partners in the group denoted by $\bar{y}_g$. Conformity to social norms is driven by either the fear of social sanctions for non-compliance, or by the desire to belong to the social group. In both cases, the perceived behavioural heterogeneity for a given activity in the social group will affect the degree of social pressure that the individual experiences to comply with the norm. A functional form that captures both the presence of a social norm and the degree of social pressure is given as:

$$-\frac{1}{\gamma_g} \cdot \lambda(y_i - \bar{y}_g) \tag{2.2}$$

where γ_g is a measure of the spread of the distribution of sexual behaviour in the social group and where $\lambda(y_i - \bar{y}_g) \geq 0$ is the social sanction function. It is assumed that $\lambda_{y_i}'(y_i - \bar{y}_g) < 0$ if $y_i < \bar{y}_g$, while $\lambda(0) = \lambda_{y_i}'(y_i - \bar{y}_g) = 0$ if $y_i \geq \bar{y}_g$. In other words, an increase in the number of sex partners

[23] This approach simplifies the analytics by excluding the sexual bargaining process between sex partners, where each individual evaluates, anticipates and incorporates their sex partners' behaviours into their own choice set. See Mannberg (2010) and Mannberg (2012).

[24] For example the costs associated with searching for a sex partner, availability of different types of sexual partners and preference for type of sexual partners to maintain simplicity.

[25] It should be noted that, to be correct, the reduction in future health is not certain since probability of acquiring a sexually transmitted disease from an unprotected sexual act is relatively small.

reduces the social cost if the individual has less sexual partners than the average in the social group, and has no effect if the individual has more sexual partners. In other words, there is no "*bliss*" or cost associated with having "*too many*" sexual partners. We further assume that $\lambda''_{y_i y_i}(y_i - \bar{y}_g) < 0$. Hence, the marginal cost of deviating from the social norm increases the farther away from the mean the individual is. Finally, we assume that $\lambda'_{\bar{y}_g}(y_i - \bar{y}_g) > 0$ and $\lambda''_{y_i \bar{y}_g}(y_i - \bar{y}_g) < 0$ for $y_i \leq \bar{y}_g$. In other words, an increase in the average number of sexual partners in the social group increases the cost of having few sexual partners. Concerning the parameter $\bar{y}_g$, we have that a reduction in the heterogeneity (a reduction in the spread) of sexual behaviour within the social group increases the cost of diverging from the social norm. $\frac{1}{\gamma_g}$ thus measures the social pressure in the social group. Combining the private utility and social utility we obtain the total utility function which is depicted in equation (2.3).

$$U^s = u_1(y_i) - \frac{1}{\gamma_g} \cdot \lambda(y_i - \bar{y}_g) \quad + \beta u_2(y_i) \tag{2.3}$$

Equation (2.3) is maximised with respect to y_i to derive the first order condition which is presented in equation (2.4).

$$u'_1(y_i) - \frac{1}{\gamma_g} \cdot \lambda'_{y_i}(y_i - \bar{y}_g) = -\beta u'_2(y_i) \tag{2.4}$$

In other words, the individual maximizes utility such that the marginal disutility of reduced future health is balanced by the direct marginal utility of sexual activity and the avoided social cost arising from non-complying behaviour. The question asked in this paper relates to how social norms and social pressure affects sexual risk taking. In terms of our simple model, this implies how changes in γ_g and $\bar{y}_g$ affect the optimal choice of y_i. Let us therefore differentiate equation (2.4) with respect to y_i, $\bar{y}_g$ and γ_g respectively. This gives us,

$$\frac{\partial y_i}{\partial \bar{y}_g} = \frac{\frac{1}{\gamma_g} \cdot \lambda''_{y_i \bar{y}_g}(y_i - \bar{y}_g)}{\left(u''_1(y_i) - \frac{1}{\gamma_g} \cdot \lambda''_{y_i y_i}(y_i - \bar{y}_g) + u''_2(y_i)\right)} > 0 \tag{2.5}$$

$$\frac{\partial y_i}{\partial \gamma_g} = -\frac{\frac{1}{\gamma_g^2} \lambda'_{y_i}(y_i - \bar{y}_g)}{\left(u''_1(y_i) - \frac{1}{\gamma_g} \cdot \lambda''_{y_i y_i}(y_i - \bar{y}_g) + u''_2(y_i)\right)} < 0 \tag{2.6}$$

since $u''_1(y)$, $u''_2(y)$, $\lambda''_{y_i y_i}(\bar{y}_g - y_i) < 0$ by assumption. In other words, an increase in the average number of sex partners in the social group increases incentives to have many sexual partners, while an increase in the spread of the number of sexual partners (heterogeneity within the group) reduces pressure to comply with the norm and thus incentives to have many sexual partners. Equation (2.6) naturally implies that $\partial y_i / \partial \frac{1}{\gamma_g} > 0$. The interaction effect between $\bar{y}_g$ and γ_g is given by equation (2.7).

$$\frac{\partial^2 y_i}{\partial \bar{y}_g \gamma_g} = \frac{-\frac{1}{\gamma_g^2} \cdot \lambda''_{y_i \bar{y}_g}(y_i - \bar{y}_g)\left(u''_1(y_i) - \frac{1}{\gamma_g} \cdot \lambda''_{y_i y_i}(y_i - \bar{y}_g) + u''_2(y_i)\right) + \frac{1}{\gamma_g^2} \cdot \lambda''_{y_i y_i}(y_i - \bar{y}_g)}{\left(u''_1(y_i) - \frac{1}{\gamma_g} \cdot \lambda''_{y_i y_i}(y_i - \bar{y}_g) + u''_2(y_i)\right)^2} < 0 \qquad (2.7)$$

Hence, an increase in the social pressure, in terms of a reduced in the spread of sexual partners in the social group, increases the marginal benefit of complying with the social norm. Equations (2.5)-(2.7) constitute the theoretical hypotheses to be tested in the empirical model. It is worth noting that the model presented here involves decision making regarding number of sex partners on the condition that there is an offer on sex. That is we assume that there is an exogenous supply of attractive and willing sexual partners, and given this pool of sexual partners each individual chooses the number of sex partners. This approach is based on the modelling ideas of Mannberg (2010) and Mannberg (2012) and excludes the sexual bargaining process between sex partners. We do acknowledge that sex and sexual relationships in general are more complex. In the literature most authors make assumptions and acknowledge that their models are a mere simplification of sexual behaviour, for example (Philipson and Posner, 1995; Mannberg, 2010; Oster (2012) and Mannberg (2012)[26]. Similarly, we also acknowledge that our model is clearly a simplification of these sexual relationships in so far that we have disregarded the complex bargaining process between sexual partners, the search costs of sexual partners, coercion between partners, availability of different types of sexual partners and preference for type of sexual partners (see Philipson and Posner 1995; Schroeder and Rojas 2002; Mannberg 2012 for some of these discussion). Ideally sexual behaviour, which is different from other behaviour such as smoking, as it involves more than one participant, is likely to be better presented as a bargaining game between sex partners. For example a game where each individual incorporates their sex partners' or potential sex partners' various risky sexual behaviours such as the possibility of other sex partners into their choice set. However, this falls outside the scope of the current paper[27]. Overall given the aforementioned assumptions and

[26] For instance Philipson and Posner (1995) note that, 'The modelling of human behaviour as a reflection of a rational calculus of gains and losses may not provide a complete satisfactory explanation of individuals' actions or even of the aggregate outcomes of such actions. Such modelling however provides valuable insights into the understanding of human phenomena, insights often inaccessible to traditional modes of social analysis, sexual behaviour is no exception' (Philipson and Posner 1995: p. 835).

[27] We do believe that the sexual bargaining process modelled as a theoretical game, and not included in this current paper, offer a great area for future research when combined together with social norms and social pressure related to number of sex partners. Especially given the presence of asymmetric sexual power, particularly in the sub-Saharan Africa context, see Christensen (1998); Wojcicki and Malala (2001); Harvey et al. (2002); Pettifor et al (2004); Wingood and DiClemente (2000); Jewkes and Morrell (2010); Wechsberg et al. (2010) and Oster (2012). The aforementioned literature highlights the lack of sexual power among women irrespective of their social or economic position. Most importantly the literature shows that women's negotiating power is likely to further decrease with decreases in economic status or upon increasing economic dependence on men. For instance Pettifor et al. (2004), referring to a study of a Xhosa township, note that 'the study showed pervasive male control over almost every aspect of women's early sexual experiences, enacted in part through violent and coercive sexual practices' (Pettifor et al 2004: p.1996). Along similar lines, Jewkes and Morell (2010) state that 'In a given relationship for example, a

simplification of the model, the main contribution of our model is that it describes how *social norms* and the associated *social pressure* will affect an individuals' choice of number of sex partners. However, we are unsure how the results will change when some of the assumptions are changed, for example the introduction of sexual bargaining process and asymmetric power into our model, hence our reason for highlighting this as an important area for future research.

2.6 Empirical Framework

The section describes our estimation strategy. Firstly a brief description of the data is provided. Thereafter, the proxies used to capture our main variables are characterised. The social groups which are the source of social interaction are then defined. The section then ends by providing a description of our econometric modelling strategy to circumvent the identification problem.

2.6.1 Area of Study and Sources of Data

The data utilised is the Cape Area Panel Study[28] (CAPS). CAPS is a longitudinal study of young adults in Metropolitan Cape Town, South Africa. The datasets capture information on young adults schooling, work, health and sexual attitudes and behaviour. There are a total of four waves of CAPS data, the first wave was conducted in 2002, while the most recent wave occurred in 2006[29]. The study used wave 1 (2002), wave 3 (2005) and wave 4 (2006). The utilisation of different CAPS datasets allows us to employ both cross sectional and panel data methods of analysis. Additionally, this permits us to observe the findings in different geographical and demographical settings. Table 2.A1 in Appendix 2.A shows the characteristics of CAPS dataset.

2.6.2 Variable Specification

2.6.2.1 Multiple Partnerships

Ideally an outcome variable that captures multiple concurrent partnerships in line with the qualitative literature is needed. However, the only variable available in all our datasets is *number of sex partners in the last 12 months*. This variable only captures *multiple partnerships* without distinguishing

man may expect to lead and control sexual relations and his woman partner to comply and he may feel entitled to have sex with other women, but expect her to remain faithful. There is strong evidence that gender power inequality in relationships, which is a cause of intimate partner violence, places women at enhanced risk of HIV infection' (Jewkes and Morell 2010: p. 2). Likewise, Wechsberg et al. (2010) states that 'the highly unequal sociocultural and economic status faced by many South African women places them at a disadvantage in their ability to be sexually assertive with regard to practising safer sex, discussing sexual risk, negotiating condom use with primary male partners and avoiding sexually coercive situations (Wechsberg et al. 2010: p. 133).

[28] http://www.caps.uct.ac.za or http://www.datafirst.uct.ac.za

[29] A recent wave was conducted in 2009 (wave 5), however this data is not yet available for public use.

whether this is multiple concurrent partnership or serial partnership[30]. Hence our outcome variable in this study is *multiple partnerships*. It is worth noting that the literature on sexual behaviour has alluded to the fact that men on average have more sex partners than women. See for example Wellings et al. (2006) who offers a global view of sexual behaviour including the differences in number of sex partners between male and females. Some of the reasons for these discrepancies are the influence of sexual relationship between men and commercial sex workers or between men and younger girls, what is commonly referred to as 'sugar daddies'. See Hunter (2002), Kaufman and Stavrou (2004), Leclerc-Madlala (2008). Going from the literature to the data, we can easily show how it is possible that men have more sex partners than women. The CAPS datasets consist of young adults. The average age is between 15 to 26 years and the average age gap between sexual partners is between 4 to 5 years. As such it is possible that the males in the datasets are having sexual relationships with younger females who are below the cut off age (younger than 15), that is, sugar daddy relationships. Ragnarsson et al (2009) and Ragnarsson et al (2010) show that often men prefer younger sex partners, additionally Selikow et al (2009) show that sexual relationship do occur in young people even as young as 13 years old. Another possibility could be the fact that these men are having relationships with individuals above the cut off age (older than 26), that is, sugar mummies or even sex workers.

Nevertheless this variable is not without flaws. Firstly, the variable is likely to experience *recall bias*, given that it is easier for a monogamous individual to remember the number of partners they have had in the last 12 months, than for those who are usually in serial or concurrent relationships, hence creating a bias towards those who are monogamous. Secondly, there is usually over reporting of number of sex partners in males and under reporting in females especially in SSA (Dinkelman *et al.* 2007; Wellings *et al.* 2006), which makes our outcome variable prone to *reporting bias*. Our strategy to overcome this bias is to use an alternative outcome variable in our regression which is a binary variable where one represents whether a respondent has more than one partner and zero otherwise. A third kind of bias which is seldom addressed in the literature is the *age bias* brought about by age and experience, that is, the older you are the more likely you are to have more sex partners (Bezabih *et al.* 2010). To address these flaws we normalise our dependent variable by dividing the variable with age.

2.6.2.2 Social Norms

In empirical application *social norms* are commonly represented as endogenous social interaction (Grodner and Kniesner 2006; Fischer and Huddart 2008; Blume *et al.* 2010), where the average occurrence of the behaviour is used as a proxy for *social norms* (Durlauf and Ioannides 2010). For

[30] Concurrent partnerships are sexual relationships with more than one person. This is contrasted with sequential or serial partnerships whereby an individual engages in a sexual relationship with only one partner, with no overlap in time with subsequent partners (Mah and Halperin 2010a; p. 12).

example Etile (2007)'s measure of *social norms* is the average weight in the reference group. Along similar lines Trogdon *et al.* (2007) uses the average weight of peers as the proxy. Giorgi *et al.* (2010) follows the trend and adopts the average subject chosen by the group, while Munshi and Myaux (2006) proxy the norm with the average contraceptive usage in the community. In Gaviria and Raphael (2001) and Fletcher (2009)'s work the authors use the average number of students who smoke as a proxy, while Lundborg (2006) measured group effects by using the average number of students who participated in binge drinking. Following these approaches *social norm* is defined as the average number of sex partners in a social group.

2.6.2.3 Social Pressure

While there exists a rich theoretical and empirical literature on *social norms*, the same cannot be stated about *social pressure*. The current empirical literature covers subjective measures of *social pressure*. Maassen *et al.* (2004) for example assesses the effect of subjective norms, and subjective *social pressure* on smoking behaviour. Mahonen *et al.* (2010) looks at the effects of perceived normative pressure on adolescent attitudes towards immigrants. While Urizar *et al.* (2010) measures the effects of socio-cultural pressure to be thin on eating disorders and Helfert *et al.* (2011) explores the effects of appearance related *social pressure* on body dissatisfaction in adolescents[31]. Because our datasets do not contain any subjective social pressure, an objective social pressure is an alternative measure. Therefore in looking for an objective proxy measure for *social pressure* social psychological literature is utilised since the perspective that *social pressure* can influence behaviour is a social psychological phenomenon[32]. Given the social psychological perspectives outlined in section 2.4 on the relationship between *social norms* and *social pressure*, and the fact that the proxy for the *social norms* is the *mean behaviour*, the *deviation from the mean*[33] is used as a proxy for *social pressure*.

2.6.2.4 Social Groups

First and foremost, it is important to point to the fact that in this paper we create a proxy social group using the available information in our dataset. We do this mainly because we do not have any information on social group composition in our datasets. In modelling these social groups we refer to qualitative research on social norms related to number of sex partners. Secondly and also worth noting is the fact that the literature on social interaction is relatively new in the economics field and gaining an exponential interest in economic research. However, much attention has been devoted to theoretical research, while empirical research still falls behind. The dominant reason behind this

[31] A common trend in these studies is that they depict a positive relationship between social pressure and the behaviour in question.

[32] We rely on an objective proxy because the current datasets do not contain any subjective measures of social pressures.

[33] This proxy is commonly used to measure volatility of financial instruments in the field of finance.

trend is the fact that social interaction models are perhaps the most difficult to identify (Brocks *et al*., 2001; Fletcher *et al*., 2006; Grodner *et al*., 2006; Lindbeck *et al*., 2009; Durlauf *et al*., 2010). This is further exacerbated by the fact that social interactions models in general lack consensus on how to best examine interdependence between individuals (Manski, 1993, 2000; Moffit, 2001; Grodner *et al*., 2006). Hence the current empirical studies are overflowing with critiques regarding lack of sound econometric theory (See Manski 1993, 2000; Moffit 2001; Grodner *et al* 2006 for some of the critiques). Currently, some of the most influential literature on the application of social interaction literature include the work of Manski (1993, 2000, 2003); Moffit (2001); Grodner *et al*., (2005); Brock *et al*., (2004); Durlauf *et al*., (2006) and Blume *et al*., (2010).

The social interaction literature states that for the identification of social interaction, *social norms* and *social pressure*, to be plausible the social groups need to be appropriately defined (Manski 2000; Etile 2007; Blume *et al*., 2010). However, the current data, like most data used in the analysis of social interaction, is *deficit data* in that it does not contain any information on the social groups which are the source of *social norms* and *social pressure* affecting multiple partnership choices. Manski (2000) notes that under such circumstance one needs to relay on a *prior knowledge* on the mechanism that constitute social interaction and use this *prior knowledge* to make assumptions regarding the social group composition given the *deficit data*. Despite this strong consensus, most of the current studies on social interaction that use such *deficit data* lack this *prior knowledge*. For example De Giorgi *et al*., (2010) note that the definition of groups varies from the most comprehensive groups such as same race groups to very restrictive groups such as roommates. This lack of *prior knowledge* is likely to lead to the improper definition of social groups and thereby lack of identification. This is a major source of measurement error in studies of social interaction (Manski 1993; Lundborg 2006; De Giorgi *et al*., 2010; Fletcher 2010; Durlauf and Ioannides 2010; Blume *et al*., 2010). To aid in the appropriate definition of social groups this study relies on a *prior knowledge* grounded on the social interaction and qualitative literature which motivated the quantitative analysis of the current research problem relating multiple partnerships to the existing *social norms* and *social pressure*.

2.6.2.4.1 Prior Knowledge on Social Group Composition

The *prior knowledge* concerning our research problem relating *social norms* and *social pressure* to multiple partnerships is founded in qualitative research. The compelling evidence in these qualitative studies suggests that the social groups that influence multiple partnerships are likely to consist of individuals with *similar* socioeconomic attributes. See Hoff *et al*., (2002); Liljeros *et al*., (2003); Carter *et al*., (2007); Kohler *et al*. (2007); Selikow *et al*., (2009); Ragnarsson *et al*., (2009); Ragnarsson *et al*., (2010); Lynch *et al*., (2010); Townsend *et al*., (2011) and Gilbert and Selikow (2011). The rationale behind this is that individuals who are socioeconomically *similar* are likely to interact on matters related to sexual behaviour. In simple terms the probability of relational ties is likely to increase as the characteristics of individuals increase in similarity.

For instance, Kohler *et al.* (2007) in their review of social networks and HIV/AIDS risk perception states that: "…the networks with whom respondents discuss issues of family planning and AIDS are characterised by a tendency to discuss these topics with others who are perceived to be similar ('like me')" (Kohler *et al.* 2007: p. 4). In a similar vein, Ragnarsson *et al.*, (2009) and Ragnarsson *et al.*, (2010) describes the social groups as consisting of males with similar socioeconomic conditions. Like Ragnarsson *et al.*, (2009) and (2010), Lynch *et al.*, (2010) referring to the 'real man' (where masculinity is characterised by multiple partners) states that this masculinity is constructed by various factors such as race and class. See also Latkin and Knowlton (2006); Carter *et al.*, (2007); Kohler *et al.* (2007); Selikow *et al.*, (2009); Gilbert and Selikow (2011) and Townsend *et al.*, (2011) for similar literature.

Smith and Christakis (2008) in a review of the impact of social networks on health state that: "Sociologist are particularly well positioned to consider and shed light on how race, gender and class might interact with network processes" (Smith and Christakis 2008: p. 406). They go further to state that "…individuals may select their network partners on the basis of qualities such as sex, socioeconomic status…" (Smith and Christakis 2008: p.417). Finally, Liljeros *et al.*, (2003) note that: "These studies demonstrate that assertive interaction is structured across sociological variables, i.e., people are more likely to sexually interact with people from the same social class, age group and ethnic group" (Liljeros *et al.*, 2003: p. 190).

2.6.2.4.2 Supporting Literature to our Prior Knowledge: Social Space and Similar Others

As previously stated it is important to have a *prior knowledge* of the formation of the social groups. The previous section identified the social groups under the concept of *similar* others based on qualitative insights. In this section we refer to additional literature that supports the formation of such a social group. We begin by identifying a *social space* and relate this *social space* to the *similar* others social group concept.

Social Space

In this sub-section *social space* is invoked to help understand the mechanisms behind social interactions. First and foremost it is important to note that most literature states that social interaction occurs when individuals are close together within an abstract *social space* (Akerlof 1997; Etile 2007; Durlauf and Ioannides 2010). It is this *social space* that in turn influences their decision making process (Akerlof 1997; Conley and Topa 2002; Grodner *et al.*, 2006; Etile 2007; Durlauf and Ioannides 2010). From a sociological perspective the definition of a *social space* is: "man's social position is the totality of his relationships towards all groups of population and within each of them, towards its members" (Sorokin, 1927 cited in Prandy and Lambert 2003: p 399). On similar lines Bourdieu (1987) states that "The social space is constructed in such a way that the closer the agents, groups, institutions which are situated within this space, the more common properties they have and the more distant the fewer (Bourdieu 1987: p16). It is this social proximity between individuals that is likely to lead to social interaction. Examples of these social proximities include social networks or social groups (Akerlof, 1997).

Secondly, it is important to note that this *social space* may contain many dimensions (for example physical geography, ethnicity, gender or education)[34] that may at times overlap with each other. Because of the fact that the *social space* can comprise of these many dimensions it becomes pivotal to identify the dimensions that are significant in defining the social interaction in the research problem (Durlauf and Ioannides 2010). Relevant to this study, this implies identifying the dimensions that determine the social groups related to our research problem of *social norms*, *social pressure* and multiple partnerships.

Thirdly, it is also worth noting that *social space* implies that physical geography is not a prerequisite for social interaction. Indeed some sociologists are of the opinion that physical geography has become redundant in today's social interaction needs (Conley and Topa 2002). This sociological view is now supported by a large body of literature. For example, Manski (2000) notes that advancement in modern telecommunication technology such as cell phones, internet and social networks has drastically diminish the role of physical geography in social interaction. Likewise, Urry (2003) notes that the average distance (physical geography) between people in their social networks has drastically increased as a result of motorisation, urbanisation, airline deregulation, advancement in internet and telecommunication. Along similar lines, Xu *et al.*, (2010) notes that social networks have expanded due to increases in migration.

Similar others

The aforementioned qualitative studies argue that these *social norms* and *social pressure* related to multiple partnership occur amongst *similar* others[35]. The same observation is made in social interaction literature arguing that in general individuals who are *similar* are likely to interact socially. While sociologists refer to this *similar* other tendency as inbreeding or homophily, economists refer to this as positive matching or assortative matching (Conley and Topa 2002). The most recent

[34] For example Case *et al.*, (1991) in their study use city blocks as the reference groups, while Conley and Topa (2002) constructs a social space based on physical geography, ethnicity similarity and socioeconomic similarity. Grodner and Kniesner (2005) group is based on individuals who are similar in age, family structure and location. Munshi *et al.*, (2006) uses religion in their group formation. On the other hand Fletcher (2009) and Giorgi *et al.*, (2010) use classmates as groups, while Etile (2007) and Grodner (2006) use individuals of the same occupation, gender and age as the group characteristics.

[35] The concept of *similar* others is distinguishable from *significant* others, whereas *similar* others is based on similar characteristics, *significant* others introduces ties from family or friends (Etile, 2007).

studies at the nexus of social group and *similar others* can be found in Etile (2007) when analysing the norms related to body weight. Other studies that have also relied on the *similar* others concept are peer effect studies where the group is usually represented by classmates, under the assumption that students interact because they share *similar* courses/classrooms. See De Giorgi *et al.*, (2010); Fletcher (2010) and Lundborg (2006). There are groups that have been formed based on religious similarity see Munshi (2006). Vendrick (2003) has also alluded to the fact that social groups are made of individuals with *similar* characteristics such as age, education and number of children. Others such as Liljer *et al.*, (2003) have noted that individuals are likely to interact if they are of the same social class, age group and ethnic group, while Carter *et al.*, (2001) notes age, gender and religious similarity.

The importance of *similar* others is accentuated in social network research. For instance Moody (2001) notes that social network research focus on peer group similarity where groups are *similar* in multiple dimensions. Conley and Topa (2002) note that: "even causal observation suggests that personal networks may be stratified along specific socio-demographic attributes such as race, ethnicity, religious affiliation, language, age, race, gender and education levels, in other words, agents are likely to draw a disproportionate share of their social contacts among sets of people that are very similar to themselves" (Conley and Topa 2002: p. 309).

The principle behind *similar* others has also been applied to peer educators so as to change risky behaviours within certain groups, such as men who have sex with men, commercial workers and drug users. Such peer educators are selected on the basis of sharing *similar* attributes to the target groups. These attributes usually include age, gender, socio-economic background and ethnicity. By using such peer educators with *similar* attributes, these interventions have proven to be successful in reducing the risky behaviours of the target group. See Lugalla *et al.*, (2004) and Latkin and Knowlton (2006).

2.6.2.4.3 Social Group Composition

To summarise this section, in unfolding the literature, it became apparent that proper definition of social group improves identification and eliminates measurement errors in models of social interaction. Secondly, the literature shows that individuals interact in a social space and that this social space encompasses varies dimensions. Thirdly, the literature also shows that physical geography is not a prerequisite for social interaction to take place. Finally, the literature also shows that the social interaction related to *social norms* and *social pressure* on multiple partnerships occurs within individuals with similar socioeconomic characteristics.

Guided by the above insights we construct the social groups based on *similar* others using the following dimensions: age, race, income and gender giving us a total of 95 social groups. This is on the assumption that individuals who are *similar* based on these attributes are likely to interact socially in relation to norms related to multiple partnerships. Moreover, the importance of these socioeconomic similarities is also evident in the literature regarding gender and sexual behaviour in sub-Saharan Africa (see Akwara *et al.*, 2003; Wellings *et al.*, 2006; Benefo *et al.*, 2008). Further to this, Kohler *et al.* (2007) observed that women in Thailand, Ghana and Kenya interact with each other on various behaviours such as family planning, sexual partnership and HIV/AIDS. Kohler *et al.* (2007) go on further to say that "The networks are highly gendered: men talk with men, women with women" (Kohler *et al.* 2007: p. 11). Conley and Topa (2002) in their review found social groups to be homogenous along socio-demographic attributes such as ethnicity, race, religion, age, gender and education levels. Race and ethnicity were particularly favourable attributes in these social contacts, in that interactions often occur along these lines. This is also evident in South Africa, hence making race a justifiable characteristic to include in the social group as a result of the apartheid regime that caused social division between racial groups (Emmett, 2003; Seekings *et al.*, 2005; Anderson *et al.*, 2007). Whereas age and income are more common characteristics, such that individuals of the same economic status or of the same age cohort are more likely to interact with each other.

It is worth noting that given the structure of our social groups (place of residence is not included in the construction of the social group) one would ask how the members of the social groups become aware of the existing *social norms* and how they would experience the *social pressure* in the group. An example of such a possibility is the influence of the popular internet social networks. This is exacerbated by the fact that access to internet social networks is possible not only through computers but via cell phones as well and where there is a high prevalence of cell phone usage in both advantaged and disadvantaged communities. See Skuse (2007), Kreutzer (2009) and Dlodlo (2009). As such social networks have made it possible for people to interact without geographical proximity. This has somewhat revolutionised social interaction (and made the model in this thesis plausible) with the idea that people can interact based on any attributes. For this literature refer to the works of Subrahmanyann *et al.* (2008); Wellman *et al.* (1996); Csanyi and Szendroi (2004); Hu and Wang (2011); Young and Rice (2011); Thelwall (2008) and Singla and Richardson (2008). Hence the aforementioned assertion that increases in residential mobility and technological advancement in communication have to some extent weakened the importance of physical geography in social interaction is supported by this literature.

It is given that age, gender and race are likely to be exogenous characteristics. The question then becomes is income also exogenous? Social interaction literature states that it is plausible for some endogenous variables to be exogenous characteristics in group formation. For example Blume *et al.*, (2010) note that: "In social interaction models, groups are typically defined in terms of exogenous categories such as ethnicity, gender or religion or endogenous categories such as residential neighbourhoods, schools and firms. The former, of course, may not literally be exogenous, but rather the determination of whether an individual is a member of the category is treated as predetermined from the perspective of the behaviours under study" (Blume *et al.*, 2010: p. 2). Similarly Brock and Durlauf (2004) state that: "The discussion assumes that the rules by which individuals are sorted into groups has no implication for empirical analysis, such an assumption implies that the group formation rule is independent of the determinants of individual choices" (Brock and Durlauf 2004: p. 3328).

Thus according to Brock and Durlauf (2004) and Blume *et al.*, (2010) the key is to consider the context under which the characteristics are assumed to be exogenous. That is, whether group formation (clusters based on socioeconomic characteristics) occurs and subsequently affect behaviour (norms related to *number of sex partners*) producing exogenous groups, or whether group formation occurs simultaneously with behaviour, where in this case there is a coevolution of social groups and behaviour producing endogenous groups. The key is then to look at the relationship between social norms related to *number of sex partners* (behaviour) and *income* (group characteristic). The problem then becomes that very little literature exists on social norms related to *number of sex partners* (Ragnarsson et al. 2010).

The literature that does exist mentions that social groups are based on socioeconomic characteristics (see Selikow *et al.* 2009; Ragnarsson *et al.*, 2009, 2010; Townsend *et al.* 2011; Gilbert and Selikow 2011) but does not state whether the groups are predetermined. As a result we turn to literature on *number of sex partners* and *income*, since the social norms that we are interested in are norms related to *number of sex partners*. In the current literature we find that *income* is likely to be a determinant of the *number of sex partners* an individual has (Ragnarsson et al. 2010). For example Ragnarsson et al. (2010) in their qualitative study record the following verbatim 'I can make an example about me, because I have eight girlfriends. It is my style and what I wear, my clothes and my money, because I have money' (Ragnarsson et al. 2010: p. 1255). This relationship is also likely to be observed in sexual relationships driven by financial gains. In such relationships we find that the more the *income* the more *number of sex partners* an individual is likely to have. See Hunter (2002), Kaufman and Stavrou (2004), Leclerc-Madlala (2008).

The problem however is that we also observe in the literature that the *number of sex partners* is likely to influence *income*. One such channel is sexual relationships driven by financial gains, where the more the *number of sex partners* the more *income* an individual is likely to spend to sustain these relationships. Additionally, in the literature we also find that *number of sex partners* has been associated with AIDS related illness and is therefore likely to affect an individuals' income, see Philipson and Posner (1995) and East et al. (2009). We do however recognise the on-going debate regarding the perceived lack of sufficient evidence to support the role of multiple partnerships in HIV transmission, see Mah and Halperin (2010a, 2010c); Lurie and Rosenthal (2010a, 2010b); Morris (2010) and Epstein (2010).

Given the aforementioned evidence from literature, we see that *income* is likely to influence *number of sex partners* and by the same token *number of sex partners* is likely to influence *income*. Thus a likely coevolution exists between this group characteristic (*income*) and behaviour (*number of sex partners*). This indicates the likelihood of endogenous group formation in so far as income is concerned. Thus unlike race, age and gender which are unlikely to produce this coevolution with behaviour and thus produce exogenous groups, *income* is likely to be endogenous as it is likely to co-evolve with behaviour. It is our view that the relationship between social norms related to *number of sex partners* and *income* would make a great area for future research. Especially empirical research that can establish whether these social groups, based on socioeconomic characteristics, are predetermined or not. Additionally, more literature is needed in the area of social interaction especially on how to best estimate such models, with practical application.

2.6.2.5 Covariates

These variables are standard in sexual behaviour research outlined in the literature and include educational attainment, knowledge of HIV prevention methods, and knowing someone with HIV/AIDS or someone who has died of AIDS.

2.6.3 Econometric Modelling

As indicated earlier the novelty of this study lays in its quantitative analysis of *social norms* and *social pressure* in influencing multiple partnerships. Section 2.1 introduced the qualitative literature findings that inspired the current study. This was followed by section 2.4 where the relationship between *social norms* and *social pressure* was provided. Section 2.5 detailed our chosen theoretical framework used to guide our expectations regarding the relationship between *social norms*, *social pressure* and number of sex partners. Using this theory together with the qualitative literature a hypothesis of a positive relationship between the *social norm* and number of sex partners and also between *social pressure* and number of sex partners was derived. This section presents the econometric modelling of this hypothesis. The section also highlights the identification challenges that our econometric model is likely to encounter and then proceeds to discuss our estimation strategy to overcome these challenges.

The hypothesis regarding the effects of *social norms* and *social pressure* on number of partners suggests the application of Manski's (1993) *linear-in-means* model as depicted by equation (2.8), where y_i is the normalised[36] number of sex partners belonging to individual *i* a member of social group g, $\mathbf{X}_i$ are the individual level characteristics, $\bar{y}_g$ is *social norms* in group g, $(1/\gamma)_g\bar{y}_g$ is the interaction of *social*

[36] Recall that we normalise the outcome variable (*multiple partnership*) by dividing this value with the age quartile so as to eliminate bias, see section 2.6.2.1. Additionally, recall that we also use an additional outcome variable which is a binary variable represented by one when a respondent has more than one sex partners and zero otherwise.

pressure and *social norm* variable in group g, given that equation (2.2) in the theoretical section suggests an interaction effect between these two variables. $\mathbf{X}_g$ is the social group characteristics included to avoid omitted variable bias (Manski 1993; Moffit 2000) and u_i is the error term. In this model the unit of analysis is an individual.

$$y_i = \beta_1 + \beta_2.\mathbf{X}_i + \beta_3.\bar{y}_g + \beta_4.(1/\gamma)_g\bar{y}_g + \beta_5.\mathbf{X}_g + u_i \qquad (2.8)$$

2.6.3.1 Econometric Model Identification

Although it is widely acknowledged that social interaction models[37] are perhaps the most difficult to identify, over the years there has been a growing amount of econometric literature covering estimation challenges and how best to overcome these challenges. The following discussion draws from the influential literature of Manski (1993, 2000, 2003); Moffit (2001); Brock and Durlauf (2004); Durlauf and Ioannides (2010) and Blume *et al.* (2010). In addition, we also make reference to empirical studies of Grodner and Kniesner (2005); Munshi and Myaux (2006); Etile (2007); Giorgi *et al.* (2010) and Grodner *et al.* (2011). Our econometric model is likely to face three identification challenges that need to be addressed.

The first being simultaneity and the related *reflection problem*, the second challenge relates to the influence of *unobservable group level* effects, and the last challenge is the effects of social group *selection bias* (Manski, 2000; Blume *et al.*, 2010). The *reflection problem* arises from the fact that the social interaction regressors, for example social norm and group characteristics, are collinear, and as a result of this linear dependence equation (2.8) becomes unidentifiable. For instance recall from equation (2.8) that y_i is the number of sex partners individual *i* in group g engages with, $\bar{y}_g$ is the social norm, $\mathbf{X}_i$ are individual characteristics and $\mathbf{X}_g$ are group characteristics. According to Manski (1993) $\bar{\mathbf{X}}_g = \mathbf{X}_g$, implying that the social group characteristics $(\mathbf{X}_g)$ is equal to the average of the individual characteristics in the group: $\bar{\mathbf{X}}_g = \frac{1}{N} \cdot \sum_{j=1} \mathbf{X}_i$. Also in equilibrium each individual's average behaviour will be equal to the expected behaviour (Blume *et al.* 2010):

$$E(\bar{y}_g) = \bar{y}_g = \frac{\beta_1 + \beta_2\bar{X}_g + \beta_4.(1/\gamma)_g\bar{y}_g + \beta_5 X_g}{1-\beta_3}, \beta_3 \qquad (2.9)$$

That is, equation (2.9)[38] states that individual *i* who is a member of group g's expectation of the average behaviour of the group $(E(\bar{y}_g))$ is equal to the average behaviour of the group $(\bar{y}_g)$, and this

[37] Empirical research on social interactions can be grouped into three hypotheses: *endogenous interactions/simultaneity* where an individuals' behaviour varies with the behaviour of the group, secondly, *contextual/exogenous interactions* where behaviour varies with the exogenous characteristics of the group, and thirdly *correlated effects/endogenous membership* where individuals in the same group behave the same as a result of being in the same environment. Whereas endogenous and contextual interaction explains social influence, the correlated interaction explains non-social interactions (Manski 1993, 2000; Moffit 2001).

[38] According to Blume *et al.*, (2010) the requirement that $\beta_3 < 1$ is guaranteed to hold, since β_2 maps the marginal rate of substitution between private and social conformity, which is a non-negative real number with interval $[0,1)$.

average behaviour is a linear function of $(1/\gamma)_g\bar{y}_g$, $\bar{\mathbf{X}}_g$ and $\mathbf{X}_g$. Since $\bar{\mathbf{X}}_g$ is the average individual characteristic in the group, it is possible to substitute $\bar{\mathbf{X}}_g$ with $\mathbf{X}_g$ (Manski's (1993) assumption), which implies that equation (2.9) can be re-written as (Blume *et al.* 2010):

$$\bar{y}_g = \frac{\beta_1 + \beta_2\bar{X}_g + \beta_4.\ (1/\gamma)_g\bar{y}_g + \beta_5 X_g}{1-\beta_3} = \frac{\beta_1 + \beta_4.\ (1/\gamma)_g\bar{y}_g + (\beta_2+\beta_5)X_g}{1-\beta_3} \quad (2.10)$$

Equation (2.10) shows that the regressor $\bar{y}_g$ is a linear function of the other social interaction regressors. This linear dependence makes identification impossible and holds for all other social interaction regressors. That is, the co-movement makes it impossible to disentangle the respective influence of social interaction regressors $\bar{y}_g$, $(1/\gamma)_g\bar{y}_g$ and $\mathbf{X}_g$ on our outcome variable y_i in equation (2.8). This is what Manski (1993) named the *reflection problem*[39] where changes in y_i could be as a result of $\bar{y}_g$ or it could be a result of $\bar{y}_g$ reflecting the changes in $\mathbf{X}_g$ (Blume *et al.* 2010). Brock and Durlauf (2004) and Blume *et al.* (2010) suggest the use of instruments to overcome the reflection problem.

The second identification challenge is the possible influence of *unobservable group level* effects which may result in similarity in behaviour. *Unobservable group level* effects are environmental or institutional conditions such as schools, healthcare or recreational activities. Recall from section 2.6.2 that the social groups were formed based on age, gender, race and income. Our social groups therefore do not contain any information on the place of residence which typically captures environmental and institutional conditions, and are usually the main source of unobservables (Blume *et al*., 2010). The advancement in modern telecommunication technology[40] such as cell phones, internet and social networks have drastically diminished the role of physical geography in social interaction, and have thus made such social group clusters plausible. See Manski (2000), Conley and Topa (2002), Urry (2003) and Xu *et al*., (2010). The social grouping has therefore had the effect of neutralising geographically related unobservables by assigning individuals into *new communities*[41]. Table 2.B1 in Appendix 2.B shows the distribution of the original communities and the *new communities* that have been formed as a result of the grouping.

[39] One of the proposed solutions to this *reflection problem* has been to estimate a 'composite coefficient' that includes both endogenous and exogenous effects. For example instead of estimating equation (2.8) we instead estimate the following equation; $y_i = \beta_1 + \beta_2.\text{social interaction}_g + \beta_5.\mathbf{X}_i + u_i$ where β_2 measures social interaction in general. Hence β_2 allows one to determine whether *social interaction* is present in the model (Manski 1993; Moffit 2000), one can then proceed to measure the impact of social interaction without being able to distinguish the source of this interaction.

[40] It is worth noting that some of the internet channels such as online chat-rooms individuals are unlikely to disclose their real information, hence grouping based on such information is likely to be tainted by such false information.

[41] The 'new' communities are defined by individuals that make up the social group, column 1 of table 2.A in Appendix 2. For example social group 54 comprises 17 individuals who were originally from the following communities; 3 from Cape Town; 3 from Goodwood; 1 from Beacon Valley; 1 from Eastridge; 2 from Tafelsig; 1 in Westridge; 1 in Woodlands; 1 in Simonstown; 2 from Wynberg; 1 from Malmesbury and 1 from Wynberg Non-Urban area have now been grouped into the social group 54 and 'new' community 54.

In this study we have ruled out the influence of geographically related unobservables. We can however not rule out the likelihood of our social groups being influenced by non-geographically related unobservables. This is based on the fact that the groups consist of individuals of the same age, gender, race and income. Such similarity implies that each social group consist of individuals who are alike, and as such these similar individuals are likely to face similar unobservable backgrounds (such as similar family structure or composition) thus causing these individuals to behave the same. Hence we need to remove this influence to ensure that it is the social norm that is causing the similarity in behaviour and not the fact that the individuals in the social group are behaving the same simply because they have similar unobservable family composition.

For instance the similar race characteristic in the social groups implies that the groups are fragmented along racial lines. As such the Black race social group for example are likely to be influenced by such unobservables as family practice of polygamy whereas monogamous family practices may influence the White race social groups. Another example is the influence of magazines and television programmes which usually differ along age lines, gender lines, income and even racial lines. Hence it is plausible that individuals may behave the same as a result of these non-geographically related unobservables. We eliminate this unobservable influence as well, as we want to ensure that the only influence emanates from *social norms* and the related *social pressure*. See Brock and Durlauf (2004) and Blume *et al.* (2010).

The third likely challenge is with regards to social group *selection bias* which arises when individuals are (un) intentionally sorted into social groups. In our case the selection bias is due to the non-random assignment of individuals into social groups where we clustered these individuals into similar socioeconomic characteristics. Hence it is difficult to view this clustering as untainted by selection bias. See Angrist and Pischke (2008) on selection bias and randomisation.

2.6.3.2 Final Models Used For Estimation

We test different model specifications. The first set of models is cross-sectional in nature using CAPS wave 3 datasets. We apply both the *linear-in-means* model (where the outcome variable is number of sex partners) and a *non-linear-in-means* (binary variable where one represents more than one partner and zero otherwise) model. This part of the analysis is based on equation (2.8). The second set of models utilises the panel nature of CAPS data where we employ wave 1, 3 and 4. In this model we employ equation (2.11) which is based on individual fixed effects where y_i^t is the normalised sex partners belonging to individual *i* a member of group g at time t, $\overline{y}_g^t$ is the relevant social norm in group g in time t, $(1/\gamma)_g^t \overline{y}_g^t$ is social pressure interacted with social norm in group g, $\mathbf{X}_g^t$ are the group characteristics, $\mathbf{X}_i^t$ are the individual characteristics, ∂_g is the social group dummy capturing the effect of time-invariant unobservable, ∂_t is the year dummy that controls for variation across the years and u_i^t the error term.

$$y_i^t = \beta_1 + \beta_2.\bar{y}_g^t + \beta_3.(1/\gamma)_g^t\,\bar{y}_g^t + \beta_4.\mathbf{X}_g^t + \beta_5.\mathbf{X}_i^t + \partial_g + \partial_t + u_i^t \quad (2.11)$$

2.6.3.3 Instrumental Variable Strategy

It is widely recognised that finding valid instruments in models of social interaction is a difficult task due to the fact that social interaction models are generally *open-endness*[42] in nature (Manski, 2000; Blume *et al*., 2010). The section outlines the instrumental variable (IV) strategy. An instrumental variable must be correlated with the endogenous regressor but be uncorrelated with the error term, that is, the IV needs to be redundant in the main equation. The narrative below provides an argument to support the proposed IV. In doing so we follow Wooldridge (2002); Cameron and Trivedi (2005) and Murray (2006), noting in particular Murray (2007) who states that the amount of credence granted to IV depends partly on the quality of the arguments.

The *trendsetter of the norms IV* is the instrumental strategy that we adopt. The proposed *trendsetter of the norms IV* extends the work of Grodner and Kniesner (2005) and Etile (2007)[43]. The key to the IV strategy is social grouping. Recall that the social groups consist of individuals with similar socioeconomic attributes (gender, race, age and income). The *trendsetter of the norms IV* is founded on the premise that the older social groups are more likely to be *trendsetters* for younger groups in *similar* socioeconomic neighbourhoods (that is, same income, race and gender). As such older social group behaviour[44] can be used as IVs for younger social groups whereas the opposite is unlikely to hold. In simple terms $\mathrm{cov}(\bar{y}_{\text{older group}}, \bar{y}_{\text{younger group}}) \neq 0, \mathrm{cov}(\bar{y}_{\text{older group}}, u_{\text{young ind}}) = 0,$ whereas, $\mathrm{cov}(\bar{y}_{\text{younger group}}, u_{\text{old ind}}) \neq 0.$

We believe that the *trendsetter of the norms IV* is correctly excluded from the main equation because norms and the fear of sanctions predict that the only thing that matters for an individual is the behaviour of the social group they belong to. See Fisher (1988); Bernheim (1994); Akerlof (1997); Cialdini and Trost (1998); Munshi and Myaux (2006); Fletschner and Carter (2008) for a discussion on norms and fear of sanctions. For example Fisher (1988) states that: 'People often conform to the attitudinal and behavioural norms of their reference group or social network-such norms constitute a potent source of social influence. That group norms affect individuals' behavioural choices regarding prevention has been documented. One reason why people adhere to group norms and

[42] Blume *et al.* (2010), define '*open-endness'* as models lacking theoretical structure that can naturally identify variables that can be included (and those that can be excluded) from the empirical model.

[43] Grodner and Kniesner (2005) and Etile (2007) IV is based on instrumentation of a variable by its lags in panel data analysis. Specifically, given an individual *i* of age A_i, then all same socio-economic individuals *j* with age A_j such that $A_i - 5 \leq A_j < A_i$ belong to social group R_i. Then same socio-economic individuals *k* such that $A_j - 5 \leq A_k < A_i - 5$ belong to R_j but not R_i. Then the social norm of individuals k should be correlated with social norms of individuals *j*. Symmetrically, $A_j + 5 \leq A_k < A_i + 5$ for all of *j's* in R_i. Therefore, adjacent group behaviour to the left and to the right of the social group can be used as IVs allowing for overidentification of the model.

[44] Social group behaviour in this case refers to both the *social norms* ($\bar{y}_g$) and the related *social pressure,* $(1/\gamma)_g\bar{y}_g$, which has an interaction with *social norm*.

espouse group values is because they fear sanctions for nonconformity (Fisher 1988: p. 914). In section 2.4 we provide literature on social norms. Hence the reason why the $\mathrm{cov}(\bar{y}_{\mathrm{older\ group}}, u_{\mathrm{young\ ind}}) = 0$. That is, the social norm of another group (older group social norm) will not affect the individual simply because an individual is unlikely to face sanctions if they deviate from the norm of the older group. However because of the fact that the older social groups in the same socioeconomic background are likely to be the *trendsetter* of the norms implies that $\mathrm{cov}(\bar{y}_{\mathrm{older\ group}}, \bar{y}_{\mathrm{younger\ group}}) \neq 0$. Admittedly the idea that older social groups are more likely to be *trendsetters* for younger groups in *similar* socioeconomic neighbourhood is not new but rather a reorientation of literature.

Indeed, support of the *trendsetter* concept comes from a series of literature. The key insights in this literature is that social behaviours are usually initiated by and manifested in a particular sub-culture or group before spreading to other groups in the social system or culture. The early adopters or initiators of the behaviour are the so called *trendsetters* of the behaviour (Pinkerton *et al.* 1995; Bertrand, 2010; Salvini and Vignoli 2011). Hence, the argument here is that behaviour in a group does not just appear but rather it emerges in some sub-groups and diffuses to other sub-groups. The importance of trendsetter is accentuated by Bertrand (2010), for instance, who contends that a new behaviour diffuses into a social system through *trendsetters* who usually model a new behaviour in the social systems. For example Kelly (1993) state that 'In experimental community-level tests of interventions intended to change HIV risk behaviours have relied on a diffusion of innovation model that posits that population behaviour and norm perceptions can be influenced by innovative trends initially exhibited and communicated by trendsetters in the population: overtime, these new trends diffuse and become normative throughout the population' (Kelly 1993: p. 2017). This is further supported by Mahdavi (2009) in a study consisting of young adults who were identified as *trendsetters* for youth in Tenran, Iran in relation to sexual, cultural and social revolution occurring in Iran. The same concept of trendsetting has been applied to influence changes in sexual behaviours related to HIV/AIDS, see Kelly *et al.* (1992), Kelly *et al.* (1993), Pinkerton *et al.* (1995), Geary *et al.* (2006), and Bertrand (2010).

The *trendsetter* notion is also found in literature on smoking, where the supposition is that young adults who smoke usually influence the smoking behaviours of youth. Based on this supposition Hendlin *et al.* (2009) observed that the *trendsetter* concept is employed by tobacco companies by using young adults to recruit and influence youth to adapt smoking behaviours. See also Alexander *et al.*, (2001) for a similar argument. For instance Hendlin *et al.* (2009) note that 'Since 1998 Master Settlement Agreement (MSA) severely restricted under-18 directed tobacco advertisement, the major tobacco companies have increasingly targeted young adults who represent an important market for tobacco companies and also set trends for adolescents, the small segment of the population who serve as innovators and early adopters of new trends influence consumer trends for the rest of society' (Hendlin *et al.* 2009: p. 213). Another example rests in the fashion industry where underweight models and celebrities, have been described as *trendsetters* of clothes and body image for teenagers (Dragone and Savorelli 2012). Along similar lines *trendsetters* have been noted to be responsible for marital disruption and changes in fertility patterns, see Salvini and Vignoli (2011) and D'Addato *et al.*, (2008) respectively. Meanwhile, another study by Neubourg and Vendrik (1994) further emphasises the concept of *trendsetters* related to norms affecting labour force participation[45].

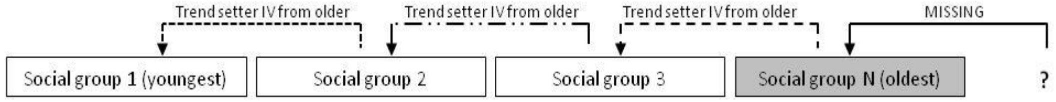
The limitation of the *trendsetter of the norms IV* is that it excludes the oldest group[46] as it will not have an older group that can be used as IV[47]. Furthermore the IVs make these models exactly identified. Importantly, our IVs depend on social groups, this implies that if the assumptions that we placed on the social grouping (for example, proper definition of the social groups, the right demarcation of age between groups) do not hold, this will in essence affect our IVs strategy such that we are likely to have redundant instruments. In Appendix 2.E we adopted the *adjacent group IV* from the works of Grodner and Kniesner (2005) and Etile (2007). Our last IV, the *lagged values* is applied in the panel data analysis and consists of *one year lag* and *two year lag* as instruments in Appendix 2.F. The *lagged values* IV extends the fear of sanctions argument and assumes that although past behaviour in the group is likely to be correlated with present behaviour, fear of sanctions dictates that it is only the current behaviour of the group that will influence behaviour of an individual. Hence past behaviour is correctly excluded from the main equation.

2.6.3.4 Including the Population that is not Sexually Active

The analysis thus far has concentrated on the *sexually active population*. This section extends the analysis to include the *sexually inactive population*. This study defines *sexually inactive population* as those individuals who have ***never*** had sex. This information is derived from the following questions: "Have you ever had sex? By sex we mean full penetration. The responses included, yes or no." In contrast the *sexually active population* refers to those who have had sex before. The motivation behind including this population is based on the fact that *abstinence* as an approach to *protective sexual behaviour*

[45] It is worth pointing out that some behaviours are likely to be set by younger groups and adopted by older groups, such that the younger groups become the trendsetters for the older groups, for example social media. Also some trends are influenced by other factors, apart from age, for example the banking sector has been described as a trendsetter in adopting IT, see Anandarajan et al. (2000). Another example is United States of America being seen as a trendsetter for other countries, see Rind (2010).

[46] We have assumed a 5 year age demarcation between social groups. Such that group g+5 is the older group of g.

[47]


We include the '*extended' trend setter of the norm IV* based on *trend setter of the norms IV*. The difference between the two is that the omitted oldest social group is included by instrumenting this group with another oldest social group with different characteristics such as income. This is to allow the inclusion of the omitted social group since this IV eliminates the last social group which contains the oldest age group.

is possibly more effective than, *condom use* or *one sex partner* approaches. That is the *sexually inactive population* are likely to have no risk of HIV/AIDS contraction (we do not include the risk of HIV/AIDS from needle sharing or blood transfusion). However the minute an individual becomes sexually active, that is in the *sexually active population* HIV/AIDS risk is introduced, and this risk varies depending on an individuals' choice of sexual behaviour. For example condom use versus not using a condom, or sex with a spouse versus a commercial worker or even having had sex once in your life time in comparison to having had sex more than 20 times in your life. All these examples introduce different variations of the risk of contracting HIV/AIDS. Thus our classification of *sexually inactive population* containing individuals with no risk and *sexually active population* consisting of individuals with various HIV/AIDS risk levels.

The motivation behind including this population is based on the fact that *abstinence* as an approach to *protective sexual behaviour* is similar to, and possibly more effective than, *condom use* or *one sex partner* approaches. Table 2.C1 in Appendix 2.C compares the characteristics of the *sexually active population* to that of the *sexually inactive population*. Since the mean age of sexual debut in the *sexually active population* is lower than the mean age of the *sexually inactive population*, is more likely to imply a choice of *abstinence* hence supporting our argument. Therefore the exclusion of this population in sexual behaviour research is likely to eliminate important analytical information related to individuals who pursue *protective sexual behaviour*.

The second motivation for including the *sexually inactive population* relates to the theme of *social norms* and *social pressure* in this study. Following from this theme the argument is that it is likely that the *sexually inactive population* face *social norms* and *social pressure* comparable to the *sexually active population* due to social interaction. Hence by eliminating this population from the analysis we would in essence be creating a *utopian world* where we assume the *sexually active population* only interact with others who are sexually active, while the *sexually inactive population* interact among themselves. The third motivation has its origin in econometric modelling and the implication thereof for non-random sample selection. Since being sexually active is likely to be correlated with unobservables that affect the number of sex partners, constraining the sample to only the *sexually active population* is likely to produce biased estimators.

However, the inclusion of the *sexually inactive population* in the analysis introduces two estimation challenges. The first challenge relates to missing information (and therefore missing variables) related to sexual behaviours among the *sexually inactive population,* for example condom use, relationship type, relevant to this study is information on *multiple partnerships*. The second challenge relates to taking into consideration the fact that there are underlying factors that cause individuals to become sexually active. That is, the choice to become sexually active or remain abstinent is not qualitatively the same but rather a function of various individual level factors (Bezabih *et al*., 2010).

To overcome the above mentioned challenges the analysis follows Wooldridge (2002); Greene (2004); Okten and Osili (2004); Bushway *et al.* (2007); Garcia *et al.* (2008); Bezabih *et al.* (2010) and Bollinger and Hirsch (2012) and adopt the censored two stage model (CTSM) as this model includes all observations from the two populations and missing outcome information are given zero values. In the first stage a PROBIT model for sample inclusion involving whether you are sexually active or not is conducted and the inverse Mills ratios are then included in the second stage IVTOBIT model as an additional regressor.

The analysis in this section follows from the *linear-in-means* model in equation (2.8), the only difference is that y_i the number of sex partners, is observed if individual *i* is sexually active as depicted by equation (2.12). Equation (2.13) depicts our selection equation outlining whether individual *i* is sexually active or not. Finally $\mathbf{Z}$ is a vector of determinants of the decision to become sexually active and include individual characteristics namely age, income, gender, education and knowledge of HIV/AIDS. The exclusion restriction is a dummy variable capturing whether abstinence is acknowledged as an HIV prevention method.

$$y_i | \, y_i^s = 1 \quad = \mathbf{X\beta} + u \qquad (2.12)$$

$$y_i^s = \begin{cases} 1 \text{ if } & y_i^{s*} = \mathbf{Z\gamma} + \nu \geq 0 \\ 0 & \text{otherwise} \end{cases} \qquad (2.13)$$

where $\mathbf{X\beta} = \beta_1 + \beta_2 . \mathbf{X}_i + \beta_3 . \bar{y}_g + \beta_4 . \frac{1}{\gamma} \bar{y}_g + \beta_5 . \mathbf{X}_g$. In summary our estimation strategy accounts for the fact that number of sex partners will depend on whether you are sexually active or not.

2.7 Estimation Results

2.7.1 Regression Results

Table 2.1 shows the results from the *linear-in-means* model of equation (2.8). The analysis is based on CAPS wave 3 dataset. The models in Panel A represent the entire sample size, that is, they include both the *sexually active population* and those who have not had sex, that is, the *sexually inactive population*. Panel A's results are based on the second stage of the censored two stage model (CTSM) which involves using the inverse Mills ratio from the first stage model as an additional regressor in the second stage as depicted in Panel A. This is so as to eliminate the sample selection bias brought about by including the *sexually inactive population*. The first stage Heckman model is shown in Table 2.D1 in Appendix 2.D.

The *non-linear-in-means model* in columns 1 to 2 uses a censored dependent variable that captures the *number of sex partners in the last 12 months* and is represented by 0 for those in the *sexually inactive population* and greater than 0 for those who are in the *sexually active population*. As such the IVTOBIT model that captures the censored nature of the dependent variable is utilised. The *linear-in-means-model* in columns 3 to 4 of panel A continues to use the censored dependent variable used in the *non-linear-in-means* model. The only difference is that the GMM is used following Angrist and Pischhke (2008) who have found that similar results are obtained with OLS or 2SLS estimation as with PROBIT or TOBIT models when the dependent variable is binary or censored. As such we use GMM estimation especially given that the linear models in STATA are less constrained and allow the use of the Sargan-Hansen statistics test for identification which is only available in linear models. Furthermore, linear models allow for easier inference than the nonlinear models.

Panel B includes only the *sexually active population*, that is, those individuals who indicated that they have had sex. *The non-linear in means model* in columns 5 and 6 uses a binary outcome variable representing 1 if an individual has multiple partners and 0 otherwise derived from the variable *number of sex partners in the last 12 months*. The *linear-in-means-model* in column 7 to 8 uses the actual *number of sex partners in the last 12 months* as the outcome variable among the population that is sexually active. Lastly, using both the *linear-in-means-model* and *non-linear-in-means model* allows us to determine how the model would behave given that the *non-linear-in-means model* has different requirements for identification from the *linear-in-means-model*. Brock and Durlauf (2004); Durlauf and Ioannides (2010) and Blume *et al*. (2010) indicate that nonlinearity improves identification.

The models presented in this paper are based on robust standard errors with clustering at the social group level. For a discussion on robust standard errors and clustered data see Moulton (1990); Williams (2000) and Wooldridge (2003). The discussion is on biasedness of standard errors in estimation models with aggregated data and micro observations. In our model this relates to the social norms and social pressure regressors which are aggregated at the social group level and where our unit of analysis is an individual (recall from section 2.6 that the aggregation of the social groups is based on individuals with similar socioeconomic characteristics). To take into account the aggregated nature of the data we used robust standard errors adjusted for clustering at social group level. It is worth noting that estimation models fitted with robust standard errors unadjusted for clustering produced consistent results. The results presented in this paper are based on robust standard errors adjusted for clustering at the social group level.

We tested for model identification using the Anderson canon correlation test under the null hypothesis that the model is underidentified, all our models reject the null hypothesis, as shown in the last rows of all the tables. Our estimation results using the *adjacent social group age bracket IV* are displayed in Appendix 2.E. This instrument allows for overidentifying restrictions, as such we used the Sargan-Hansen test in the GMM models and Amemiya-Lee Newey in the IVTOBIT/IVPROBIT models on the joint null hypothesis that the instruments are valid, that is, the instruments are uncorrelated with the error term and are correctly excluded from our equation.

The results are displayed in the last rows of the tables. Overall the models are overidentified. Appendix 2.F shows the results by gender. Specifically table 2.F1 is based on the male population while table 2.F2 is based on the female population. Similar results for both genders are observed, indicating that *social norms* and *social pressure* are significant across gender.

The final analysis is based on panel data using CAPS waves 1 (2002), 3 (2005) and 4 (2006). These results are depicted in Table 2.2. The first column of this table is based on the fixed effects model of equation (2.11) using wave 1, 3 and 4. Column 2 is based on wave 3 where we use wave 1 as a lagged instrument for the endogenous social interaction regressors in wave 3. In column 3 we use the variables of wave 3 as instruments for endogenous variables of wave 4. Column 4 uses a *two year lag* as instruments. As such the analysis is based on wave 4 where wave 1 is used as the *two year lag* instrument. As shown in the table the panel data analysis also support our hypothesis of a positive relationship between *social norms*, *social pressure* and sex partners.

In general the results vary by type of instrument, type of population and type of model specification. The results also show that the *social norm* which captures the average number of sex partners in the social groups has a positive and significant influence on individual own choice of number of sex partners. This effect is enhanced as *social pressure* mounts for any given *social norm* prevalent with the social group. This is shown by the positive coefficient on the interaction term between *social norm* and *social pressure*. The dispersion and the mean behaviour, therefore play pivotal roles in influencing an individual's tendency to conform to the prevailing *social norm* within a group. In any of our models either the *social norm* itself or the interaction between the *social norm* and the *social pressure* prevalent in the social group is positive and significant.

It is worth noting that given the structure of our social groups (place of residence is not included in the construction of the social group) one would ask how the members of the social groups become aware of the existing norms and how they would experience the *social pressure* in the group. An example of such a possibility is the influence of the popular internet social networks such as Facebook and Twitter, where individuals are exposed to the same norms without necessarily being in the same place of residence. Social networks have made it possible for people to interact without geographical proximity. This has somewhat revolutionised social interaction (and made the model in this thesis plausible) with the idea that people can interact based on any attributes. In this paper we argue interaction is driven by people with similar socio-economic background, motivated by the qualitative literature related to the influences of multiple partnerships.

Table 2.1: Social Norm and Social Pressure Results

	Panel A Both *Sexually Active Population* and *Sexually Inactive Population*				Panel B *Sexually Active Population* Only			
	IVTOBIT CTSM 2^{nd} stage Nonlinear in means model		GMM CTSM 2^{nd} stage Linear in means model		IVPROBIT Nonlinear model		GMM Linear in means model	
	No. of sex partners incl. not sexually active		No. of sex partners incl. not sexually active		Multiple Partnership (0/1)		No. of sex partners	
	Trend setter IV	'Extended' Trend setter IV	Trend setter IV	'Extended' Trend setter IV	Trend setter IV	'Extended' Trend setter IV	Trend setter IV	'Extended' Trend setter IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Social norm	**1.191*****	**1.076*****	**0.0218**	**0.0991**	**1.372*****	**1.486*****	**0.0105*****	**0.0137*****
	(0.321)	**(0.243)**	**(0.216)**	**(0.174)**	**(0.246)**	**(0.225)**	**(0.00365)**	**(0.00301)**
Social norm*social pressure	**0.451**	**0.822***	**3.027*****	**3.029*****	**12.18****	**11.57****	**0.424*****	**0.414*****
	(0.580)	**(0.432)**	**(0.397)**	**(0.312)**	**(4.937)**	**(5.126)**	**(0.0952)**	**(0.0878)**
Social group characteristics	-0.00323	-0.0165	0.0616	0.0316	0.0534	0.138	0.00122	0.00134
	(0.0776)	(0.0680)	(0.0542)	(0.0505)	(0.134)	(0.151)	(0.00231)	(0.00224)
Education years	-0.00829	-0.00884	-0.00570	-0.00684	-0.00272	-0.0103	-5.14e-05	-0.000233
	(0.0120)	(0.00933)	(0.00836)	(0.00689)	(0.0332)	(0.0325)	(0.000540)	(0.000452)
Mills ratio	75.50***	61.86***	38.44***	34.44***				
	(11.86)	(9.289)	(7.707)	(6.501)				
HIV prevention, abstain	0.291***	0.242***	0.170***	0.149***	-0.0178	-0.0162	-0.000496	0.000469
	(0.0478)	(0.0377)	(0.0331)	(0.0277)	(0.120)	(0.117)	(0.00213)	(0.00178)
HIV prevention, one partner	-0.139*	-0.129**	-0.0336	-0.0399	-0.0551	-0.00433	0.00180	0.00147
	(0.0749)	(0.0594)	(0.0514)	(0.0434)	(0.195)	(0.192)	(0.00339)	(0.00285)
Knows someone died of AIDS	0.146***	0.112***	0.108***	0.0864***	0.00197	0.0320	0.00138	0.000961
	(0.0515)	(0.0403)	(0.0363)	(0.0302)	(0.128)	(0.125)	(0.00218)	(0.00183)
Majority population, Coloured	-0.0267	-0.0263	-0.316***	-0.252***	-0.0145	-0.0297	-0.00537	-0.00502
	(0.129)	(0.0920)	(0.0872)	(0.0670)	(0.222)	(0.226)	(0.00379)	(0.00347)
Majority population, White	0.0424	0.0692	-0.433**	-0.305*	-0.125	-0.328	-0.00191	-0.00245
	(0.273)	(0.237)	(0.188)	(0.175)	(0.440)	(0.481)	(0.00754)	(0.00721)
HIV prevention, condom					-0.169	-0.187	-0.00463	-0.00475*
					(0.209)	(0.206)	(0.00330)	(0.00271)
Constant	-11.55***	-9.456***	-5.205***	-4.699***	-2.913***	-2.948***	0.0414***	0.0401***
	(1.736)	(1.358)	(1.127)	(0.948)	(0.538)	(0.529)	(0.00855)	(0.00739)
Observations	2,302	2,769	2,302	2,769	1,524	1,962	1,524	1,962
R-squared			0.161	0.181			0.106	0.121
Anderson canon correlation statistics (p-value)			0.0000	0.0000			0.0000	0.0000

- Robust Standard errors in parentheses

- *** p<0.01, ** p<0.05, * p<0.1 indicates significance at 1%, 5% and 10%
- Majority population reference is Black
- Instrumented for social norm and social norm*social pressure
- Not that the excluded instruments: social norm of older social group, social norm*social pressure of older social group

Table 2.2: Panel Data Analysis

	(1) FE Model[1]	(2) 2SLS[2]	(3) 2SLS[3]	(4) 2SLS[4]
	Based on wave 4, 3 and 1	Based on wave 3 and 1	Based on wave 4 and 3	Based on wave 4and 1
Dependent variable No. of sex partners	IVs 1 year lag (1)	IVs 1 year lag (2)	IVs 1 year lag (3)	IVs 2 year lag (4)
Social norm	**1.131**	**0.896*****	**0.907*****	**1.060*****
	(0.926)	**(0.163)**	**(0.159)**	**(0.263)**
Social norm*social pressure	**0.539****	**0.170**	**0.0818**	**0.347***
	(0.268)	**(0.245)**	**(0.0849)**	**(0.194)**
Social group characteristics	0.212	-0.00975	-0.0164	0.151**
	(0.586)	(0.0295)	(0.0572)	(0.0730)
Education years	-0.107	-0.000308	-0.0375***	-0.0286**
	(0.105)	(0.00586)	(0.0114)	(0.0119)
Knows someone died of AIDS	-0.0957*	0.0474*	0.0526	0.0679
	(0.0526)	(0.0260)	(0.0517)	(0.0547)
Majority population, Coloured		0.00414	-0.0104	0.204**
		(0.0354)	(0.0723)	(0.0879)
Majority population, White		0.0332	0.0249	0.361**
		(0.0676)	(0.150)	(0.181)
year	-0.0816			
	(0.0821)			
Constant	0.860	0.0311	0.432**	-0.000791
	(1.043)	(0.0950)	(0.184)	(0.215)
Observations	3,071	1,289	1,782	1,280
R-squared		0.375	0.277	0.274
Number of person id	2,094			
Anderson canon correlation statistics (p-value)	0.0000	0.0000	0.0000	0.0000

- Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
- Instrumented for social norm and social norm*social pressure
- Excluded instruments: lag social norm and lag social norm*social pressure

[1] The Fixed Effects (FE) model uses CAPS waves 1, 2 and 3, where the endogenous variables are instrumented for by one year lags. Wave 4 lacks information on HIV prevention; hence we have not included this wave in the panel data analysis. Further the panel data models are exactly identified that is, they have the exact number of instrumental variables for our endogenous variables, hence we cannot derive the Cragg-Donald Wald F statistic and the Sargan Hansen tests. However, all the models pass the Anderson canon LM statistic test, implying that the model is identified.

[2] This model is based on wave 3, and in this model we instrument for the endogenous variables using a one year lag (that is, the values in wave 1).

[3] The wave 4 dataset was used in this model, where the endogenous variables were represented by a one year lag as instrumental variables (wave 3 values).

[4] The analysis is based on wave 4, however in this model instead of a one year lag as before the endogenous variables are instead instrumented for using a two year lag (that is, the wave 1 values).

2.7.2 Social Group Falsification Test

The current empirical literature covers various factors that have been used to define groups[48]. For example, ethnicity, socioeconomic position and close proximity in the study by Conley and Topa (2002). In another study ethnicity and religiosity were used to construct a group, see Manski and Mayshar (2003). Grodner and Kniesner (2005) defined groups using demographic similarity and close proximity, while Etile (2007)'s groups were based on gender, age and occupation. Classmates were used in a study by Fletcher (2009) and students taking the same courses form the basis of groups in Giorgi *et al.* (2010). Census tract is another example, see Reagan and Salsberry (2005). Finally communities have also been used to define groups (Bezabih *et al.* 2010). A major criticism of social interaction models is the existence of measurement error caused by lack of distinction *a priori* concerning appropriate definitions of social groups (Manski 1993; Blume *et al.* 2010). Another criticism is that most of the empirical studies do not offer evidence that supports their definitions of the social groupings.

The purpose of this section is to validate our social group composition. To achieve this, the motivation of our social group composition is first outlined, thereafter the section proceeds to perform a social group falsification test by constructing a *placebo social group* using ad hoc characteristics to determine whether evidence of social interaction will be found when ad hoc information for our social group composition. Lastly a comparison of the social group and the *placebo social group* using various sexual behaviour variables is offered.

As previously mentioned a major criticism of empirical models is their lack of clear, *a priori* mechanisms of social interaction. Unlike most studies, the advantage of our study is the existence of sufficient background information to create a valid social group composition. This background information is based on qualitative research, which suggests that social groups consist of individuals with the same socioeconomic status. Furthermore the characteristics of our social group are supported by the literature, for example Akwara *et al.* (2003); Wellings *et al.* (2006) and Benefo (2008).

The next step is to test whether there is evidence of social interaction if different social group composition is used. This social group falsification test is conducted using '*placebo treatments*' following Lavy and Schlosser (2007). The test is based on ensuring that the social interaction evidence found from the proposed social grouping is not a result of measurement errors, such as omitted unobservable confounders. As such the expectation is that insignificant social interaction will be found when ad hoc variables are used to construct social groups (Lavy and Schlosser 2007).

[48] Researchers studying social effects rarely offer empirical evidence to support their reference-group specifications. If researchers do not know how individuals form reference groups and perceive reference-group outcomes, then it is reasonable to ask whether observed behaviour can be used to infer these unknowns (Manski 1993: p. 536).

A similar approach was adopted by Fletcher (2009) and Giorgi *et al.* (2010). Giorgi *et al.* (2010) study sought to determine how student's choice of courses affected individuals' main course selection (where the groups consisted of students in the same class taking the same courses). Thus for the falsification test Giorgi *et al.* (2010) replaced the groups with placebo groups that were composed by randomly and artificially allocating students to hypothetical classes. The placebo social grouping findings showed no evidence of peer effects as per the authors' expectations. Fletcher's (2009) study analysed how classmates' smoking behaviour influences youth smoking decisions. For the falsification test Fletcher (2009) replaced classmates (peers who smoke) with those in lower or higher grades from the same school. The author found no evidence of peer effects when lower or higher grades classes are used.

Recall that our social groups are based on the *similar* others (age, race, income and gender) concept on the assumption that individuals who have these similar attributes are likely to interact in relation multiple partnerships. In this vein, the falsification test uses *placebo social groups* which assign individuals into groups using ad hoc *similar* others variables. Basically we are arguing that not any *similar* others attributes will lead to social interaction that can produce *social norms* and *social pressure* related to multiple partnerships. These placebo *similar* others variables include: whether the youth shops for the household; chances of living in Cape Town three years from now; and young adults' month of birth. Our expectation is that there should be no evidence of social interaction based on this *similar* others *placebo social groups*. Table 2.2 shows the results based on the *placebo social groups*[49]. As is evident from the table we find no significant effects of social interaction (*social norms* and the interaction between *social norms* and *social pressure*) variables on number of partners. This confirms that our findings in table 2.1 are not spurious but are indeed picking up the effects of social interactions.

[49] In Appendix 2.H we also use the *adjacent social group age bracket IV* in addition to the *trendsetter IV on placebo social groups.*

Table 2.2: Falsification Test - Placebo Social Groups Results

	Panel A Both *Sexually Active Population* and *Sexually Inactive Population*				Panel B *Sexually Active Population* Only			
	IVTOBIT CTSM 2nd stage Nonlinear in means model No. of sex partners incl. not sexually active		GMM CTSM 2nd stage Linear in means model No. of sex partners incl. not sexually active		IVPROBIT Nonlinear model Multiple Partnership (0/1)		GMM Linear in means model No. of sex partners	
	Trend setter IV	'Naïve' Trend setter IV	Trend setter IV	'Naïve' Trend setter IV	Trend setter IV	'Naïve' Trend setter IV	Trend setter IV	'Naïve' Trend setter IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Social norm	**1.407**	**0.861**	**2.665**	**1.793**	**-0.0902**	**-0.0475**	**8.942**	**-4.003**
	(1.809)	**(0.892)**	**(2.588)**	**(1.238)**	**(0.375)**	**(0.343)**	**(16.93)**	**(17.51)**
Social norm*social pressure	**-2.381**	**-0.0596**	**-4.466**	**-0.830**	**0.0898**	**0.0476**	**-3.808**	**7.129**
	(4.674)	**(0.810)**	**(6.760)**	**(1.115)**	**(0.279)**	**(0.245)**	**(12.58)**	**(12.41)**
Social group characteristics	0.238	0.139	0.371	0.224	0.00420	0.0108	-0.153	-0.0835
	(0.274)	(0.144)	(0.394)	(0.198)	(0.00888)	(0.0261)	(0.419)	(1.354)
Education years	-0.0297**	-0.0332***	-0.0276	-0.0324***	-0.000798	-0.00170	-0.0492	-0.0812
	(0.0126)	(0.00877)	(0.0179)	(0.0119)	(0.000761)	(0.00140)	(0.0376)	(0.0730)
Mills ratio	32.91***	32.79***	63.70***	63.02***				
	(10.98)	(7.062)	(16.14)	(10.13)				
HIV prevention, abstain	0.152***	0.120***	0.285***	0.248***	-0.000975	0.00111	0.0844	-0.0536
	(0.0438)	(0.0340)	(0.0617)	(0.0464)	(0.00283)	(0.00297)	(0.141)	(0.155)
HIV prevention, one partner	-0.0194	-0.0419	-0.118	-0.116*	0.00443	0.00112	0.0892	-0.0682
	(0.0683)	(0.0466)	(0.0968)	(0.0639)	(0.00444)	(0.00384)	(0.236)	(0.202)
Knows someone died of AIDS	0.124***	0.0748**	0.178***	0.121***	0.00534*	0.000781	0.135	-0.0245
	(0.0478)	(0.0321)	(0.0663)	(0.0431)	(0.00291)	(0.00195)	(0.143)	(0.102)
Majority population, Coloured	-0.281***	-0.282***	-0.481***	-0.443***	-0.00515	-0.0108	-0.185	-0.280
	(0.0596)	(0.0540)	(0.0846)	(0.0735)	(0.00337)	(0.00683)	(0.172)	(0.356)
Majority population, White	-0.250*	-0.222**	-0.491***	-0.417***	-0.000204	-0.00725	0.0997	0.0291
	(0.130)	(0.108)	(0.188)	(0.148)	(0.00609)	(0.0168)	(0.288)	(0.868)
HIV prevention, condom					-0.00748*	-0.00694**	-0.0646	-0.248
					(0.00411)	(0.00309)	(0.211)	(0.176)
Constant	-3.967**	-4.351***	-8.730***	-9.306***	0.0962	0.0957	-5.898	0.190
	(1.561)	(1.066)	(2.293)	(1.525)	(0.172)	(0.171)	(7.763)	(8.754)
Observations	1,408	2,739	1,408	2,739	1,030	2,063	1,030	2,063
R-squared	0.051	0.089			0.025	0.005		
Anderson statistics (p-value)	0.0159	0.0000			0.0003	0.0320		

• Standard errors in parentheses • *** p<0.01, ** p<0.05, * p<0.1 indicates significance at 1%, 5% and 10%
• Definition of instruments follows from table 2.2.
• Majority population reference is Black
[1] The analysis in this table replaces our original social groups (income, gender, race and age) with p**lacebo social groups** using ad-hoc variables. In comparison with our original model, none of the social interaction variables in this table are significant.

Our last falsification test of our social group composition involves using alternative variables in our datasets. Specifically, data from two questions in the CAPS wave 3 questionnaire is used: "*Have you ever had sex*?" where the response option were yes or no and the second question is "*How many of your friends have had sex*?" where the response options were "*most of them*", "*some of them*", "*few of them*", and "*none of them*". Our first expectation is that there is likely to be a strong positive correlation between these two questions as a result of social interaction. That is, whether the young adult is sexually active will be correlated with whether the young adult's friends are sexually active. Secondly, if the average of the first question ("*Have you ever had sex*?") by our social groups is derived, then this average behaviour should also be correlated with the first question. That is, if our social grouping is a true representation of the youths' *social circle* then the average behaviour of the social group in terms of how sexually active the youths are will also be correlated with whether the youth is sexually active. Thirdly, if indeed our social grouping is a true representation of the youths' *social circle*, then the average behaviour based on the *placebo social grouping* will not be correlated with whether the youth is sexually active or not. Table 2.3 depicts the correlation of the above scenarios.

Table 2.3: Tetrachoric Correlation[1]

		(1)	(2)	(3)	(4)
• Have had sex (1)	Tetrachoric correlation	1.000			
	Standard error	0.000			
• Friends have had sex (response from young adult) (2)	Tetrachoric correlation	**0.695**	1.000		
	Standard error	0.021	0.000		
	2-sided exact P[2]	0.000			
• "Friends" have had sex (average of social group) (3)	Tetrachoric correlation	**0.568**	0.428	1.000	
	Standard error	0.025	0.024	0.000	
	2-sided exact P	0.000	0.000		
• "Friends" have had sex (average of *placebo social group*) (4)	Tetrachoric correlation	**0.251**	0.288	0.178	1.000
	Standard error	0.031	0.028	0.029	0.000
	2-sided exact P	0.000	0.000	0.000	

• Number of observation: 2919

[1] Tetrachoric correlation is specifically designed for measuring correlation between binary variables. We therefore converted the responses from "*How many of you friends have had sex*?" to binary representing "none of my friends" and "some of my friends". Tetrachoric correlation is the only one applicable to our data which is mostly categorical in nature, the other correlations offered by STATA are designed for continuous data (pairwise correlation, and partial correlation), and hence using the correlation would bias our results.

[2] The p-value is based on the null hypothesis that the variables are independent (2-sided exact P); that is, 1 and 2; or 2 and 3; or 3 and 4; or 1 and 4.

As can be seen and as per our expectation the correlation between whether the young adult has had sex (1) and whether their friends have had sex (2) is quite high at 0.695, furthermore the p-value of 0.0000 implies the failure to accept the null hypothesis of independence between (1) and (2).

Additionally and as per our expectation the correlation between whether the youth has had sex (1) and the average of whether the social group is sexually active (3) is also relatively high, 0.568, and positive and also fails to accept the null hypothesis.

Finally the correlation between whether the youth has had sex (1) and the average of the *placebo social group* (4) is as expected very low, 0.251. These findings imply that the *social group radius* and the *youth friendship radius* are likely to be equivalent, hence validating our social group classification.

2.8 Conclusion

To reiterate, this study sought to investigate whether *social norms* and the associated *social pressure* influence an individuals' choice to engage in multiple sexual partnerships. To address the question alternative estimation strategies are used. These included using different estimation strategies – cross sectional and panel data methods to establish whether our findings will be consistent. The data used is the CAPS data on young adults in Cape Town, South Africa. Further to this, the analysis was based on different types of populations, namely the *sexually active population* and the *sexually inactive population*. To control for sample selection bias which can occur as a result of including the *sexually inactive population* a censored two stage model (CTSM) is employed, where in the first stage the Mills ratio is derived and then used as an additional regressor in the second stage outcome model (Wooldridge 2002; Greene 2004; Okten and Osili 2004; Bushway *et al.* 2007; Bezabih *et al.* 2010).

To circumvent the identification problem inherent in models of social interaction various model specifications and a menu of IV strategies are employed. The model specifications include using both *linear-in-mean*s models and *non-linear-in-means* models owing to the fact that these models require different identification requirements (Blume *et al.*, 2010). These different model specifications are deployed in both cross sectional and panel data methods taking advantage of the cross sectional nature of the CAPS (wave 3) data and the panel nature of CAPS (using waves 1, 3 and 4).

The instruments based on the work of Grodner and Kniesner (2005); Etile (2007) and Grodner *et al.* (2011) include the *trend setter of the norms*, *adjacent social group age bracket* and *one and two year lags*. After providing a background to the validity of our instruments the Anderson canon correlation statistics test of underidentification and the Sargan-Hansen and Amemiya-Lee Newey statistics test of overidentification were employed. Lastly to validate our social grouping the falsification test of a '*placebo treatment*', based on the work of Lavy and Schlosser (2007) is employed, where social groups are constructed using ad hoc variables. Under this specification no evidence of social interaction is expected.

The current literature on multiple partnerships discusses the persistent lack of change in multiple partnerships in sub-Saharan Africa (Wellings *et al.* 2006; Morris 2010). The literature also points to the fact that 80% of new infection is through heterosexual transmission (UNAIDS 2010a). As such the role of multiple partnerships in driving heterosexual transmission of HIV/AIDS has been at the heart of most research on HIV/AIDS prevention initiatives in the region. It becomes almost impossible to disentangle multiple partnership prevention from HIV/AIDS prevention in regions such as sub-Saharan Africa.

A question that has persisted over time is what drives multiple partnerships in sub-Saharan Africa. Most studies have in the past looked at individual level factors. The study explores the possibility of social interaction being an equally powerful determinant. The results do indeed show that *social norms* and *social pressure* that occurs as a result of social interaction are significant in influencing multiple partnerships. The same conclusion is obtained using alternative estimation strategies. As such our findings shed some light on the reasons for the persistence of multiple partnerships in sub-Saharan Africa by going beyond an investigation of individual level factors.

The results support the findings of the current qualitative research of Ragnarsson *et al.* (2009), Selikow *et al.* (2009) and Ragnarsson *et al.* (2010) that suggest that *social norms* and the related *social pressure* are responsible for multiple partnerships. In addition the results also show that for both men and women the choice of number of sex partners is affected by prevailing *social norms* and *social pressure*. These findings confirm the results of studies in Latkin *et al.* (2005)'s literature review and also qualitative research by Selikow *et al.* (2009). From a policy perspective this has important implications for HIV prevention initiatives such as the '*zero new HIV infection*' vision promoted by UNAIDS. This is because the *social norms* influence the number of sex partners an individual is likely to have. As such it becomes apparent that in order to change sexual behaviour such as multiple partnerships and achieve a zero new infection rate interventions that can change prevailing *social norms* need to be established.

We acknowledge the following methodological limitations. Firstly, it is well documented that data and modelling limit the social interaction empirical literature. Our dataset is not an exception in that it does not provide information on social groups. As such great efforts were placed in modelling these social groups. However, although all efforts were made to construct the relevant social groups we cannot deny the fact that theory *open-endness* (Blume *et al.*, (2010) defines *open-endness* as social interaction models lacking theoretical structures that can naturally identify variables that can be included or excluded in the empirical model.) on social interaction literature makes it difficult to determine social groups composition with certainty. This makes it easy for such social interaction models to be vulnerable to measurement error. This problem continues to plague social interaction models. A further limitation is that we do not have any information that can substantiate that individuals do indeed interact based on the aforementioned characteristics in our dataset, for example evidence from telephone records. As such our groups are only proxies for social groups and should be taken as just that. Future research could overcome this limitation by using datasets that contains social network data, especially social network data that are not geographically influenced, for example social interactions via Facebook dataset as modelled (where geography is not a determinant in social group formation) in the current paper.

2.9 Reference

Adetunji, J. and Meekers, D. (2001). Consistency in condom use in the context of HIV/AIDS in Zimbabwe, *Journal of Biosocial Science,* 33: 121-138.

Adimora, A. and Schoenbach, V. (2005). Social context, sexual networks, and racial disparities in rates of sexually transmitted infections, *The Journal of Infectious Diseases*, 191: 115-122.

Airhihenbuwa, C. and Webster, J. (2004). Culture and African contexts of HIV/AIDS prevention, care and support, *Journal of Social Aspects of HIV/AIDS Research Alliance*, 1 (1): 4-13.

Akerlof, G. (1997). Social distance and social decisions, *Econometrica*, 65 (5): 1005-1027.

Akerlof, G. and Kranton, R. (2005). Identity and the economics of organisations, *The Journal of Economic Perspective*, 19 (1): 9-32.

Akwara, P., Madise, N. and Hinde, A. (2003). Perception of risk of HIV/AIDS and sexual behaviour in Kenya, *Journal of Biosocial Science,* 35: 385-411.

Alexander C., Piazza M., Mekos D. and Valente T. (2001). Peers, schools and adolescent cigarette smoking, Journal of Adolescent Health, 29: 22-30.

Anderson, K., and Beutel, A. (2007). HIV/AIDS prevention knowledge among youth in Cape Town, South Africa, *Journal of Social Sciences*, 3 (3): 143-151.

Anandarajan M., Igbaria M. and Anakwe U. (2000). Technology acceptance in the banking industry: A perspective from a less developed country, Information Technology and People, 13 (4): 298-312.

Angrist, J. and Pischhke, J. (2008). Mostly harmless econometrics: An empiricist's companion, Massachusetts Institute of Technology and London School of Economics.

Bamberg, S., Huneke, M., and Blobaum, A. (2007). Social context, personal norms and use of public transportation: Two field studies, *Journal of Environmental Psychology*, 27: 190-203.

Bamberg, S. and Moser, G. (2007). Twenty years after Hines, Hungerford and Tomera, *Journal of Environmental Psychology*, 27: 14-25.

Benefo, K. (2008). Determinants of Zambian men's extra-marital sex: A Multi-level Analysis, *Arch Sex Behaviour*, 37: 517-529.

Bernheim, B. (1994). A theory of conformity, *Journal of Political Economy*, 102 (5): 841-877.

Bertrand J. (2010). Diffusion of Innovations and HIV/AIDS, Journal of Health Communication: International Perspectives, 9 (S1): 113-121.

Benz, S. (2005). The killing of the fittest, quantitative analysis of HIV/AIDS and conflict. Assessed 10 June 2010. hhtp://kops.ub.uni-konstanz.de/voltexte/2005/1473.

Bezabih, M., Mannberg, A. and Visser, M. (2010), Sexual risk taking among young adults in Cape Town: Effects of expected health and income, SALDRU Working Paper Number 43.

Blume, L., Durlauf, S., Brock, W. and Ionnides Y. (2010), Identification of social interactions, *Handbook of Social Economics*.

Bollinger, C., and Hirsch, B. (2012). Is earnings nonresponse ignorable? *The Review of Economics and Statistics*.

Bongaarts, J. (2006). Late marriage and the HIV epidemic in sub-Saharan Africa, Policy Research Division Working Papers, *The Population Council*, 216: 1-24.

Bound, J., Jaeger, D. and Baker, R. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogeneous explanatory variable is weak, *Journal of the American Statistical Association*, 90 (430): 443-450.

Bourdieu P. (1987). On the theoretical and practical existence of groups, Berkeley Journal of Sociology, 32: 1-17.

Bracher, M., Santow, G. and Watkins, S. (2003). "Moving" and marrying: Modelling HIV infection among newly-weds in Malawi, *Demographic Research,* 1 (7): 208-244.

Brock, W. and Durlauf, S. (2004). Chapter 54 interaction-base models, *Handbook of Econometrics*, 5: 3297-3380.

Brock, W. and Durlauf, S. (2007). Identification of binary choice models with social interactions, *Journal of Econometrics*, 140: 52-75.

Buhi, E. and Goodson, P. (2007). Predictors of adolescent sexual behaviour and intention: A theory-guided systematic review, *Journal of Adolescent Health*, 40: 4-21.

Burgard, S. and Lee-Rife, S. (2009). Community characteristics, sexual initiation, and condom use among young Black South Africans, *Journal of Health and Social Behaviour,* 50: 293-309.

Bushway, S., John, B. and Slocum, L. (2007). Is the magic still there? The use of the Heckman two-step correction for selection bias in criminology, *Journal of Quantitative Criminology*, 23 (2): 9024-9040.

Callegari, L., Harper, C., Straten, A., Kamba, M., Chipato, T. and Padian, N. (2008), Consistent condom use in married Zimbabwean women after a condom intervention, *Sexually Transmitted Diseases*, 35 (6): 624-630.

Cameron C. and Trivedi P. (2005). Microeconometrics: Methods and applications. Cambridge University Press, Cambridge.

Carter, M., Kraft, J., Koppenhaver, T., Galavotti, C., Roels, T., Kilmarx, P. and Fidzani, B. (2007). "A bull cannot be contained in a single kraal": Concurrent sexual partnerships in Botswana, *AIDS Behaviour*, 11: 822-830.

Christensen K. (1998). Economics without money, sex without gender: A critique of Philipson and Posner's private choices and public health: the AIDS epidemic in an economic perspective, *Feminist Economics*, 4(2): 1-24.

Cialdini R., and Trost, M. (1998). Social influence: Social norms, conformity and compliance, *Handbook of Social Psychology Influence*, 2: 151-192.

Cleland, J., Ali, M. and Shah, I. (2006). Trends in protective behaviour among single vs. married young women in Sub-Saharan Africa: The Big Picture, *Reproductive Health Matters*, 14 (28): 17-22.

Coates, T., Richter, L. and Caceres, C. (2008). Behavioural strategies to reduce HIV transmission: How to make them work better, *Lancet*, 372: 669-684.

Cochran, J. and Beeghley, L. (1991). The influence of religion on attitudes toward non-marital sexuality: Assessment of reference group theory, *Journal for the Scientific Study of Religion*, 30 (1): 45-62.

Coleman, E. (2002). Promoting sexual health and responsible sexual behaviour: An introduction, *The Journal of Sex Research*, 39 (1): 3-6.

Conley, T. and Topa, G. (2002). Socio-economic distance and spatial patterns in unemployment, *Journal of Applied Econometrics*, 17: 303-327.

Courtenay-Quirk, C., Pals S., Colfax, G., McKirnan, D., Gooden, L., and Dogan, Eroglu D. (2008). Factors associated with sexual risk behaviour among persons living with HIV: Gender and sexual identity group differences, *AIDS Behaviour*, 12: 685-694.

Csanyi G. and Szendroi B. (2004). Structure of a large social network, Physical Review, 69: 036131-036135

D'Addato A., Vignoli D. and Yavuz S. (2008). Towards smaller family size in Egypt, Morocco and Turkey: Overall change over time or socio-economic compositional effect? Population Review, 47 (1): 21-40.

Deaton, A. (2005). Measuring poverty in a growing world, *Review of Economics and Statistics*, 87 (1): 1-19.

Desgrees, du Lou A. (2005). The couple and HIV/AIDS in Sub-Saharan Africa: Telling the partner, sexual activity and childbearing, *Population,* 60 (3): 179-198.

Dinkelman, T., David, L., and Murray, L. (2007). Household and community income, economic shocks and risky behaviour of young adults: Evidence from the Cape Area Panel Study 2002 and 2005, *AIDS*, 21 (7), S49-S56.

Djamba, Y. (2003). Social capital and premarital sexual activity in Africa: The case of Kinshasa, Democratic Republic of Congo, *Archives of Sexual Behaviour,* 32 (4): 327-337.

Donald, S. and Newey, W. (2001). Choosing the number of instruments, *Econometrica*, 69(5): 1161-1191.

Dunkle, K., Stephenson, R., Karita, E., Chomba, E., Kayitenkore, K., Vwalika, C., Greenberg L. and Susan Allen S. (2008). New heterosexually transmitted HIV infections in married or cohabiting couples in Urban Zambia and Rwanda: An analysis of survey and clinical data, *Lancet,* 371 (9631): 2148-2150.

Durlauf, S. and Ioannides, Y. (2010). Social interactions, *Annual Reviews of Economics*, 2 (45): 451-478.

Dragone D. and Savorelli L. (2012). Thinness and obesity: a model of food consumption, health concerns and social pressure, Journal of Health Economics, 31: 243-256.

East L., Jackson D., Peters K. and O'Brien L. (2009). Disrupted sense of self: young women and sexually transmitted infections, Journal of Clinical Nursing, 19, 1995-2003.

Elster, J. (1985). Rationality, morality and collective action, *Ethics*, 96 (1): 136-155.

Elster, J. (1989). Social norms and economic theory, *The Journal of Economic Perceptive*, 3 (4): 99-117.

Epstein, H. (2010). The mathematics of concurrent partnerships and HIV: A commentary on Lurie and Rosenthal, 2009, *AIDS Behaviour*, 14: 29-30.

Etile, F. (2007). Social norms, ideal body weight and food attitudes, *Health Economics*, 16: 945-966.

Evans, W., Oates, W., Schwab, R. (1992). Measuring peer group effects: A study of teenage behaviour, *The Journal of Political Economy*, 100 (1): 966-991.

Fenton, K., Johnson, A., and McManus, S. (2001). Measuring sexual behaviour: Methodological challenges in survey research, *Sexually Transmitted Infections,* 77: 84-92.

Fehr, E., and Fischbacher ,U. (2004). Social norms and human cooperation, *Trends in Cognitive Sciences*, 8 (4): 186-190.

Ferguson, A., Pere, M., Morris, C., Ngugi, E. and Moses, S. (2004). Sexual patterning and condom use among a group of HIV vulnerable men in Thika, Kenya, *Sexually Transmitted Infection*, 80: 435-439.

Filmer, D. and Prichett, L. (2001). Estimating wealth effects without expenditure data-or tears: An application to education enrolment in states of India, *Demography*, 38 (1): 115-132.

Finn, A., Franklin, S., Keswell, M., Leibbrandt, M. and Levinsohn J. (2009). National income dynamic survey: expenditure, Report on NIDS Wave.

Fischer, P. and Huddart S. (2008). Optimal contracting with endogenous social norms, *American Economic Review*, 98 (4): 1459-1475.

Fischbacher, U., Gachter S. and Fehr E. (2001). Are people conditionally cooperative? Evidence from a public goods experiments, *Economics Letter*, 71: 397-404.

Fisher, J. (1988). Possible effects of reference group-based social influence on AIDS-Risk behaviour and AIDS prevention, *American Psychologist*, 43 (11): 914-920.

Fletcher, J. (2009). Social interaction and smoking: Evidence using multiple student cohort, instrumental variables and school fixed effects, *Health Economics*, 19: 466-484.

Fletschner, D. and Carter M. (2008). Constructing and reconstructing gender: reference group effects and women's demand for entrepreneurial capital, *The Journal of Socio-Economics*, 37: 672-693.

Gachter, S. and Herrmann B. (2009). Reciprocity, culture and human cooperation: previous insights and new cross-cultural experiment, *Philosophical Transactions of the Royal Society*, 364: 791-806.

Galea, S. and Vlahov, D. (2005). Urban health: Evidence, challenges, and directions, *Annual Review of Public Health*, 26: 341-365.

Gausset, Q. (2001). AIDS and cultural practices in Africa: The case of the Tonga (Zambia), *Social Science and Medicine*, 52: 509-518.

Gillespie, G., Greener, R., Whitworth, J. and Whiteside A. (2007). Poverty, HIV and AIDS: vulnerability and impact in Southern Africa, *AIDS*, 21 (7): S1-S104.

Gilbert A. and Selikow T. (2011). 'The epidemic in this country has the face of a woman': Gender and HIV/AIDS in South Africa, African Journal of AIDS Research, 10 (sup1): 325-334.

Glick, P. and Sahn, D. (2007). Changes in HIV/AIDS knowledge and testing behaviour in Africa: how much and for whom? *Journal of Population Economy*, 20: 383-422.

Galavotti, C., Cabral, R., Lansky, A., Grimley, D., Riley, G. and Prochaska, J. (1995). Validation of measures of condom and other contraceptive use among women at high risk for HIV infection and unintended pregnancy, *Health Psychology*, 14 (6): 570-578.

Garcia, S., Harou, P., Montagne, C. and Stenger, A. (2009). Models for sample selection bias in contingent valuation: An application to forest biodiversity, *Journal of Forest Economics*, 15: 59-78.

Gaviria, A. and Raphael, S. (2001). School-based peer effects and juvenile behaviour, *The Review of Economics and Statistics*, 83 (2): 257-268.

Geary C., Burke H., Neupane S., Catelnau L. and Brown J. (2006). Does MTV reach an appropriate audience for HIV prevention message? Evidence from MTV viewership data in Nepal and Brazil, Journal of Health Communication: International Perspective, 11 (7): 665-681.

Giorgi, G., Pellizari, M. and Redaelli S. (2010). Identification of social interactions through partially overlapping peer groups, *American Economic Journal: Applied Economics*, 2 (2): 241-275.

Gracia, E. and Herrero, J. (2006). Acceptability of domestic violence against women in the European Union: A Multilevel Analysis, *J Epidemiol Community Health*, 60: 123-129.

Greene, W. (2002). *Econometric Analysis*, 5th Edition. Upper Saddle River, New Jersey: Prentice-Hall.

Gregson, S., Todd, J. and Zaba, B. (2009). Sexual behaviour change in countries with generalised HIV epidemics?, Evidence from population cohort studies in sub-Saharan Africa, *Sexually Transmitted Infections,* 85: i1-i2.

Grosse, M., Klasen, S. and Spatz, J. (2005). Creating national poverty profiles and growth incidence curves with incomplete income or consumption expenditure data: An application to Bolivia, Institute for Economic Research (IAI) Discussion Paper Number 129, University of Gottingen.

Groenewold, G., Horstman, R. and Bruijn, B. (2004). Gender and the role of men in reproductive health: Applications in studies on HIV sexual risk-behaviour in Zambia, safe motherhood in Nepal, Netherlands Interdisciplinary Demographic Institute.

Grodner, A. and Kniesner, T. (2005). An empirical model of labour supply with social interactions: Econometric issues and tax policy implications, Centre for Policy Research, Working Paper No. 69.

Grodner, A. and Kniesner, T. (2006). Social interactions in labour supply, *Journal of European Economic Association*, 4 (6): 1226-1248.

Grodner, A., Kniesner, T. and Bishop J. (2011). Social interactions in labour markets, Discussion paper series, IZA DP No. 5934.

Heckman, J. (1979). Sample selection bias as a specification error, *Econometrica*, 47: 153-61.

Helfert, S. and Warschburger, P. (2011). A prospective study on the impact of peer and parental pressure on body dissatisfaction in adolescent girls and boys, *Body Image*, 8: 101-109.

Hendlin Y., Anderson S. and Glantz S. (2009). 'Acceptable rebellion': marketing hipster aesthetics to sell Camel cigarettes in the US, Tobacco Control, 19: 213-222.

Hill, Z., Cleland, J., Ali, M. (2004). Religious affiliation and extramarital sex among men in Brazil, *International Family Planning Perspectives*, 30 (1): 20-26.

Hoff P., Raftery A. and Handcock M. (2002). Latent space approaches to social network analysis, Journal of the American statistical association, 97 (460): 1090-1098.

Hounton, S., Carabin, H. and Henderson, N. (2005). Towards an understanding of barriers to condom use in rural Benin using the health belief model: A cross sectional survey, *BMC Public Health,* 5 (8): 1-8.

Horton, R. and Pam, D. (2008). Putting prevention at the forefront of HIV/AIDS, *Lancet*, 372: 421-422.

Hu H. and Wang X. (2011). How people make friends in social networking sites-A microscopic perspective, Phyica, 391: 1877-1886.

Huebner, A. and Howell, L. (2003). Examining the relationship between adolescent sexual risk-taking and perceptions of monitoring, communication, and parenting, *Journal of Adolescent Health*, 33: 71-78.

James, S., Reddy, P., Ruiter, R., McCauley, A., and van den Borne B. (2006). The impact of an HIV and AIDS life skills program on secondary school student in KwaZulu-Natal South Africa, *AIDS Education and Prevention,* 18 (4): 281-294.

Jesko, H., Lanjouw, J., Lanjouw, P., and Poggi, J. (2000). Combing census and survey data to trace the spatial dimensions of poverty: A case study of Ecudor, *World Bank Economic Review*, 14 (1): 147-165.

JewkesR. and Morrell R. (2010). Gender and sexuality: Emerging perspectives from the heterosexual epidemic in South Africa and implications for HIV risk and prevention, *Journal of the International AIDS Society*, 13 (6): 1-11.

Kelly J., Larence J., Stevenson Y., Hauth A., Kalichman S., Diaz Y., Brasfield T., Koob J. and Morgan M. (1992). Community AIDS/HIV risk reduction: The effects of endorsements by popular people in three cities, American Journal of Public Health, 82 (11): 1483-1489.

Kelly J., Murphy D., Sikkeme K. and Kalichman S. (1993). Psychological interventions to prevent HIV infection are urgently needed, American Psycologist, 48 (10): 1023-1034.

Kennedy, S., Nolen, S., Applewhite, J., Pan, Z., Shamblen, S., and Vanderhoff, K. (2007). A quantitative study on the condom-use behaviours of eighteen- to twenty-four–year-old urban African American males, *AIDS Patient Care and STDs*, 21 (5): 306-320.

Kohler H., Behrman J. and Watkins S. (2007). Social networks and HIV/AIDS risk perceptions, Demography, 44 (1): 1-33.

Kolenikov, S. and Angeles, G. (2008). Socioeconomic status measurement with discrete proxy variables: Is principal component analysis a reliable answer? *Review of Income and Wealth*, 55(1): 128-165.

Kolenikov, S. (2009). Confirmatory factor analysis: confa, *Stata Journal*, 9 (3): 329-373.

Kongnyuy, E. and Wiysonge, C. (2007). Alcohol use and extramarital sex among men in Cameroon, *BMC International Health and Human Rights*, 7 (6): 1-7.

Lam, D., Ardington, C., Branson, N., Case, A., Leibbrandt, M., Menendez, A., Seekings, J. and Sparks, M. (2008). *The Cape Area Panel Study: A Very Short Introduction to the Integrated Waves 1-2-3-4 Data,* The University of Cape Town.

Lamptey, P., Johnson, J., and Khan, M. (2006). The global challenge of HIV and AIDS, *Population Bulletin*, 61 (1): 1-24.

Larson, A. (1989). Social context of HIV transmission in Africa: Historical and cultural bases of East and Central African sexual relations, *Reviews of Infectious Diseases*, 11 (5): 716-731.

Latkin C. and Knowlton A. (2006). Micro-social structural approaches to HIV prevention: A social ecological perspective, AIDS Care: Psychological and Socio-medical Aspects of AIDS/HIV, 17 (S1): 102-113.

Latkin, C., Donnell, D., Celentano, D., Aramrattna, A., Liu, T., Vongchak, T., Wiboonnatakul, K., Davis-Vogel, A., and Metzger, D. (2009). Relationships between social norms, social network characteristics, and HIV risk behaviours in Thailand and the U.S, *Health Psychology,* 28(3): 323-329.

Lavy, V. and Schlosser, A. (2007). Mechanism and impacts of gender peer effects at school, NBER Working Paper No. 13292.

Liljeros F., Edling C. and Amaral L. (2003). Sexual networks: Implication for the transmission of sexually transmitted infections, Microbes and Infection, 5:189-196.

Logan, T., Cole, J. and Leukefeld, C. (2002). Women, sex, and HIV: Social and contextual factors, meta-analysis of interventions, and implications for practice and research, *Psychological Bulletin*, 128 (6): 851-885.

Lundborg, P. (2006). Having the wrong friends? Peer effects in adolescent substance use*, Journal of Health Economics*, 25: 214-233.

Lurie, M. and Rosenthal, S. (2010a). Concurrent partnerships as a driver of the HIV epidemic in Sub-Saharan Africa?, The evidence is limited, *AIDS Behaviour*, 14: 17-24.

Lurie, M. and Rosenthal, S. (2010b). The concurrency hypothesis in Sub-Saharan Africa: Convincing empirical evidence is still lacking. response to Mah and Halperin, Epstein, and Morris, *AIDS Behaviour*, 14: 34-37.

Lynch I., Brouadard P. and Visser M. (2010). Constructions of masculinity among a group of South Africa men living with HIV/AIDS: Reflections on resistance and change, An International Journal for Research, Intervention and Care, 12 (1): 15-27.

Maasen, I., Kremer, S., Mudde, A. and Joof, B. (2004). Smoking initiation among Gambian adolescent: social cognitive influences and the effect of cigarette sampling, *Health Education Research*, (19)5: 551-560.

McDonald, J. and Moffit, R. (1980). The uses of Tobit Analysis, *The Review of Economics and Statistics*, 62: 318-312.

Macintyre, S., Ellaway, A. and Cummins, S. (2002). Place effects on health: How can we conceptualise, operationalise and measure them? *Social Science and Medicine*, 55: 125-139.

Magnani, R., Seiber, E., Gutierrez, E. and Vereau, D. (2001). Correlates of sexual activity and condom use among secondary-school students in urban Peru, *Studies in Family Planning*, 32 (1): 53-66.

Mah, T. and Halperin, D. (2010a). Concurrent sexual partnerships and the HIV epidemics in Africa: evidence to move forward, *AIDS Behaviour*, 14: 11-16.

Mah, T. (2010b). Prevalence and correlates of concurrent sexual partnerships among young people in South Africa, *Sexually Transmitted Diseases,* 37 (2): 105-108.

Mah, T. and Halperin, D. (2010c). The evidence for the role of concurrent partnerships in Africa's HIV epidemics: A response to Lurie and Rosenthal, *AIDS Behaviour*, 14: 25-28.

Mahdavi P. (2009). But what if someone sees me?: Women, risk and the aftershock of Iran's sexual revolution, Journal of Middle East Women's Studies, 5 (2): 1-22.

Mahonen, T., Jasinskaya-Lahti, I., Liebkind, K. and Finell, E. (2010). Perceived normative pressure and majority adolescents' implicit and explicit attitudes towards immigrants, *International Journal of Psychology*, 45 (3): 182-189.

Mannberg, A. and Sjogren, T. (2010). Conflicting identities and social pressure: Effects on the long-run evolution of female labour supply, Umea Economic Studies Number 805.

Mannberg, A. (2010). Risky sex in a risky world: Sexual behaviour in a HIV/AIDS environment, *Scandinavian Journal of Economics*, 114 (2): 296-322.

Mannberg A. (2012). Risk and rationalization: The role of affect and cognitive dissonance for sexual risk taking, *European Economic Review*, 56: 1325-1337.

Manski, C. (1993). Identification of endogenous social effects: The reflection problem, *The Review of Economic Studies*, 60 (3): 531-542.

Manski, C. (2000). Economic analysis of social interactions, *The Journal of Economic Perspectives*, 14 (3): 115-136.

Manski, C. (2003). Identification problems in the social sciences and everyday life, *Southern Economic Journal*, 70 (1): 11-21.

Manski, C. and Mayshar, J. (2003). Private incentives and social interactions: Fertility puzzles in Israel, *Journal of the European Economic Association*, 1 (1): 181-211.

McQueen, D. (1996). The Research for theory in health behaviour and health promotion, *Health Promotion International*, 11 (1): 27-32.

Meek, W., Pacheco, D. and York, J. (2009). The impact of social norms on entrepreneurial action: Evidence from the environmental entrepreneurship context, *Journal of Business Venturing*, 25 (5): 1-17.

Moffitt, R. (2003). Policy interventions, low-level equilibria and social interactions, *Journal of Applied Econometrics,* 18 (5): 45-82.

Morris, M. (1997). Sexual network and HIV, *AIDS*, 11 (Supplement A): 209-216.

Morris, M. (2010). Barking up the wrong evidence tree. Comment on Lurie and Rosenthal, "Concurrent partnerships as a driver of the HIV epidemic in Sub-Saharan Africa? The evidence is limited", *AIDS Behaviour*, 14: 31-33.

Moulton R. (1990). An illustration of a pitfall in estimating the effects of aggregated variables on micro units, The Review of Economics and Statistics, 72 (2): 334-338.

Munshi, K. and Myaux, J. (2006). Social norms and the fertility transition, *Journal of Development Economics*, 80: 1-38.

Nattrass, N. (2009). Poverty, sex and HIV, *AIDS Behaviour*, 13: 833-840.

Neubourg C. and Vendrik M. (1994), An extended rationality of social norms in labour supply, Journal of Economic Psychology, 15: 93-126.

Noar, S. and Zimmerman, R. (2005). Health behaviour theory and cumulative knowledge regarding health behaviours: Are we moving in the right direction? *Health Education Research*, 20 (3): 275-290.

Njau, B., Mtweve, S., Barongo, L., Manongi, R., Chugulu, J., Msuya, M, Mwampeta, S., Kiwale, B., Lekule, J. and Jalipa, H. (2006). The influence of peers and other significant persons on sexuality and condom-use among young adults in Northern Tanzania, *African Journal of AIDS Research,* 5(3): 33-40.

Obermeyer, C. (2005). Refraining research on sexual behaviour and HIV, *Studies in Family Planning*, 36 (1): 1-12.

Okten, C. and Osili, U. (2004). Social networks and credit access in Indonesia, *World Development*, 32(7): 1225-1246.

Pettifor A., Measham D., Rees H. and Padian N. (2004). Sexual power and HIV risk, South Africa, *Emerging Infectious Diseases*, 10 (11): 1996-2004.

Philipson T. and Posner R. (1995). The microeconomics of the AIDS epidemic in Africa, *Population and Development Review*, 21 (4): 835-848.

Pinkerton S., Holtgrave D., DiFranceisco W., Stevenson Y. and Kelly J. (1998). Cost-effectiveness of a community level HIV risk reduction intervention, American Journal of Public Health, 88: 1239-1242.

Prandy K. and Lambert P. (2003). Marriage, social distance and the social space: An alternative derivation and validation of the Cambridge scale, Sociology, 37 (3): 397-411.

Pries L. (2001). The disruption of social and geographic space: Mexican-US migration and the emergence of transnational social spaces, International Sociology, 16 (1):55-74.

Ragnarsson, A., Townsend, L., Thorson, A., Chopra, M. and Ekström, A. (2009). Social networks and concurrent sexual relationships: A qualitative study among men in an urban South African community, *AIDS Care*, 21 (10): 1253-1258.

Ragnarsson, A., Townsend, L., Ekström, A., Chopra, M., Thorson, A. (2010). The construction of an idealised urban masculinity among men with concurrent sexual partners in a South African township, *Global Health Action*, 3 (5092): 1-7.

Reagan, P. and Salsberry, P. (2005). Race and ethnic differences in determinants of preterm birth in the USA: Broadening the social context, *Social Science and Medicine*, 60: 2217-2228.

Reniers, G. and Watkins, S. (2009). Polygamy and the spread of HIV in Sub-Saharan Africa: A Case of Benign concurrency, California Centre for Population Research, On-Line Working Paper Series.

Rind B. (2010). Social response to age-gap sex involving minors: Empirical, historical, cross-cultural and cross-species considerations, Journal of Boyhood Studies, 4 (2): 113-115.

Rossem, R., Meekers, D. and Akinyemi, Z. (2001). Consistent condom use with different types of partners: Evidence from two Nigerian surveys, *AIDS Education and Prevention*, 13 (3): 252-267.

Rosenzweig, M. and Wolpin, K. (2000). Natural 'natural experiments', *Journal of Economic Literature*, 38: 827-874.

Salvini S. and Vignoli D. (2011). Things change: Women's and men's marital disruption dynamics in Italy during a time of social transformation, 1970-2003, Demography, 24 (5): 145-174.

Schroeder K. and Rojas F. (2002). A game theoretical analysis of sexually transmitted disease epidemics, *Rationality and Society*, 14: 353-383.

Selikow, T., Ahmed, N., Flisher, A., Mathews, C., Mukoma, W. (2009). I am not "Umqwayito": A qualitative study of peer pressure and sexual risk behaviour among young adolescents in Cape Town, South Africa, *Scandinavian Journal of Public Health*, 37 (Suppl 2): 107-112.

Singla and Richardson (2008). Yes, there is a correlation: From social networks to personal behaviour on the web, Social Networks and Web 2.0- Analysis of Social Networks and Online Interactions.

Smith, R. and Blundell, W. (1986). An exogeneity test for a simultaneous equation Tobit model with an application to labor supply, *Econometrica*, 54: 679-686.

Smith P. and Christakis N. (2008). Social networks and health, Annual Review of Sociology, 34: 405-429.

Staiger, D. and Stock, J. (1997). Instrumental variables regression with weak instruments, *Econometrica*, 65 (3): 557-586.

Subrahmanyann K., Reich S., Waechter N., and Espinoza G. (2008). Online and offline social networks: Use of social networking sites by emerging adults, Journal of Applies Development Psychology, 29: 420-433.

Takyi, B. (2003). Religion and women's health in Ghana: Insights into HIV/AIDS preventive and protective behaviour, *Social Science and Medicine*, 56: 1221-1234.

Thelwall M. (2008), Social networks, gender and friending: An analysis of MySpace Member Profiles, Journal of American Society For Information Science and Technology, 59 (8): 1321-1330.

Townsend L., Ragnarsson A., Mathews C., Johnston L., Ekstrom A., Thorson A. and Chopra M. (2011). 'Taking care of business': Alcohol as currency in transactional sexual relationships among players in Cape Town, South Africa, Qualitative Health Research, 21 (1): 41-50.

Trogdon, J. Nonnemaker, J., Pais, J., (2008). Peer effects in adolescent overweight, *Journal of Health Economics*, 27: 1388-1399.

UNAIDS. (1999). Sexual behavioural change for HIV: Where have theories taken us?, UNAIDS Best Practice Collection, Geneva, Switzerland.

UNAIDS. (2007). UNAIDS Expert consultation on behaviour change, 25-26 September 2006, Joint United Nations Programme on HIV/AIDS, International Conference Centre, Geneva, Switzerland.

UNAIDS. (2008). 2008 Report on global AIDS epidemic - preventing new HIV infections: The key to reversing the epidemic, Geneva, Switzerland.

UNAIDS. (2009). AIDS epidemic update, Joint UNAIDS and WHO, Geneva, Switzerland.

UNAIDS. (2010a). Joint action for results: UNAIDS Outcome Framework, 2009-2011, Joint United Nations Programme on HIV/AIDS Geneva, Switzerland.

UNAIDS. (2010b). Global report: UNAIDS Report on the global AIDS epidemic 2010, Joint United Nations Programme on HIV/AIDS, Geneva, Switzerland.

UNDP. (2003). Millennium development goals: A compact among nations to end human poverty, human development report 2003, United Nations Development Programme.

Urizar, A., Garcia, M., Toro J., Maldonado, J., Penaloza, C., Cuadros-Sosa, Y., and Madrid, J. (2010). Associations between sociocultural pressure to be thin, body distress and eating disorder symptomatology among Chilean adolescent girls, *Body Image*, 8: 78-81.

Vendrick, M. (2001). Dynamics of a household norm in female labour supply, *Journal of Economic Dynamics and Control*, (27): 823-841.

Veentra G. (2007). Social space, social class and Bourdieu: Health inequalities in British Columbia, Canada, Health and Place, 13: 14-31.

Visser, M., Makin, J. and Lehobye, K. (2006). Stigmatizing attitudes of the community towards people living with HIV/AIDS. *Journal of Community and Applied Social Psychology*, 16: 42-58.

Voeten, H., Egesah, O., and Habbema, D. (2004). Sexual behaviour is more risky in rural than in urban areas among young women in Nyanza, Kenya, *Sexually Transmitted Diseases,* 31 (8): 481-487.

Wechsberg W., Luseno W., Kline T., Browne F. and Zule W. (2010). Preliminary findings of an adapted evidence-based woman-focused HIV intervention on condom use and negotiation among at risk women in Pretoria, South Africa, *Journal of Prevention and Intervention in the Community*, 38 (2): 132-146.

Weir, S., Tate, J., Zhusupov, B. and Boerma, J. (2004). Where the action is: Monitoring local trends in sexual behaviour, *Sexually Transmitted Infections,* 80: ii63-ii68.

Wellings, K., Collumbien, M., Slaymaker, E., Singh, S., Hodges, Z., Patel, D. and Bajos, N. (2006). Sexual behaviour in context: A global perspective, *Lancet,* 368: 1706-1728.

Wellman B., Salaff J., Dimitrova D., Garton L., Gulia M. and Haythornwaite C. (1996). Computer networks as social networks: Collaborative work, telework and virtual community, Annual Review of Sociology, 22: 213-238.

Williams R. (2000). A note on robust variance estimation for clustered-correlated data, *Biometrics*, 56: 645-646.

Wilson, D. and Halperin, D. (2008). "Know your epidemic, know your response": A useful approach, if we get it right, *Lancet*, 372: 423-426.

Wingood G. and DiClemente R. (2000). Application of the theory of gender and power to examine HIV related exposure, risk factors and effective interventions for women, *Health Education and Behaviour*, 27 (5): 539-565.

Wojcicki J. and Malala J. (2001). Condom use, power and HIV/AIDS risk: Sex workers bargain for survival in Hillbrow/Joubert Park/Berea, Johannesburg, *Social Science and Medicine,* 53 (1): 99-121.

Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data*, Cambridge, the MIT Press.

Wooldridge, J. (2003). *Introductory Econometrics: A Modern Approach*, 2nd Edition, Thomson South-Western.

Wooldridge J. (2003). Clustered-sample methods in applied econometrics, *The American Economic Review*, 93 (2): 133-138.

World Health Organisation (WHO). (2008). Priority interventions: HIV/AIDS prevention, treatment and care in the health sector, Geneva, Switzerland.

World Health Organisation (WHO). (2010). World health statistics 2010, Geneva, Switzerland.

Young S. and Rice E. (2011). Online social networking technologies, HIV knowledge and sexual risk and testing behaviours among homeless youth, AIDS Behaviour, 15: 253-260.

Zanella, G. (2004). Social interaction and economic behaviour, University of Wisconsin-Madison and University of Siena.

Appendix 2.A: Characteristics of our Sample

Table 2.A1: Overall Characteristics of the CAPS Surveys

	Black	Colured	White	Total	Male	Female
2002 (wave 1)						
Age (mean)	18.1	17.9	17.8	18.0	17.911	18.015
Education (mean years)	8.9	9.3	10.4	9.3	8.9751	9.4764
Household per capita income (mean)	423.82	977.88	3814.48	1112.24	1176.4	1055.2
Marital status (%yes)	0.2	0.9	0.0	0.6	0.3	0.9
Has had sex (%yes)	61.6	34.2	28.6	46.0	45.8	46.2
Age of first intercourse (mean)	16.8	18.1	19.3	17.4	17.3	17.5
Partners in past 12 months (mean)	1.52	1.49	1.54	1.51	2.0	1.1
Knows someone with AIDS (% yes)	21.0	10.1	9.9	15.0	13.2	16.4
Knows someone died from AIDS (% yes)	33.0	13.6	6.7	21.5	18.6	23.9
AIDS: abstain (% yes)	37.3	47.2	55.1	43.7	40.1	46.7
AIDS: use condom (% yes)	92.4	87.8	87.1	89.8	90.3	89.4
AIDS: one sex partner (% yes)	14.4	19.4	24.8	17.8	19.6	21.1
AIDS: limit partners (% yes)	4.2	6.0	13.8	6.2	5.2	7.0
2005 (wave 3)						
	Black	**Colured**	**White**	**Total**	**Male**	**Female**
Age (mean)	21.13	20.76	20.76	20.92	20.86	20.98
Education (mean years)	9.90	10.20	11.61	10.24	9.98	10.46
Household per capita income (mean)	602.73	1244.68	5787.96	1411.01	1517.5	1321.2
Marital status (%yes)	0.9	3.9	0.0	2.4	1.4	3.3
Has had sex (%yes)	88.2	63.6	51.9	73.3	74.1	72.7
Age of first intercourse (mean)	18.0	19.9	20.4	18.9	18.951	18.904
Partners in past 12 months (mean)	1.2	1.1	1.2	1.2	1.3463	1.0213
Partners in lifetime (mean)	2.5	2.0	2.5	2.3	2.7732	1.9483
Condom use with 'last partner' (% yes)						
Always	56.2	55.8	61.73	56.22	64.8	47.41
Usually	5.11	6.91	11.11	6.09	5.61	6.67
Sometimes	21.02	15.19	16.05	18.71	12.41	25.74
never	17.66	22.1	11.11	18.53	17.18	20.19
Relationship with 'last partner'						
spouse/married	4.77	2.2	0.0	3.3	2.16	4.3
someone loved but not married to	85.83	67.3	68.18	76.28	67.68	83.82
someone knew well but didn't love	3.51	19.21	21.59	11.79	17.75	6.57
someone you knew, but not well	5.89	11.29	10.23	8.63	12.41	5.31
Time knew 'last partner' (mean)						
> 1 year	30.1	50.7	47.7	40.6	34.9	45.7
>1 month but < 1 year	38.3	34.1	42.1	36.7	36.0	37.2
> 2 days but < 1 month	24.8	10.0	6.8	16.8	20.1	14.0
two days and <	6.8	5.3	3.4	5.9	9.0	3.2
Knows someone with AIDS (% yes)	46.5	14.1	10.7	27.7	22.1	32.4
Knows someone died from AIDS (% yes)	54.9	14.9	10.7	31.7	27.3	35.4
AIDS: abstain (% yes)	37.1	45.3	70.2	44.3	46.0	48.8
AIDS: use condom (% yes)	89.0	79.7	82.7	84.0	86.8	81.7
AIDS: one sex partner (% yes)	9.8	26.6	37.6	20.5	19.6	21.1
AIDS: limit partners (% yes)	3.6	13.3	26.3	10.4	10.9	10.2
2006 (wave 4)						
	Black	**Colured**	**White**	**Total**	**Male**	**Female**
Age (mean)	22.03	21.7246	21.761	21.862	21.807	21.911
Education (mean years)	10.08	10.2849	11.761	10.376	10.112	10.591
Marital status (%yes)	1.5	6.6	0.0	3.9	2.6	5.6
Has had sex (%yes)	93.9	70.4	61.2	79.4	82.1	77.0
Age of first intercourse (mean)	16.6	19.4	22.6	18.0	18.126	17.939
Partners in past 12 months (mean)	1.6	1.3	1.3	1.5	1.9	1.1
Partners in lifetime (mean)	3.6	2.1	2.3	3.0	3.9	2.5

Appendix 2.B: Distribution of Social Groups

Table 2.B1: Distribution of social groups by place and sub-place

	Atlantis	Bellville	CT-Bellville	CT-Cape	CT-Goodwoo	CT-Kuilsrivier	CT-Mitchells Plain(MD)																																				CT-Simons T	Gordons B	Macassar	Malmesbur				Total
Social Groups							Beacon	Browns farm	Browns farm mac	Colorado	Eastridge	Gugulethu				Khayelits ilitha	Khayelits maca.	Khayelits sectA	Khayelits sectC	Khayelits sectE	Khayelits sectG	Khayelits sectJ	Khayelits siteB	Khayelits siteC	Khayelits vila.	Khayelits town2	Lentegeur	Lower Cro	New Cross	Nyanga	Nyanga-KT	Nyanga-Mk	Old Cross	Old Cross gqob.	Portlands	Rocklands	Strandfon	Tafelsig	Weltevred	Westridge	Woodlands								
1	0	0	0	0	2	3	0	3	1	0	0	5	•	•	•	0	7	1	0	2	3	0	4	6	0	1	0	0	2	2	4	1	2	0	0	0	0	0	0	0	0	1	0	0	0	•	•	•	63
2	0	0	1	0	0	1	0	2	1	0	0	2				1	3	0	2	0	0	1	3	6	2	1	0	1	0	0	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0				33
4	0	0	0	1	3	6	0	4	0	0	0	4				1	2	1	1	1	0	1	2	0	3	0	0	0	2	3	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0				42
5	0	0	0	1	3	2	0	1	0	0	0	3				0	2	0	0	0	0	0	3	4	3	0	0	0	0	0	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0				28
6	0	0	0	0	0	0	0	0	0	0	0	3				1	0	1	1	1	1	1	2	5	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0				20
7	0	0	0	1	3	1	0	1	0	0	0	4				1	1	1	1	2	0	0	4	1	2	1	0	0	4	6	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0				40
8	0	0	0	0	2	1	0	1	0	0	0	5				2	1	0	0	1	0	0	1	3	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0				23
9	0	0	0	0	1	0	0	0	0	0	0	1				0	1	0	0	2	1	0	3	2	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0				16
10	0	0	0	0	4	3	0	4	0	0	0	6				0	3	0	0	0	0	0	3	2	1	1	0	0	3	3	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0				38
11	0	0	0	0	1	2	0	2	0	0	0	4				0	4	0	0	1	0	0	6	4	3	1	0	0	2	3	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0				40
12	0	0	0	0	2	1	0	0	0	0	0	1				0	0	0	1	1	0	0	4	6	0	1	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0				21
13	1	0	2	0	7	2	1	0	0	0	1	0				0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	1	0	1	0	0	3				29
14	2	1	3	2	5	7	1	0	0	0	0	0				0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	3	0	1	2	0	1	0				46
15	2	1	5	2	9	6	1	0	0	0	1	0				0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	1	0	0	3	0	2	0	2	0	0	0				57
16	2	0	2	0	5	11	1	0	0	0	0	0				0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	2	2	0	1	1	0	0	0	0				49
18	0	0	5	0	9	1	1	0	0	0	0	0				0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	3	0	2	1	1	0	0	0				35
19	6	0	5	1	14	7	4	0	0	0	0	0				0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	1	2	0	5	0	3	1	4	0	1	0				81
20	0	0	3	5	7	5	0	0	0	1	1	0				0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	2	0	0	1	1	1	0	1	0				44
21	2	0	5	1	6	1	0	0	0	0	0	0				0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	1	0	0	1	0	1	0				31
22	5	0	4	1	11	5	1	0	0	1	0	0				0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	2	1	0	2	0	0	1	2				47
23	2	0	2	3	2	7	0	0	0	0	0	0				0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	1	0	2	0	0	0				43
24	3	0	1	1	7	1	2	0	0	0	0	0				0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0				22
25	2	0	1	1	7	3	1	0	0	0	1	0				0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	3	0	4	0	3	4	1	0	1	1				40
26	2	0	2	5	3	10	0	0	0	1	1	0				0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	1	1	2	0	0	1	0	0	1	0				49
27	0	0	7	4	8	4	0	0	0	0	0	0				0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0				43
28	0	0	6	3	8	2	0	0	0	0	0	0				0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	0	0				35
29	0	0	5	2	4	3	0	0	0	0	0	0				0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0				27
30	0	0	0	0	5	1	0	4	1	0	0	3				0	5	0	1	1	1	0	5	7	0	1	0	0	1	6	1	0	0	2	0	0	0	1	0	0	0	0	0	0	0				56
31	0	0	0	0	1	1	0	2	0	0	0	4				0	2	0	0	1	0	0	6	8	3	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0				37
•																																																	
•																																																	
•																																																	

Appendix 2.C: Comparison between Sexually Active and Sexually Inactive Population

Table 2.C1: Comparison between sexually active and sexually inactive population

	Sexually Active population	Sexually Inactive population	Total population
wave 1 (2002)			
age	18.9	16.4	17.3
population group	1.5	2.0	1.8
gender	1.5	1.5	1.5
education in years	9.6	8.7	9.0
age of sexual debut	17.4		17.4
wave 3 (2005)			
age	20.8	18.9	20.2
population group	1.6	2.1	1.8
gender	1.5	1.6	1.5
education in years	10.3	10.6	10.4
age of sexual debut	18.7		18.7

Appendix 2.D: Results of the First Stage CTSM

Table 2.D1: First Stage - Heckman Model

	(2)
Dependent variable: Had Sex	CAPS
Age	0.293***
	(0.0686)
Age squared	-0.00577***
	(0.00165)
Gender	-0.0282*
	(0.0149)
Income	7.72e-07
	(3.58e-06)
Education	-0.0221***
	(0.00455)
HIV prevention, one partner	-0.128***
	(0.0179)
HIV prevention, abstain	0.0549**
	(0.0262)
Coloured	-0.181***
	(0.0209)
White	-0.201***
	(0.0340)
Knows someone with HIV	-0.0462**
	(0.0208)
Constant	-2.373***
	(0.695)
Select	
HIV prevention, condom	-0.0365
	(0.0750)
Constant	1.608***
	(0.0811)
Mills	
lambda	-0.874***
	(0.0154)
Observations	2,315

- Standard errors in parentheses
- *** p<0.01, ** p<0.05, * p<0.1 indicates significance at 1%, 5% and 10%
- Population group reference is Black

Appendix 2.E: Social Norm and Social Pressure using adjacent social group age bracket IV

In this section we adopt the *adjacent social group age bracket IV* from the works of Grodner and Kniesner (2005); Etile (2007) and Grodner *et al.* (2011). The assumption is that the behaviour of age group $\bar{y}_{21-25}$ for example will be correlated with behaviour of adjacent age group to the left, $\bar{y}_{15-20}$ and to the right, $\bar{y}_{26-30}$ because of similar socioeconomic neighbourhoods, however the adjacent social group behaviour $\bar{y}_{15-20}$ and $\bar{y}_{26-30}$ are uncorrelated with unobservables of individuals in social group $\bar{y}_{21-25}$ (for example unobservables of a 21year old) because these individuals are not members of the adjacent groups (Grodner and Kniesner 2005; Etile 2007; Grodner *et al.* 2011). Therefore, adjacent group behaviour to the left and to the right of the social group can be used as IVs. The advantage of the IVs is that it allows for overidentification of the models[50]. The limitation is that the IVs exclude the youngest group and oldest group as these groups lack IVs[51].

[50] We also employ the *'extended' adjacent social group IV*. This is based on *adjacent social group age bracket,* with the only difference being that we include the youngest social group and oldest social group by using social groups with a different characteristics, for example replace the oldest social group if males with female oldest social group. This is so as to include the total sample size since the adjacent social group IV omits two age groups; the first age group and the last age group.

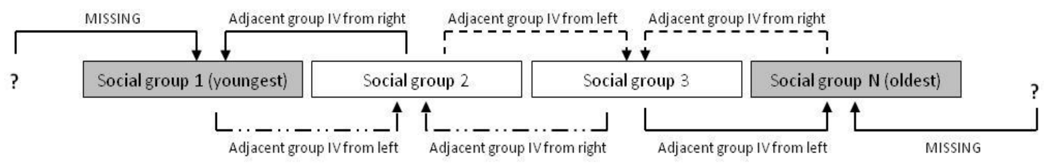
[51] 

Table 2.E1: Social Norm and Social Pressure Results

	Panel A							
	Both *Sexually Active Population* and *Sexually Inactive Population*							
	IVTOBIT CTSM 2nd stage				GMM CTSM 2nd stage			
	Nonlinear in means model No. of sex partners incl. not sexually active				Linear in means model No. of sex partners incl. not sexually active			
	Adjacent group IV[1]	'Extended' Adjacent group IV[1]	Trend setter IV[2]	'Extended' Trend setter IV[2]	Adjacent group IV[1]	'Extended' Adjacent group IV[1]	Trend setter IV[2]	'Extended' Trend setter IV[2]
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Social norm	**0.754*****	**1.138*****	**1.191*****	**1.076*****	**0.546*****	**0.538*****	**0.0218**	**0.0991**
	(0.186)	**(0.251)**	**(0.321)**	**(0.243)**	**(0.155)**	**(0.184)**	**(0.216)**	**(0.174)**
Social norm*social pressure	**2.445*****	**0.740**	**0.451**	**0.822***	**3.138*****	**2.050*****	**3.027*****	**3.029*****
	(0.851)	**(0.467)**	**(0.580)**	**(0.432)**	**(0.707)**	**(0.341)**	**(0.397)**	**(0.312)**
Social group characteristics	-0.109	0.00425	-0.00323	-0.0165	-0.137	0.157**	0.0616	0.0316
	(0.111)	(0.0933)	(0.0776)	(0.0680)	(0.0919)	(0.0672)	(0.0542)	(0.0505)
Education years	-0.0120	-0.00942	-0.00829	-0.00884	-0.00557	-0.0142*	-0.00570	-0.00684
	(0.00889)	(0.00990)	(0.0120)	(0.00933)	(0.00739)	(0.00725)	(0.00836)	(0.00689)
Mills ratio	41.19***	61.66***	75.50***	61.86***	28.74***	33.16***	38.44***	34.44***
	(7.245)	(9.298)	(11.86)	(9.289)	(5.848)	(6.532)	(7.707)	(6.501)
HIV prevention, abstain	0.120***	0.244***	0.291***	0.242***	0.0832***	0.152***	0.170***	0.149***
	(0.0290)	(0.0377)	(0.0478)	(0.0377)	(0.0242)	(0.0279)	(0.0331)	(0.0277)
HIV prevention, one partner	-0.104**	-0.126**	-0.139*	-0.129**	-0.0768**	-0.0255	-0.0336	-0.0399
	(0.0469)	(0.0600)	(0.0749)	(0.0594)	(0.0390)	(0.0440)	(0.0514)	(0.0434)
Knows someone died of AIDS	0.0343	0.114***	0.146***	0.112***	0.0151	0.0979***	0.108***	0.0864***
	(0.0316)	(0.0410)	(0.0515)	(0.0403)	(0.0266)	(0.0307)	(0.0363)	(0.0302)
Majority population, Coloured	-0.121	-0.0245	-0.0267	-0.0263	-0.143**	-0.218***	-0.316***	-0.252***
	(0.0781)	(0.0874)	(0.129)	(0.0920)	(0.0647)	(0.0644)	(0.0872)	(0.0670)
Majority population, White	0.163	0.0156	0.0424	0.0692	0.188	-0.596***	-0.433**	-0.305*
	(0.295)	(0.301)	(0.273)	(0.237)	(0.242)	(0.216)	(0.188)	(0.175)
HIV prevention, condom								
Constant	-6.155***	-9.458***	-11.55***	-9.456***	-4.262***	-4.655***	-5.205***	-4.699***
	(1.063)	(1.358)	(1.736)	(1.358)	(0.858)	(0.953)	(1.127)	(0.948)
Observations	1,256	2,769	2,302	2,769	1,256	2,769	2,302	2,769
R-squared					0.162	0.174	0.161	0.181
Anderson canon correlation statistics (p-value)					0.0000	0.0000	0.0000	0.0000
Sargan-Hansen statistic (p-value)					0.9636	0.4115		
Amemiya-Lee Newey statistic (p-value)	0.6178	0.6741						

	Panel B							
	Sexually Active Population Only							
	IVPROBIT				GMM			
	Nonlinear model Multiple Partnership (0/1)				Linear in means model No. of sex partners			
	Adjacent group IV[1]	'Extended' Adjacent group IV[1]	Trend setter IV[2]	'Extended' Trend setter IV[2]	Adjacent group IV[1]	'Extended' Adjacent group IV[1]	Trend setter IV[2]	'Extended' Trend setter IV[2]
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Social norm	**4.550***	**1.190*****	**1.372*****	**1.486*****	**0.0200**	**0.00964*****	**0.0105*****	**0.0137*****
	(2.634)	**(0.183)**	**(0.246)**	**(0.225)**	**(0.0288)**	**(0.00309)**	**(0.00365)**	**(0.00301)**
Social norm*social pressure	**15.48**	**13.01****	**12.18****	**11.57****	**0.490*****	**0.405*****	**0.424*****	**0.414*****
	(13.82)	**(5.640)**	**(4.937)**	**(5.126)**	**(0.136)**	**(0.0837)**	**(0.0952)**	**(0.0878)**
Social group characteristics	-1.330	0.650	0.0534	0.138	-0.00156	0.0106*	0.00122	0.00134
	(1.836)	(0.461)	(0.134)	(0.151)	(0.0194)	(0.00629)	(0.00231)	(0.00224)
Education years	0.123	-0.0507	-0.00272	-0.0103	6.83e-05	-0.000957	-5.14e-05	-0.000233
	(0.137)	(0.0469)	(0.0332)	(0.0325)	(0.00143)	(0.000643)	(0.000540)	(0.000452)
Mills ratio								
HIV prevention, abstain	-0.0946	0.0200	-0.0178	-0.0162	-0.000357	0.00116	-0.000496	0.000469
	(0.248)	(0.122)	(0.120)	(0.117)	(0.00272)	(0.00184)	(0.00213)	(0.00178)
HIV prevention, one partner	-0.220	0.0877	-0.0551	-0.00433	0.000374	0.00316	0.00180	0.00147
	(0.488)	(0.212)	(0.195)	(0.192)	(0.00516)	(0.00307)	(0.00339)	(0.00285)
Knows someone died of AIDS	-0.343	0.115	0.00197	0.0320	-0.00214	0.00230	0.00138	0.000961
	(0.316)	(0.141)	(0.128)	(0.125)	(0.00332)	(0.00200)	(0.00218)	(0.00183)
Majority population, Coloured	1.814	-0.674	-0.0145	-0.0297	-0.00285	-0.0167**	-0.00537	-0.00502
	(2.363)	(0.592)	(0.222)	(0.226)	(0.0250)	(0.00823)	(0.00379)	(0.00347)
Majority population, White	3.859	-1.760	-0.125	-0.328	0.00400	-0.0287	-0.00191	-0.00245
	(5.072)	(1.332)	(0.440)	(0.481)	(0.0535)	(0.0182)	(0.00754)	(0.00721)
HIV prevention, condom	-0.478	-0.142	-0.169	-0.187	-0.00727*	-0.00398	-0.00463	-0.00475*
	(0.463)	(0.211)	(0.209)	(0.206)	(0.00423)	(0.00276)	(0.00330)	(0.00271)
Constant	-7.009*	-2.015**	-2.913***	-2.948***	0.0350	0.0569***	0.0414***	0.0401***
	(3.939)	(0.930)	(0.538)	(0.529)	(0.0414)	(0.0131)	(0.00855)	(0.00739)
Observations	1,016	1,962	1,524	1,962	1,016	1,962	1,524	1,962
R-squared					0.097	0.099	0.106	0.121
Anderson canon correlation statistics (p-value)					0.0283	0.0000	0.0000	0.0000
Sargan-Hansen statistic (p-value)					0.9229	0.4684		
Amemiya-Lee Newey statistic (p-value)	0.4346	0.0852						

• Robust Standard errors in parentheses
• Majority population reference is Black
• Instrumented for social norm and social norm*social pressure
• *** p<0.01, ** p<0.05, * p<0.1 indicates significance at 1%, 5% and 10%

[1] Excluded instruments: social norm of left social group, social norm of right social group, social norm*social pressure of right social group and social norm*social pressure of left social group.

[2] Excluded instruments: social norm of older social group, social norm*social pressure of older social group

Appendix 2.F: Social Norm and Social Pressure Results by Gender

Table 2.F1: Social Norm and Social Pressure Results among the Male Population

	Both Sexually Active and Not Sexually Active Population								Sexually Active Population Only							
	IVTOBIT CTSM 2nd stage				GMM CTSM 2nd stage				GMM				IVPROBIT			
	Nonlinear in means model No. of sex partners incl. not sexually active				Linear in means model No. of sex partners incl. not sexually active				Linear in means model No. of sex partners incl. not sexually active				Nonlinear in means model No. of sex partners incl. not sexually active			
	(1)[1]	(2)[2]	(3)[3]	(4)[4]	(5)[1]	(6)[2]	(7)[3]	(8)[4]	(9)[1]	(10)[2]	(11)[3]	(12)[4]	(13)[1]	(14)[2]	(15)[3]	(16)[4]
Social norm	**2.821**	**2.853****	**3.035*****	**3.303*****	**0.689**	**0.922****	**0.995****	**1.006***	**49.33****	**4.352**	**18.17****	**10.08**	**3.968**	**3.990****	**8.041**	**1.388**
	(1.979)	**(1.432)**	**(0.798)**	**(1.136)**	**(1.222)**	**(0.441)**	**(0.491)**	**(0.543)**	**(24.01)**	**(10.77)**	**(9.038)**	**(9.295)**	**(6.403)**	**(1.755)**	**(27.88)**	**(0.885)**
Social norm*social pressure	**0.778****	**1.365****	**0.804*****	**0.772****	**0.181**	**0.0747**	**0.0632**	**0.0656**	**0.127**	**1.619****	**0.658*****	**0.803*****	**0.242**	**1.262**	**0.156**	**0.0996**
	(0.340)	**(0.561)**	**(0.236)**	**(0.361)**	**(0.549)**	**(0.492)**	**(0.145)**	**(0.139)**	**(0.426)**	**(0.644)**	**(0.204)**	**(0.188)**	**(0.489)**	**(1.166)**	**(0.569)**	**(0.153)**
Social group characteristics	-0.00868	0.0483	0.0591	0.0772	-0.0456	0.0207	0.0490	0.0533	-0.965*	-0.253	-0.0331	0.00976	-1.731*	-0.553	0.332	-0.0688
	(0.454)	(0.115)	(0.128)	(0.151)	(0.415)	(0.0866)	(0.0815)	(0.108)	(0.565)	(0.211)	(0.133)	(0.153)	(1.016)	(0.599)	(0.933)	(0.264)
Education years	-0.0108	-0.0269*	-0.0164	-0.0211	-0.00155	-0.0197	-0.00914	-0.00900	0.0383	0.0118	-0.00756	-0.00577	0.118*	0.0562	0.0299	0.0247
	(0.0404)	(0.0151)	(0.0183)	(0.0230)	(0.0372)	(0.0126)	(0.0131)	(0.0159)	(0.0316)	(0.0154)	(0.0130)	(0.0163)	(0.0715)	(0.0510)	(0.0560)	(0.0313)
Mills ratio	105.5***	64.33***	104.5***	121.4***	59.99***	48.42***	58.62***	62.44***								
	(19.67)	(16.38)	(18.65)	(23.13)	(14.11)	(13.29)	(12.70)	(14.70)								
HIV prevention, abstain	0.223***	0.146**	0.225***	0.260***	0.127**	0.102**	0.131***	0.140**	-0.0376	-0.0292	-0.0101	-0.0333	-0.396	-0.423	-0.111	-0.142
	(0.0708)	(0.0580)	(0.0691)	(0.0856)	(0.0518)	(0.0484)	(0.0502)	(0.0587)	(0.0702)	(0.0443)	(0.0499)	(0.0615)	(0.271)	(0.321)	(0.233)	(0.213)
HIV prevention, condom									-0.191	-0.171*	-0.0624	-0.0597	0.0851	0.101	0.131	0.0933
									(0.139)	(0.0921)	(0.0860)	(0.110)	(0.157)	(0.156)	(0.121)	(0.116)
HIV prevention, one partner	-0.148	-0.111	-0.139	-0.145	-0.0802	-0.0729	-0.0719	-0.0678	-0.178	-0.0109	0.00804	0.0249	-0.354	0.0875	0.0173	-0.0232
	(0.120)	(0.0893)	(0.109)	(0.134)	(0.0869)	(0.0743)	(0.0782)	(0.0909)	(0.159)	(0.0733)	(0.0829)	(0.0989)	(0.291)	(0.269)	(0.227)	(0.189)
Knows someone died AIDS	0.121	0.0498	0.123	0.161*	0.0968*	0.0241	0.0978*	0.124*	-0.0954	-0.0714	0.0218	0.0458	-0.239	-0.162	0.0436	-0.0155
	(0.0761)	(0.0632)	(0.0757)	(0.0959)	(0.0562)	(0.0530)	(0.0562)	(0.0672)	(0.103)	(0.0586)	(0.0553)	(0.0694)	(0.242)	(0.191)	(0.146)	(0.129)
Majority population, Coloured	0.0114	-0.0936	-0.0458	-0.0521	0.000233	-0.0532	-0.0499	-0.0378	1.476*	0.275	0.159	0.0903	2.391	0.732	-0.799	-0.0685
	(0.336)	(0.131)	(0.161)	(0.201)	(0.291)	(0.112)	(0.135)	(0.133)	(0.823)	(0.310)	(0.198)	(0.223)	(2.008)	(0.837)	(1.120)	(0.390)
Majority population, White	0.137	-0.0784	-0.0198	-0.0564	0.142	-0.0356	-0.0423	-0.0314	3.185*	0.763	0.316	0.164	5.272	1.882	-1.355	0.0266
	(0.978)	(0.270)	(0.348)	(0.375)	(0.897)	(0.224)	(0.249)	(0.235)	(1.769)	(0.638)	(0.426)	(0.481)	(3.722)	(1.793)	(2.640)	(0.816)
Constant	-16.12***	-9.956***	-15.99***	-18.54***	-8.556***	-6.758***	-8.470***	-9.062***	-3.546*	-0.184	-0.686	-0.192	-8.352	-7.243***	-1.158	-2.750**
	(2.773)	(2.430)	(2.729)	(3.451)	(1.942)	(1.932)	(1.863)	(2.194)	(2.016)	(0.817)	(0.612)	(0.655)	(8.983)	(2.688)	(1.601)	(1.304)
Observations	1,225	549	1,225	1,030	1,225	549	1,225	1,030	873	458	873	691	873	458	873	691
Anderson canon correlation statistics (p-value)					0.0089	0.0000	0.0000	0.0000	0.0362	0.0068	0.0000	0.0000				
Sargan-Hansen statistic (p-value)					0.7711	0.7006			0.1970	0.6308						
Amemiya-Lee Newey (p-value)	0.7813	0.3582												0.8709		

- Standard errors in parentheses
- *** p<0.01, ** p<0.05, * p<0.1 indicates significance at 1%, 5% and 10%
- Definition of instruments follows from table 2.2 in section 2.7.1
- Majority population reference is Black

Table 2.F2: Social Norm and Social Pressure Results among the Female Population

	Both Sexually Active and Not Sexually Active Population								Sexually Active Population Only							
	IVTOBIT CTSM 2nd stage				GMM CTSM 2nd stage				GMM				IVPROBIT			
	Nonlinear in means model No. of sex partners incl. not sexually active				Linear in means model No. of sex partners incl. not sexually active				Linear in means model No. of sex partners incl. not sexually active				Nonlinear in means model No. of sex partners incl. not sexually active			
	(1)[1]	(2)[2]	(3)[3]	(4)[4]	(5)[1]	(6)[2]	(7)[3]	(8)[4]	(9)[1]	(10)[2]	(11)[3]	(12)[4]	(13)[1]	(14)[2]	(15)[3]	(16)[4]
Social norm	**2.588*** **	**2.666*** **	**2.206*** **	**2.069*** **	**0.956*** **	**1.149*** **	**0.897*** **	**1.006***	28.57***	0.401	33.51***	139.3**	0.391	9.304**	-1.203	-2.349
	(0.612)	**(0.704)**	**(0.494)**	**(0.333)**	**(0.259)**	**(0.317)**	**(0.343)**	**(0.543)**	(8.720)	(0.455)	(12.76)	(65.59)	(1.410)	(4.430)	(1.004)	(1.557)
Social norm*social pressure	**0.925*** **	**1.663*** **	**0.720*** **	**0.793*** **	**0.0377**	**0.0109**	**0.0670**	**0.0656**	0.612	2.521***	0.610	-4.722	0.500	3.015	4.461	6.472
	(0.182)	**(0.546)**	**(0.196)**	**(0.141)**	**(0.129)**	**(0.324)**	**(0.107)**	**(0.139)**	(0.549)	(0.780)	(0.846)	(3.435)	(3.889)	(6.058)	(2.943)	(4.166)
Social group characteristics	0.120	0.117	0.0188	0.00967	0.0411	0.125	0.0286	0.0533	0.271***	0.0419	-0.192***	-0.164***	-0.850***	0.617	0.645*	-0.0444
	(0.142)	(0.103)	(0.0667)	(0.0509)	(0.0781)	(0.0886)	(0.0713)	(0.108)	(0.0832)	(0.0315)	(0.0644)	(0.0621)	(0.327)	(0.462)	(0.340)	(0.251)
Education years	-0.0155	-0.0233**	-0.00448	-0.00771	-0.00607	-0.0227**	-0.00501	-0.00900	-0.0341***	-0.00928	0.00395	-0.0527	0.0916*	-0.0985	-0.0174	0.0180
	(0.0149)	(0.0116)	(0.0143)	(0.0110)	(0.00943)	(0.00959)	(0.00819)	(0.0159)	(0.00989)	(0.00642)	(0.0109)	(0.0324)	(0.0485)	(0.0659)	(0.0420)	(0.0477)
Mills ratio	44.15***	41.83***	52.92***	43.29***	26.72***	31.42***	26.59***	62.44***								
	(10.15)	(9.361)	(13.08)	(10.07)	(6.921)	(7.247)	(6.936)	(14.70)								
HIV prevention, condom									0.0288	0.0247	-0.0159	-0.0121	-0.0705	0.0599	0.0289	-0.138
									(0.0251)	(0.0195)	(0.0229)	(0.0437)	(0.147)	(0.214)	(0.141)	(0.161)
HIV prevention, abstain	0.297***	0.188***	0.349***	0.291***	0.203***	0.159***	0.203***	0.140**	-0.0286	-0.0303	-0.0230	0.00475	-0.425*	-0.417	-0.431*	-0.507*
	(0.0439)	(0.0404)	(0.0558)	(0.0430)	(0.0310)	(0.0322)	(0.0309)	(0.0587)	(0.0339)	(0.0271)	(0.0311)	(0.0651)	(0.253)	(0.378)	(0.247)	(0.288)
HIV prevention, one partner	-0.139**	-0.129**	-0.169*	-0.154**	-0.0851*	-0.0892*	-0.0868*	-0.0678	0.0892**	0.00534	0.00425	0.0822	0.104	0.407	0.344	0.145
	(0.0701)	(0.0630)	(0.0871)	(0.0677)	(0.0487)	(0.0502)	(0.0485)	(0.0909)	(0.0417)	(0.0324)	(0.0397)	(0.0829)	(0.278)	(0.373)	(0.273)	(0.288)
Knows someone died AIDS	0.125***	0.0871**	0.148**	0.115**	0.0852**	0.0732**	0.0840**	0.124*	0.0537**	-0.00378	-0.00987	0.0433	-0.0720	0.180	0.142	0.119
	(0.0473)	(0.0413)	(0.0577)	(0.0449)	(0.0334)	(0.0334)	(0.0326)	(0.0672)	(0.0264)	(0.0190)	(0.0233)	(0.0467)	(0.143)	(0.196)	(0.143)	(0.152)
Majority population, Coloured	-0.0794	-0.221**	-0.0754	-0.0471	0.0168	-0.0727	0.0118	-0.0378	-0.173	-0.0355	0.393***	0.815***	0.493	-0.264	-1.176***	-0.316
	(0.0987)	(0.102)	(0.116)	(0.0802)	(0.0679)	(0.0886)	(0.0755)	(0.133)	(0.110)	(0.0680)	(0.0616)	(0.298)	(0.446)	(0.899)	(0.416)	(0.406)
Majority population, White	-0.276	-0.356	-0.0392	-0.0207	-0.0350	-0.224	-0.0154	-0.0314	-0.807***	-0.165*	0.486***	0.648**	2.307**	-2.384	-2.183**	-0.323
	(0.358)	(0.257)	(0.198)	(0.156)	(0.186)	(0.195)	(0.150)	(0.235)	(0.233)	(0.0910)	(0.159)	(0.307)	(0.908)	(1.503)	(1.080)	(0.631)
Constant	-6.915***	-6.456***	-8.176***	-6.675***	-3.854***	-4.382***	-3.812***	-9.062***	-0.408	0.0442	-1.433***	-5.922**	-3.500***	-10.16**	-0.573	-0.866
	(1.502)	(1.425)	(1.924)	(1.483)	(1.025)	(1.088)	(1.060)	(2.194)	(0.409)	(0.428)	(0.453)	(2.847)	(0.889)	(4.350)	(0.767)	(0.671)
Observations	1,544	707	1,272	1,544	1,544	707	1,544	1,030	1,090	559	1,090	834	1,090	559	1,090	834
Anderson canon correlation statistics (p-value)					0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000				
Sargan-Hansen statistic (p-value					0.6852	0.8656			0.0112	0.0000						
Amemiya-Lee Newey (p-value)	0.3692	0.5685												0.9320		

• Standard errors in parentheses
• *** p<0.01, ** p<0.05, * p<0.1 indicates significance at 1%, 5% and 10%
• Definition of instruments follows from table 2.2 in section 2.7.1
• Majority population reference is Black

Appendix 2.H: Social Norm and Social Pressure using Placebo groups

Table 2.H1: Placebo Social Groups

	Placebo Social Group															
	Both Sexually Active and Not Sexually Active Population								Sexually Active Population Only							
	IVTOBIT CTSM 2nd stage				GMM CTSM 2nd stage				IVPROBIT				GMM			
	Nonlinear in means model No. of sex partners incl. not sexually active				Linear in means model No. of sex partners incl. not sexually active				Nonlinear model Multiple Partnership (0/1)				Linear in means model No. of sex partners			
	Adjacent group IV	'Naïve' Adjacent group IV	Trend setter IV	'Naïve' Trend setter IV	Adjacent group IV	'Naïve' Adjacent group IV	Trend setter IV	'Naïve' Trend setter IV	Adjacent group IV	'Naïve' Adjacent group IV	Trend setter IV	'Naïve' Trend setter IV	Adjacent group IV	'Naïve' Adjacent group IV	Trend setter IV	'Naïve' Trend setter IV
	(1)[1]	(2)[2]	(3)[3]	(4)[4]	(5)[1]	(6)[2]	(7)[3]	(8)[4]	(9)[1]	(10)[2]	(11)[3]	(12)[4]	(13)[1]	(14)[2]	(15)[3]	(16)[4]
Social norm	**1.766**	**1.818**	**1.407**	**0.861**	**3.747**	**0.596**	**2.665**	**1.793**	**-0.0407**	**0.0698**	**-0.0902**	**-0.0475**	**12.64**	**-2.757**	**8.942**	**-4.003**
	(3.972)	**(12.20)**	**(1.809)**	**(0.892)**	**(5.623)**	**(16.79)**	**(2.588)**	**(1.238)**	**(0.215)**	**(0.114)**	**(0.375)**	**(0.343)**	**(12.30)**	**(6.016)**	**(16.93)**	**(17.51)**
Social norm*social pressure	**-0.336**	**-0.700**	**-2.381**	**-0.0596**	**-1.602**	**0.105**	**-4.466**	**-0.830**	**0.0355**	**-0.0340**	**0.0898**	**0.0476**	**-3.901**	**5.393**	**-3.808**	**7.129**
	(3.224)	**(8.990)**	**(4.674)**	**(0.810)**	**(4.590)**	**(12.37)**	**(6.760)**	**(1.115)**	**(0.122)**	**(0.101)**	**(0.279)**	**(0.245)**	**(6.596)**	**(5.366)**	**(12.58)**	**(12.41)**
Social group characteristics	0.235	0.226	0.238	0.139	0.436	0.0703	0.371	0.224	0.00624	0.00245	0.00420	0.0108	-0.322	0.0552	-0.153	-0.0835
	(0.521)	(1.374)	(0.274)	(0.144)	(0.738)	(1.891)	(0.394)	(0.198)	(0.00863)	(0.00283)	(0.00888)	(0.0261)	(0.487)	(0.151)	(0.419)	(1.354)
Education years	-0.0248*	-0.0339	-0.0297**	-0.0332***	-0.0199	-0.0295	-0.0276	-0.0324***	-0.00100	-0.00127***	-0.000798	-0.00170	-0.0458	-0.0863***	-0.0492	-0.0812
	(0.0139)	(0.0219)	(0.0126)	(0.00877)	(0.0189)	(0.0301)	(0.0179)	(0.0119)	(0.000857)	(0.000469)	(0.000761)	(0.00140)	(0.0453)	(0.0247)	(0.0376)	(0.0730)
Mills ratio	31.97**	31.25	32.91***	32.79***	58.68***	64.79**	63.70***	63.02***								
	(14.41)	(19.89)	(10.98)	(7.062)	(20.73)	(27.49)	(16.14)	(10.13)								
HIV prevention, abstain	0.194***	0.123	0.152***	0.120***	0.350***	0.238**	0.285***	0.248***	0.00574	0.000399	-0.000975	0.00111	0.312	-0.0387	0.0844	-0.0536
	(0.0587)	(0.0822)	(0.0438)	(0.0340)	(0.0807)	(0.113)	(0.0617)	(0.0464)	(0.00389)	(0.00187)	(0.00283)	(0.00297)	(0.214)	(0.0989)	(0.141)	(0.155)
HIV prevention, one partner	-0.0248	-0.0399	-0.0194	-0.0419	-0.145	-0.122	-0.118	-0.116*	0.00511	0.000387	0.00443	0.00112	0.0684	-0.0556	0.0892	-0.0682
	(0.0914)	(0.0653)	(0.0683)	(0.0466)	(0.126)	(0.0895)	(0.0968)	(0.0639)	(0.00590)	(0.00296)	(0.00444)	(0.00384)	(0.332)	(0.158)	(0.236)	(0.202)
Knows someone died of AIDS	0.147**	0.0691	0.124***	0.0748**	0.192**	0.127	0.178***	0.121***	0.00992***	0.000723	0.00534*	0.000781	0.210	-0.0212	0.135	-0.0245
	(0.0628)	(0.0731)	(0.0478)	(0.0321)	(0.0845)	(0.1000)	(0.0663)	(0.0431)	(0.00371)	(0.00191)	(0.00291)	(0.00195)	(0.192)	(0.100)	(0.143)	(0.102)
Majority population, Coloured	-0.240***	-0.275***	-0.281***	-0.282***	-0.436***	-0.435***	-0.481***	-0.443***	-0.00425	-0.00881***	-0.00515	-0.0108	-0.177	-0.322***	-0.185	-0.280
	(0.0644)	(0.0339)	(0.0596)	(0.0540)	(0.0875)	(0.0460)	(0.0846)	(0.0735)	(0.00402)	(0.00219)	(0.00337)	(0.00683)	(0.219)	(0.119)	(0.172)	(0.356)
Majority population, White	-0.119	-0.219	-0.250*	-0.222**	-0.344**	-0.382	-0.491***	-0.417***	0.00907	-0.00209	-0.000204	-0.00725	0.0866	-0.0713	0.0997	0.0291
	(0.105)	(0.192)	(0.130)	(0.108)	(0.147)	(0.265)	(0.188)	(0.148)	(0.00701)	(0.00367)	(0.00609)	(0.0168)	(0.353)	(0.191)	(0.288)	(0.868)
HIV prevention, condom									-0.00304	-0.00730***	-0.00748*	-0.00694**	0.0697	-0.236	-0.0646	-0.248
									(0.00600)	(0.00283)	(0.00411)	(0.00309)	(0.311)	(0.166)	(0.211)	(0.176)
Constant	-5.052*	-4.633	-3.967**	-4.351***	-10.17***	-9.009*	-8.730***	-9.306***	0.0739	0.0374	0.0962	0.0957	-8.584	-0.0869	-5.898	0.190
	(2.665)	(3.421)	(1.561)	(1.066)	(3.773)	(4.720)	(2.293)	(1.525)	(0.113)	(0.0490)	(0.172)	(0.171)	(6.497)	(2.575)	(7.763)	(8.754)
Observations	824	2,739	1,408	2,739	824	2,739	1,408	2,739	613	2,063	1,030	2,063	613	2,063	1,030	2,063
R-squared	0.091	0.082	0.051	0.089					0.043	0.028	0.025	0.005				
Anderson statistics (p-value)	0.0000	0.5336	0.0159	0.0000					0.0002	0.0000	0.0003	0.0320				
Sargan-Hansen (p-value)	0.8797	0.7224							0.8048	0.9338						
Amemiya-Lee Newey (p-value)					0.6838	0.9098							0.7289	0.6281		

- Robust Standard errors in parentheses
- Majority population reference is Black
- Instrumented for social norm and social norm*social pressure
- *** p<0.01, ** p<0.05, * p<0.1 indicates significance at 1%, 5% and 10%

[1] Excluded instruments: social norm of left social group, social norm of right social group, social norm*social pressure of right social group and social norm*social pressure of left social group.

[2] Excluded instruments: social norm of older social group, social norm*social pressure of older social group

Chapter 3
'Know Your Epidemic': The Effects of Expected Health and Contextual Effects of Health Uncertainty on Risky Sex

Abstract

This paper measures the effects of *expected income*, *expected health* and the contextual effects of *health uncertainty* on sexual behaviours associated with the risk of HIV infection using six cross sectional Demographic and Health Surveys. We argue that the concept of social interactions as critical guiding individual behaviour is applicable to our analysis of the determinants of sexual behaviour. We apply an alternative estimation strategy from the current existing literature to overcome the identification problem arising from selection bias, unobservables and simultaneity, which is often ignored. As expected the results show a significant negative relationship between *expected health* and *risky sex* and a significant positive relationship between *health uncertainty* and *risky sex*. The results show exposure to HIV information and education as important determinants of risky sexual behaviour, although to a lesser extent. These results point to the fact that limited economic prospects and uncertain future health status are even more important factors in engaging in risky sexual behaviour than the widely held view that lack of information and education contributes to deadly diseases such as HIV infection.

Keywords: Contextual factors, Risk Taking, HIV/AIDS, Sexual behaviour, Expected health, Health Uncertainty

"While we cannot hold that uncertainty is something specific to Africa alone, the scale and impact of various kinds of catastrophes on the continent has been disproportionately huge….on one hand recurrent phenomena across the continent such as drought, famine and epidemics of various kinds and on the other predictable consequences of poor government and civil wars…resulted in a weakening of social fabric and traditional support networks, making life uncertain for the majority of Africans" (Haram and Yamba, 2009, p. 11). "*Interviewer: Why is that, men think about pleasure first before thinking about their health? Informant: The dangers and risks of the job we are doing are such that no one can afford to be motivated with life-so the only thing that motives us is pleasure*" (Interview with South African mine worker, Campbell 1997, p. 227 cited in Mannberg 2010).

3.1 Introduction and Structure of the Study

A number of studies point to the importance of the *local social and physical environment* in shaping sexual behaviour in general (Adimora and Schoenbach, 2005; Peterman *et al.*, 2005) and HIV/AIDS in particular (Gabrysh *et al.* 2008; Nattrass 2009). These observations are also complemented by policy measures in which HIV/AIDS prevention initiatives are called on to match underlying contextual factors, particularly known as the United Nations Programme on HIV and AIDS (UNAIDS) '*know the local epidemic*'[52] (UNAIDS, 2007; 2008). Approaches focusing on individual behaviour are unlikely to produce improvements in sexual health. This is particularly true in poor countries where personal choice is more constrained than in wealthier countries. Numerous calls have been made to pay greater heed to the *local social and physical environment* within which sexual intercourse occurs. Most economic research on behaviour is however mainly methodologically individualistic (Logan *et al.* 2002; Coates *et al.* 2008; Benefo 2010; Durlauf and Ioannides 2010), as famously cited, 'economics is all about how people make decisions, sociology is about how they have no decisions to make' (Duesenberry 1960 cited in Morgan and Niraula 1995: p. 542).

In this vein, this study examines individual risky sexual behavioural responses to *expected income*, *expected health* and the contextual effects of *health uncertainty*. The study is closely aligned to the empirical studies of Oster (2009, 2012) and Bezabih *et al.* (2010). The novel feature of this study however is in the application of an alternative estimation strategy to overcome the identification problem arising from selection bias, unobservables and simultaneity. The analysis is based on data from six cross sectional Demographic and Health Surveys (DHS), three from East Africa (Tanzania, Kenya and Uganda) and the others from Southern Africa (Swaziland, Lesotho and Zimbabwe) to establish whether the findings are similar by region. Taking a different approach from most of the sexual behaviour literature our outcome variable is a *risky sex* index. The *risky sex* index is an aggregation of various sexual behaviours namely relationship with last partner, condom use with last partner, number of sex partners and age of sexual debut using confirmatory factor analysis (Bezabih *et al.* 2010). The rationale for this approach is that individuals who perceive themselves as practicing more *risky sexual behaviour*,

[52] This means knowing where the epidemic exists (regional realities and populations most affected) and also what are its main drivers and where it is moving. In-depth understanding of the social and behavioural context is central to knowing your epidemic (UNAIDS 2007: p vi).

are more likely to adopt other *protective sexual behaviour* to counteract their overall riskiness. As such it becomes paramount to aggregate behaviours and establish overall sexual risk.

In addition, individuals who are not sexually active should also be included in the analysis as the decision to abstain also constitutes a *protective sexual behaviour* (Bezabih *et al.* 2010). However, their inclusion in our model implies the presence of sample selection bias, because the decision to become sexually active is non-random. To overcome this we use a censored two stage model (Greene 2002; Wooldridge 2003; Bushway *et al.* 2007; Bezabih *et al.* 2010). In the first stage a selectivity model for sample selection is estimated and the inverse Mills ratios used as an additional regressor in the second stage model.

The main regressors in this model are *expected income*, *expected health* and the contextual *health uncertainty*. The study uses two different proxies for *expected income*: household income and wealth indices. To obtain household income the regression based approach is adopted using the Living Standards Measure Surveys (LSMS) following the work of Grim *et al.* (2006) and Oster (2009), while the wealth indices are calculated following the work of Rusten and Johnson (2004) and Grim *et al.* (2009). The departure from their work is based on the fact that confirmatory factor analysis is used to derive the indices, instead of principal component analysis since principal component analysis produces biased estimates in the case of categorical data (Kolenikov 2009a; Bezabih *et al.* 2010).

The second main regressor is *expected health*. Our proxy measure for *expected health* is life expectancy following Bezabih *et al.* (2010). However, because DHS datasets do not contain any information on life expectancy the INDEPTH model life tables are used. In this case Oster's (2009) work is extended by calculating three different life expectancies which vary by mortality rates and the life table pattern. The first life expectancy is named *HIV-life expectancy 2* and this is based on the *HIV mortalities* (using 2000s DHS data) and *pattern 2* of the INDPETH model life tables, where *pattern 2* captures the East and Southern African regions. The *HIV-life expectancy 2* is unlikely to be *non-HIV life expectancy* since *HIV mortalities* and *pattern 2* of East and Southern African regions with high HIV rates were used. As such instruments are used in the final analysis when using this proxy.

The second life expectancy is called *HIV-life expectancy 1* and this is based on the *HIV mortalities* with the only difference being that *pattern 1* of the INDPETH model life tables which is based on West and Central African regions with low HIV rates is used. The *HIV-life expectancy 1* is also unlikely to be *non-HIV life expectancy* as the current mortality rates that are likely to be influenced by HIV/AIDS are utilised. However the reverse causality is likely to be '*lower*' compared to the *HIV-life expectancy 2* because *pattern 2* is employed. The third and final life expectancy is the *pre-HIV-life expectancy 1*. This is calculated using the mortality rates of DHS data from the 1980s, what is called the *pre-HIV mortalities* and *pattern 1* based on the West and Central African regions with low HIV. As such this life expectancy is likely to be *non-HIV life expectancy* and thus does not require instruments in the final analysis.

The last main regressor captures the contextual effects of *health uncertainty*. Three proxies of *health uncertainty* are tested which are captured by the variance of life expectancy following Bezabih *et al.* (2010) and derived from the three measures of life expectancy described above. The departure from Bezabih *et al.* (2010) is that while they calculated this at the community level the current proxy is instead calculates at a social group level. This is because place of residence brings with it geographically related unobservables (Blume *et al.* 2010) resulting in the identification problem. Following the literature by Manski (1993, 2000, 2003); Moffit (2001); Brock and Durlauf (2004); Grodner and Kniesner (2005); Brock and Durlauf (2007) and Blume *et al.* (2010) the *social space* is represented by a social group and not represented by a community. The social groups consist of individuals with the same gender, age and income and this eliminates any trace of geographically related unobservables associated with place of residence. As such the study looks at the *social environment* and not the *physical environment*. The instruments used to circumvent the identification problem are based on the works of Grodner and Kniesner (2005); Etile (2007) and Grodner *et al.* (2011). The study deploys two types of instruments. The first is the *trend setter IV* which extends the aforementioned works and the second instrument used is the *adjacent social group age bracket IV* is borrowed from the wok of Etile (2007).

The *health uncertainty* proxy was tested by comparing the performance of the proxy with other indicators of *health uncertainty* such as World Health Organisation (WHO) mortality rates (Montgomery and Casterline 1996) of each country and we found a similar pattern. The social group composition was tested using the '*placebo treatment*' which creates social group falsification tests following Lavy and Schlosser (2007). The results from the selected DHS data generally reveal a significant negative relationship between *expected health* and *risky sex* and a significant positive relationship between *health uncertainty* and *risky sex*. Our findings do indeed support Mannberg's (2010) theoretical framework.

The rest of this paper is organised as follows: section 3.2 gives a brief background and provides details on current existing empirical studies in sub-Saharan Africa and the contribution of the current study. This is followed by section 3.3 which outlines the theoretical model. Section 3.4 discusses the estimation challenges and the approach used to circumvent them. Section 3.5 shows the results, and a conclusion is provided in section 3.6.

3.2 Background, Motivation and Contribution of the Study

'*This place is killing me*', a captivating title of a relatively recent research by Peterman *et al.* (2005) that points to the fact that the *local social and physical environment* may have a role to play in influencing the spread of HIV/AIDS. Another study by Adimora and Schoenbach (2005) establishes that social context is significant in determining sexual behaviours. Similarly, Gabrysh *et al.* (2008) found that contextual factors influence HIV risk in Zambia. In response to such evidence the UNAIDS initiated a project '*Know your epidemic, know your response'*, calling for countries to match their HIV/AIDS prevention initiatives to the underlying contextual factors that may be driving their local epidemic (UNAIDS 2007; 2008). This approach is not surprising because after more than thirty years of HIV/AIDS initiatives, the world still remains challenged by the epidemic, with sub-Saharan Africa remaining the most affected region.

Nattrass (2009) established that contextual factors play a significant role in the African epidemic, concluding that results do offer support for the validity of UNAIDS '*know the local epidemic' challenge.* A comprehensive study by Wellings *et al.* (2006) reviewed global sexual behaviour patterns and found that social context does play a role in HIV/AIDS transmission[53]. The same view is shared by amongst others Gillespie *et al.* (2007); Wilson and Halperin (2008); Mah and Halperin (2010a) and Mannberg (2010). However, there is a paucity of quantitative empirical literature in this area (Macintyre *et al.* 2002; Morisky *et al.* 2006; Benefo 2008; Coates *et al.* 2008; Painter *et al.* 2008; Burgard and Lee-Rife 2009). To fill this gap, this study examines individual sexual behavioural responses to *expected income*, *expected health* and the contextual effects of *health uncertainty*[54] using Mannberg's (2010) theoretical framework.

The aforementioned arguments have long been put forth by sociologists who have often argued that an individual's *health and health related behaviour* are affected not only by their individual characteristics but also by the *local social and physical environment* (Macintyre *et al.* 2002; Diez-Roux and Aiello 2005; Gabrysch *et al.* 2008; Dembo 2009; Durlauf *et al.* 2010; Diez-Roux 2010), resulting in individuals from the same region having correlated health outcomes (Macintyre *et al.* 2002; Gabrysch *et al.* 2008; Dembo 2009). Therefore it is argued that, in order to understand individual health outcomes, it is useful to analyse the *local social and physical environment* (UNAIDS 1999; Diez-Roux 1998; Logan *et al.* 2002; Jipguep *et al.* 2004; Adimora and Schoenbach 2005; Painter *et al.* 2008; DeRose 2009). Research dating back to the early twentieth century has demonstrated the importance of the *local social and physical environment* in determining an individual's health outcomes. For example, Goldberger in 1916 found the risk of pellagra to be related to the availability of fresh fruits and vegetables at the village

[53] ...approaches focusing on individual behaviour are unlikely to produce improvements in sexual health. They are especially inappropriate to a poor country... In wealthier countries, personal choice is greater than in poor countries....Numerous calls have been made to pay greater heed to the social context within which sexual intercourse occurs (Wellings *et al.* 2006: p.1716).

[54] *Health uncertainty* has been highlighted as an influence on various *health and health related behaviours*, such as higher birth rates in younger women from disadvantaged populations (Geronimus 1996, Geronimus *et al.* 2006). Another example is parental uncertainty regarding HIV/AIDS status which has been associated with a decrease in children's school enrolment (Grant 2008). Correspondingly, in relation to sexual behaviour, the high prevalence of uncertainty of future health outcomes in sub-Saharan Africa implies that the perceived future benefits of abstaining from risky sexual behaviour are outweighed by the perceived current benefits of pleasure from risky sexual pleasure (Mannberg 2010).

level, independent of individual incomes, illustrating a link between individual and village level factors (Macintyre *et al.* 2002).

It is only relatively recently that their incorporation into economic literature gained favour (Macintyre *et al.* 2002; Obermeyer 2005; Morisky *et al.* 2006; Benefo 2008; Coates *et al.* 2008; Painter *et al.* 2008; Burgard and Lee-Rife 2009). To date the general consensus is that, to understand individual *health and health related behaviour*, it is imperative to go beyond the study of individual characteristics (UNAIDS 1999; Diez-Roux 1998; Jipguep *et al.* 2004; Adimora and Schoenbach 2005; Painter *et al.* 2008; DeRose 2009; Diez-Roux 2010). However, as earlier mentioned empirical literature still lags behind (Macintyre *et al.* 2002; Morisky *et al.* 2006; Coates *et al.* 2008; Painter *et al.* 2008; Burgard and Lee-Rife 2009).

The most recent studies that are closely related to the current research theme of health and sexual behaviour are that of Oster (2009, 2012); Mannberg (2010) and Bezabih *et al.*, (2010). Oster's (2009) theory suggests that responsiveness of sexual behaviour to HIV/AIDS will be lower for individuals with lower income and lower life expectancy. To test this theory Demographic Household Surveys (DHS) data for the periods 1998 to 2003 were utilised which include Benin, Burkina Faso, Ethiopia, Ghana, Kenya, Malawi, Mali, Namibia and Zimbabwe. The theory is based on the following proposition: $\frac{d\sigma_1}{dh} < 0$, $\frac{d\sigma_1/dh}{dp} < 0$ and $\frac{d\sigma_1/dh}{dy} < 0$, where σ_1 is the sexual partners, h is the HIV prevalence, p is the life expectancy and y is income. On the other hand, Oster (2012) includes 14 DHS datasets from the following countries: Burkina Faso, Cameroon, Ethiopia, Ghana, Guinea, Kenya, Lesotho, Malawi, Mali, Niger, Senegal, Swaziland, Zambia and Zimbabwe. In this study the theory measures responsiveness of sexual behaviour to HIV prevalence, life expectancy and HIV/AIDS knowledge. In this study the theoretical framework is based on the following proposition replacing income from the previous study with HIV/AIDS knowledge: $\frac{d\sigma_1}{dh} < 0$, $\frac{d\sigma_1/dh}{dp} < 0$ and $\frac{d\sigma_1/dh}{d\gamma} < 0$, where γ is the HIV/AIDS knowledge.

Mannberg (2010) contends that it is not only *expected health* and *expected income* that influence sexual behaviour but also *health uncertainty* and extends Oster's (2009; 2012) theory by including the contextual effects of *health uncertainty*. The theory postulates that in poor countries, the future is more likely to be uncertain. Individuals in these countries are therefore less likely to invest in their health and instead live for the present. Their sexual behaviours are therefore more likely to be risky. The study is based on the following proposition: $\frac{dx_{us}}{d\gamma} > 0$, $\frac{dx_{us}}{d\tilde{h}_2} < 0$ and $\frac{dx_{us}}{dm_2} \lessgtr 0$, where x_{us} is unsafe sex, γ is uncertainty of future health, $\tilde{h}_2$ is future health/life expectancy and m_2 is income. Lastly, Bezabih *et al.* (2010) applied Mannberg's (2010) theory using data from the Cape Area Panel Study (CAPS) of young adults in the Metropolitan area of Cape Town, South Africa. They found that indeed the greater the uncertainty with regard to future health the more risky the sexual behaviour, although their findings regarding expected health and sexual risk taking were not significant.

3.3 Theoretical Framework

This study adopts the theoretical framework of Mannberg (2010), to include contextual effects of *health uncertainty* in previous models of health and sexual behaviour.

The theory is based on a utility maximising individual who trades off future health for pleasure derived from sexual activities in the present. The theory considers a two period utility model where an individual consumes goods (c), health (h) and sex (x). The preferences in period *t* are captured by the following lifetime utility function: $\mathrm{U}\,(\mathbf{c}, \mathbf{h}, \mathbf{x_{us}}, \mathbf{x_s})$, where $\mathbf{c} = (c_1, c_2)$ is consumption of goods in period one and two: $\mathbf{h} = (h_1, h_2)$ is consumption of health: $\mathbf{x_s} = (x_{s,1}, x_{s,2})$ is consumption of *safe sex* and $\mathbf{x_{us}} = (x_{us,1}, x_{us,2})$ is consumption of *risky sex*. The utility function is assumed to be twice continuously differentiable, increasing in each argument and strictly concave. The consumption of *risky sex* yields more utility than the consumption of *safe sex*. Furthermore the consumption of *safe sex* and *risky sex* are assumed to be additively separable. In addition to this, the consumption of *risky sex* in period two does not have any implication on health since the individual only lives for two periods (Mannberg 2010).

The theory further assumes that the individual faces two types of risks. The first being the risk of contracting HIV and the second is the exogenous risk associated with a change in health due to a stochastic health shock. The probability of HIV infection through unprotected sex is $\Pr(\mathrm{HIV})$, while $1 - \Pr(\mathrm{HIV}) = \varphi$ is the probability of not being infected and hence staying HIV negative. However, if infected an individual is assumed to die in period one and therefore does not reach period two. On the other hand, the changes in health brought about by the stochastic health shock are reflected by high or low health represented by probability ρ and $(1 - \rho)$ respectively. Additionally, the difference between high and low health, which Mannberg (2010) refers to as the '*spread*' is given by $\gamma = (h_{high} - h_{low})$. Finally, if the individual lives up to period two then the expected value of future health in period two, $\tilde{h}_2$ implies that $h_{high} = \tilde{h}_2 + (1 - \rho) \cdot \gamma$ and $h_{low} = \tilde{h}_2 - \rho \cdot \gamma$. This suggests that expected utility in period one can be expressed formally as follows:

$$E[U] = u(c_1) + g(x_{us}) + \alpha\, g(x_{s,1}) + \beta \cdot E\,[u(c_2, h_2)] \tag{3.1}$$

From equation (3.1) $0 < \alpha < 1$ is the scale factor of the utility of *safe sex* and $0 < \beta < 1$ is the exogenous discount parameter[55]. Subsequently, the expected utility in period two is depicted in

[55] According to Mannberg's (2010) theory living in an area with a high risk of dying from causes other than AIDS implies that there is no guarantee of returns on investing in *protective sexual behaviour*. The implication is that the presence of future health uncertainty may result in risk averse individuals acting short-sightedly and practicing *riskier sexual behaviours*. Mannberg (2010) acknowledges that practicing *riskier sexual behaviours* may be the results of hyperbolic discounting, that is, present bias preference. The theory however does not include this in an individuals' sexual behaviour decision. Furthermore, Mannberg (2010) argues that hyperbolic discounting is likely to be present in all cultures and that there is lack of evidence to show that individuals who are susceptible to HIV have a higher hyperbolic utility function. Hence the theory omits present bias preference from the analysis and concentrates on a comparison between behaviours in areas that have high health uncertainty and those that have low health uncertainty.

equation (3.2), while the inter-temporal budget constraints in period one and two are shown in equation (3.3) and (3.4) respectively:

$$E[U\,(c_2, h_2)] = \varphi_{x_{us}} \cdot \left[\rho \cdot u\left(c_2, h_{high}\right) + \;(1-\rho)\cdot u(c_2, h_{low})\right] \tag{3.2}$$

$$m_1 = c_1 + p\,x_s + q\,x_{us} + S \tag{3.3}$$

$$m_2 + (1+r)\;\cdot S = c_2 \tag{3.4}$$

In the equation above *p* and *q* are the price of protected sex and unprotected sex respectively. Therefore the first order conditions for *safe sex*, *risky sex* and savings take the following form as revealed in equation (3.5), (3.6) and (3.7) respectively:

$$\frac{\partial E[U_t]}{\partial x_s} = U_{x_s} = \;-p\cdot u_{c_1} + \;\alpha \cdot g_{x_s} = 0 \tag{3.5}$$

$$\frac{\partial E[U_t]}{\partial x_{us}} = U_{x_{us}} = \;-q\cdot u_{c_1} + g_{x_{us}} + \;\beta\ln(\varphi)\varphi^{x_{us}}\left[\rho\cdot U_{c_2}\left(c_2, h_{high}\right) + \;(1-\rho)\cdot U_{c_2}(c_2, h_{low})\right] = 0 \tag{3.6}$$

$$\frac{\partial E[U_t]}{\partial S} = U_S = \;-u_{c_1} + \beta\varphi^{x_{us}}\left[\rho\cdot U_{c_2}\left(c_2, h_{high}\right) + \;(1-\rho)\cdot U_{c_2}(c_2, h_{low})\right] \le 0 \tag{3.7}$$

Accordingly the model yields the following predictions: $\frac{\partial x_{us}}{\partial \gamma} > 0$, $\frac{\partial x_{us}}{\partial \tilde{h}_2} < 0$ and $\frac{\partial x_{us}}{\partial m_2} \gtreqless 0$. That is, an increase in uncertainty of future health reduces the incentives to abstain from *risky sex*. Secondly, an increase in the expected level of future health increases incentives to abstain from *risky sex*. Finally, an increase in future income has ambiguous effects on incentives to engage in risky sexual practices (Mannberg 2010). The theoretical framework implies the following econometric model that specifies the relationship between risky sex, expected income, expected health and health uncertainty.

$$y_i = \alpha + \;\beta_1.\tilde{h}_2 + \;\beta_2.\gamma_{health} + \;\beta_3.E[m_2] + \;\mathbf{X}'_i\,\boldsymbol{\beta} + u_i \tag{3.8}$$

Where y_i is *risky sex*, $\tilde{h}_2$ is *expected health*, γ_{health} is *health uncertainty*, $E[m_2]$ is *expected income*, $\mathbf{X}$ are our control variables, and u_i is the error term.

3.4 Empirical Framework

The following sub-section introduces the data. The second part of this section describes in more detail the methodology used to define proxies for *expected health*, *health uncertainty* and *expected income*. The final sub-section outlines the estimation and identification strategies.

3.4.1 Area of Study

The data employed in the analysis is the Demographic and Health Surveys[56] (DHS). The surveys focus on population, nutrition and health, with recent surveys collecting data on HIV/AIDS including knowledge of the disease and sexual relationships, and in some countries the survey incorporates HIV testing. In order to account for heterogeneity in HIV prevalence in the African continent (James *et al*, 2006; UNAIDS, 2007, 2008, 2009, 2010b), we include Tanzania, Kenya and Uganda in East Africa; and Swaziland, Lesotho, and Zimbabwe from Southern Africa. The focus is on the Southern African and East African regions given that the Southern African region has the highest HIV prevalence of the African regions (James *et al.* 2006; UNAIDS 2007, 2008, 2009, 2010b), in contrast in East Africa the prevalence is much lower. Table 3.A1 of Appendix 3.A depicts the characteristics of the DHS data.

3.4.2 Variable Specification

3.4.2.1 Risky Sex Index

The outcome variable in this study is the *risky sex index*. This index is derived by aggregating different sexual behaviours for each individual. This section outlines the motivation for this approach. We then proceed to identify the types of sexual behaviours included in the index. Thereafter the procedure used to aggregate the index is provided. We also give our rationale for including the population that is not sexually active.

While most sexual behaviour studies have in the past concentrated on measuring HIV risk using isolated individual behaviours, such as condom use, sexual debut, and multiple partners. However, this study adopts the view of Bezabih *et al.* (2010) that to gain a more accurate understanding, one needs to aggregate these individual behaviours. The main advantage of such an aggregation is that individual measures are unlikely to represent the full range of actions that reflect sexual behaviour that an individual takes. For example an analysis of condom use that divides the population according to non-users (risky behaviour) and users (safe behaviour) is unlikely to present the overall HIV riskiness or safeness of the behaviour of populations. This is because an individual who does not use a condom with a spouse might not have the same risk as another individual who does not use a condom with a commercial sex worker. Further to this, individuals who deem themselves as practicing more *risky*

[56] DHS are cross sectional surveys conducted in low and middle income countries in five year intervals. The surveys focus on population, nutrition and health, with recent surveys collecting data on HIV/AIDS including knowledge of the disease and sexual relationships, and in some countries the survey incorporates HIV testing. For details on the surveys, see http://www.measuredhs.com

sexual behaviour, are more likely to adopt other *protective sexual behaviour* to counteract their overall riskiness[57].

Notwithstanding, the aggregation of sexual behaviours is a relatively new phenomenon, and those who have attempted to do so have used ad hoc methods to allocate weights to the different types of sexual behaviour. Bezabih *et al.* (2010) recently constructed a *risky sex* index using LISREL confirmatory factor analysis (CFA) to assign weights to different variables. Hence following Bezabih *et al.* (2010) a *risky sex* index is constructed to capture the overall riskiness of an individual's sexual choices. The variables used to calculate this index include type of relationship with last partner, number of sex partners, condom use with last partner and age of sexual debut. The STATA CFA package[58] is used to construct our risky sex index because principal component analysis (PCA) will yield biased estimates because our sexual behaviour variables are categorical in nature (Kolenikov 2009a; Bezabih *et al.* 2010). Further details of the CFA procedure are provided in Appendix 3.C, while Table 3.C1 of the same Appendix shows the CFA output.

It is worth noting that rational individuals with many partners are likely to use condoms more consistently (Bezabih *et al.* 2010). Due to this correlation condom use will have a positive loading in the CFA analysis and might be wrongly interpreted as an increased risk in the construction of the index (Bezabih *et al.* 2010). Condom use is therefore not included in the CFA index[59] because it is likely to be correlated with the number of sex partners. Bezabih *et al.*, (2010) rely on qualitative aggregation and qualitative weights when including the behaviour associated with condom use in the CFA index. As such condom use is included in the index using qualitative weights, the index is then multiplied by 1 if condoms were used and 0 otherwise.

Another sexual behaviour worth noting emanates from those who have not had sexual intercourse. Bezabih *et al.* (2010) put forth two arguments why this population should be included in the analysis. Firstly the decision to abstain also constitutes an HIV prevention strategy and secondly, sexual debut is likely to be influenced by age, such that older individuals are more likely to be sexually active than

[57] A typical example is the relationship between condom use and type of sex partner. The data shows that there is very low condom use with the '*last partner*' (15.9% of the total sample size). This is quite surprising considering that 77.8% believe that one of the methods to prevent HIV is to use a condom every time you have sexual intercourse. A further analysis of condom use by marital status showed that condom use was highest among those who have never been married (52.7%), followed by the divorced (31.4%), not living together (29.0%), widowed (25.9%), living together (10.3%) and lowest among the married (6.3%).

We also created an indicator for multiple concurrent partners (MCP) by establishing the number of sex partners among those who have been married for more than 12 months. We find that 6.3% of those who have been married have had more than one partner, implying that they are in a MCP. This highlights the link between relationship type and sex partners also supporting our construction of an index.

[58] CFA was conducted using the STATA confa package. For more detail please refer to the Stata Journal volume 9 number 3, further details can be found in the reference section of this paper.

[59] The CFA index consists of type of relationship with last partner, number of sex partners and age of sexual debut.

young individuals[60]. Hence these individuals are included and are represented with zeros in the *risky sex* index (Bezabih *et al.* 2010). Therefore the overall *risky sex* index implies that the higher the score on the index the riskier the sexual behaviour, while zero represents no risk. The distribution of the *risky sex* index is depicted in table 3.C2 in Appendix 3.C[61].

3.4.2.2 Expected Future Health

This section identifies the proxy used for *expected health* and outlines the procedure used to derive this proxy including the challenges of our proposed proxy.

The proxy measure for *expected health* is life expectancy. The rationale being that life expectancy represents one of the best indicators of the overall health of a population (WHO 2010). DHS datasets do not however have direct questions on life expectancy. Therefore the life expectancy is obtained by using the INDEPTH[62] model life tables (MLT) [63]. Our measure overcomes the drawback of bias in self-reported information. The INDEPTH tables are based on the brass relationship life table system (BRLTS). The BRLTS assumes that any two mortality patterns are related through the following relationship: $Y_x = \alpha + \beta Y_{sx} \equiv \text{logit}\,(l_x) = \alpha + \beta\text{logit}\,(l_{sx})$, where l_x and l_{sx} are the survivorship probabilities of the two mortality patterns, and l_{sx} is the standard survivorship pattern. Given this relationship one can determine a complete life table of any population from a *standard life table*. INDEPTH has generated *standard life tables* specific to sub-Saharan Africa which have been classified into *pattern 1* and *pattern 2*. *Pattern 1* is based on populations of West and Central Africa where HIV prevalence is less than 10%, while *pattern 2* depicts those of East and Southern Africa where HIV prevalence is usually above 10% (INDEPTH, 2004). Appendix 3.D provides the detailed methodology of how the life expectancy was derived. Table 3.D1, 3.D2 and 3.D3 of Appendix 3.D shows the distribution of the life expectancies in each country which were calculated using this methodology.

---

60 We tested the argument with DHS data. In these data sources the average age of those who are sexually active is 26 years while those who are not sexually active have an average age of 18 years. Also 89% of those who are not sexually active stated that they intend to postpone sexual intercourse until marriage.

61 The table shows that the population in the study from East African countries, Kenya (the majority 32.7% in the first quartile of the risky sex index), Tanzania (29.8% in the first quartile which represents the majority) and Uganda appear to have less risky behaviour than the population in the Southern Africa countries; Lesotho (the majority, 34.8% are in the fourth quartile), Swaziland (46.5%), although Zimbabweans appear to exhibit less risky behaviour as well. This is consistent with findings recorded in the literature.

62 http://www.indepth-network.org. We employ INDEPTH tables because they allow the calculation of life tables even in cases where not all mortality rates are available for each age/age group, which is the case in the DHS data. DHS data only give information on child mortality rates which have been calculated using the reproduction section of the women's questionnaire, capturing data on total child births and deaths. Additionally, although there are other life tables (for example the United Nations Model Life Table; The Coale and Demeny Model Life Tables; The Lederman Model Life Table System; and The United Nations Model Life Table for Developing Countries), INDEPTH tables are specifically designed for SSA countries, hence our preferred option (INDEPTH 2004).

63 Model life tables present mortality indicators that are typical of given population groups, by levels and patterns of the risks of dying, they are a series of reference life tables that can be used to estimate mortality when only a few indicators are known (INDEPTH 2004: p. 3).

The challenge is however the reverse causality between the INDEPTH life expectancy and our outcome variable *risky sex*. It is explicitly stated that the INDEPTH *standard life tables* are influenced by the prevailing HIV/AIDS mortality (INDEPTH, 2004) and therefore unlikely to be a *non-HIV life expectancy*. Furthermore the *standard life tables* have been classified into *pattern 1* and *pattern 2*, where *pattern 1* is based on populations of West and Central Africa with a HIV prevalence of less than 10%, while *pattern 2* depicts those of East and Southern Africa where HIV prevalence is usually above 10% (INDEPTH, 2004). The section below outlines three different approaches that are employed to overcome this problem of reverse causality where we match different mortality rates (*pre-HIV mortality* rates and *HIV mortality* rates) to either *pattern 1* or *pattern 2* of the life table.

The life expectancy that was obtained from the first approach is called *HIV-life expectancy 2*. This follows the work of Oster (2009)[64], and the second and third proxies are extensions of Oster (2009) work. The child mortality rates from the *2000s DHS datasets* (current datasets) are referred to as the *HIV mortalities*. Using these mortalities *pattern 2* from the INDEPTH *standard life tables* is applied which corresponds with the dataset from East and Southern Africa and reflects high prevalence rates of HIV/AIDS. The *HIV-life expectancy 2* is shown in Table 3.D1 in Appendix 3.D. Once this life expectancy is calculated, instrumental variables (IVs) are used in our final analysis since *HIV-life expectancy 2* is unlikely to be *non-HIV life expectancy*. The second approach maintains the *HIV mortalities* from the 2000s, although in this case *pattern 1* is used instead of *pattern 2* of the model life table, which is based on low incidence of West and Central African mortality patterns. This is named the *HIV-life expectancy 1* and is depicted in Table 3.D2 in Appendix 3.D. However, because the *HIV mortalities* is used this life expectancy is also unlikely to be *non-HIV life expectancy*. Although the reverse causality in the *HIV-life expectancy 1* is likely to be '*lower'* than in the first approach since *pattern 1* is used.

The third and final approach uses child mortality rates from the *1980s DHS datasets*[65] (what we refer to *as pre-HIV mortality* rates) and *pattern 1* from West and Central Africa countries with much lower prevalence rates than *pattern 2* of the Southern and East Africa of the model life table. This is named the *pre-HIV-life expectancy 1* and this proxy is likely to give us *non-HIV life expectancy*. As such IVs are not used when analysing this measure of life expectancy. Table 3.D3 in Appendix 3.D depicts the *pre-HIV-life expectancy 1*.

The use of the different approaches affords us the opportunity to work with three different types of *expected health* proxies: *HIV-life expectancy 2*, *HIV-life expectancy 1* and *pre-HIV-life expectancy 1*. It is worth noting that life expectancy varies at age-gender-country level; please refer to Tables 3DI, 3D2 and

[64] It is worth noting that Oster (2012) uses malaria frequency and child mortality as proxies for life expectancy.

[65] Since DHS surveys were not conducted in Swaziland and Lesotho in the 1980s, we do not have infant and under-five mortality rates for these periods in these countries. We instead use Botswana 1988 DHS mortality rates as we anticipate that these countries had similar health characteristics in the 1980s. The life expectancy for Botswana, Swaziland and Lesotho in 1980 was 65, 61 and 60 respectively showing that the countries had similar health indicators in the 1980s, (WHO: http://www.who.int/healthinfo/statistics/mortality_life_tables).

3D3 in Appendix 3.D of chapter 3. Based on this and on the fact that the social groups are based on income-gender-age and excludes place of residence (country) implies ample variation for analysis.

3.4.2.3 Uncertainty of Future Health

In their study Bezabih *et al.* (2010)'s proxy for *health uncertainty* is the variance of life expectancy calculated at the community level. The likely challenge with this measure is the influence of geographical related unobservables as a result of including place of residence (Blume *et al.* 2010). To avoid this we create a *social space* that is not based on place of residence but rather on the 'similar others' concept using age, gender and income in the social group formation.

The concept *social space* has its origin in sociological literature, the definition of which is: "man's social position is the totality of his relationships towards all groups of population and within each of them, towards its members" (Sorokin, 1927 cited in Prandy and Lambert 2003: p 399). This *social space* may contain many dimensions[66]. However, it is important to note that physical geography is not a necessary dimension in a *social space*. Indeed some school of thought among sociologists is of the opinion that physical geography has become redundant (Conley and Topa 2002). This sociological perspective is supported by others who cite advancement in transportation, telecommunication such as cell phones, internet and social networks as the major players in this perspective. See Manski (2000), Urry (2003) and Xu *et al.*, (2010).

Our social group composition is motivated by the fact that there is a general consensus among social proximity literature that individuals who are *similar* are likely to interact socially. This *similar* other tendency is dubbed as inbreeding or homophily by sociologists, while economists refer to this as positive matching or assortative matching (Conley and Topa 2002). The principle behind *similar* others can be found in Etile (2007) analysis of norms related to body weight. In peer effects studies where classmates are used as groups, under the proposition that they interact because of sharing *similar* courses/classrooms. See De Giorgi *et al.*, (2010); Fletcher (2010) and Lundborg (2006). The adoption of the *similar* others concept can also be found in the work of Carter *et al.*, (2001); Moody (2001); Vendrick (2003); Liljer *et al.*, (2003); Lugalla *et al.*, (2004); Munshi (2006) and Latkin and Knowlton (2006). Finally, the importance of *similar* others is accentuated in Conley and Topa (2002) who note that: "even causal observation suggests that personal networks may be stratified along specific socio-demographic attributes such as race, ethnicity, religious affiliation, language, age, race, gender and education levels, in other words, agents are likely to draw a disproportionate share of their social contacts among sets of people that are very similar to themselves" (Conley and Topa 2002: p. 309).

[66] For example Case *et al.*, (1991) use city blocks, while Conley and Topa (2002) constructs a social space based on physical geography, ethnicity and socioeconomic similarity. Grodner and Kniesner (2005) group is based on individuals who are similar in age, family structure and location. Munshi *et al.*, (2006) uses religion in their group formation. On the other hand Fletcher (2009) and Giorgi *et al.*, (2010) use classmates as groups, while Etile (2007) and Grodner (2006) use individuals of the same occupation, gender and age as the group characteristics.

Table 3.B1 in Appendix 3.B shows the composition of the social groups. As can be seen in Table 3.B1 social group 1 for example comprises 59 individuals with similar ages, gender and income. However these individuals are dispersed across countries: 15 individuals live in Lesotho, 12 live in Swaziland, 13 are in Zimbabwe, 8 individuals are in Tanzania, while 6 are Kenyan and finally the remaining 5 are Ugandan. Additionally within each country the individuals reside in different regions, for example the 12 individuals from Swaziland are distributed as follows within Swaziland: 3 individuals reside in Hhohho, 1 in Manzini, while 5 live in Shiselwen and the last 3 in Lubombo.

Thus instead of looking at the *physical environment* the study looks at the *social environment*, noting the fact that both the *social and physical environment* can produce contextual effects. In addition, defining *health uncertainty* at group level instead of community level maintains the theoretical framework of Mannberg (2010) outlined in section 3.2 where *health uncertainty* is measured as a contextual factor[67]. Furthermore since contextual effects are present in *social and physical environment* (Manski 1993), having health uncertainty proxy as the variance of life expectancy at social group level suffices[68].

Recall that in the previous section three different life expectancies were created: *HIV-life expectancy 2*, *HIV-life expectancy 1* and *pre-HIV-life expectancy 1*. The fact that our proxy measure for *health uncertainty* is the variance of life expectancy means that there are three proxies. These include variance of *HIV-life expectancy 2*, variance of *HIV-life expectancy 1* and variance of *pre-HIV-life expectancy 1*.

3.4.2.4 Expected Income

DHS datasets do not contain any information on either income or expenditure. However, we need a measure for *expected income* as per the theoretical framework. We use different proxy measures of expected income. These proxies include wealth indices (DHS wealth indices and own wealth indices) and household income. One of the motivations for using different measures is because of the fact that these proxies capture essential but different dimensions of expected income. That is, the household income is more reflective of the current earning capacity of the household. In this case we use household expenditure under the assumption that this is likely to be equal to household income. In contrast, durable goods ownership which reflects earnings not easily converted to current income is captured by wealth indices better. Another motivation is that these various measures possess different

[67] We tested our health uncertainty proxy by comparing it with other indicators of health uncertainty. Montgonery and Casterline (1996) found a strong link between subjective health uncertainty and mortality rates. Hence we compare our health uncertainty proxy with the WHO mortality rates of each country. Graph 5.A in appendix 5 shows the distribution of health uncertainty by country. We find health uncertainty to be highest in Zimbabwe and Swaziland. Graph 5.B shows the World Bank mortality rates between 2002 and 2007 in each country. A comparison between the two graphs shows a similar pattern amongst the six countries, where Zimbabwe and Swaziland are ranked the highest, which validates our health uncertainty proxy.

[68] In addition, several studies have alluded to the fact that an individuals' assessment of uncertainty and risk associated with HIV infection, and mortality in general, are linked to the uncertainty and risk of the individuals' social network circles such as family, social groups or community (Montgomery et al. 1996; Grant 2007).

drawbacks which we discuss below. Lastly, we use these different measures to check robustness of our results to different expected income definition.

The DHS datasets come with information on wealth indices. For a discussion on the DHS wealth indices see Rusten and Johnson (2004). However there are two main shortcomings with these indices. Firstly, the DHS wealth indices were calculated using dichotomised[69] principal component analysis (PCA), this is despite the fact that PCA is specifically designed for continuous data. Consequently, applying PCA to categorical data is likely to render the indices biased (Kolenikov, 2009b; Bezabih *et al.*, 2010). Secondly, in calculating these indices, country specific information on household durable goods ownership and housing characteristics is used, which is likely to increase the inaccuracy of comparing the wealth indices across countries.

We derive our own wealth indices following the work of Rusten and Johnson (2004) and Grimm *et al.* (2008). These indices are constructed using household characteristics and durable goods ownership data. Confirmatory Factor Analysis (CFA), with Satorra-Bertler robust standard errors, is employed to aggregate the data into a single index. In the CFA output all the variables had positive factor loading. Additionally, although the RMSEA is slightly higher than expected (0.08), the Tucker-lewis non-normed fit index (TLI) and Bentler's comparative fit index (CFI) were both 0.9 indicating a relative good fit. A more detailed discussion is provided in Appendix 3.5. To address the biasedness of the DHS wealth indices we constructed our own indices using CFA, this measure is however not perfect. Without purporting to be exhaustive, wealth indices do not capture the differences in the quality of durable goods, for example an old black and white television is not distinguished from ownership of a seventy two inch colour television. Importantly, the indices do not adequately capture the differences in preferences for durable goods, which are more prominent, between the rural and urban areas. That is, in rural areas preference is more likely to be towards livestock keeping or farmland ownership whereas the durable goods used in wealth indices are more preferred in the urban areas. Additionally, provision of electricity, sanitation, water and housing are more likely to take place in urban areas than in rural areas in most countries in sub-Saharan Africa, making these indices biased against rural areas. Moreover, access to services such as electricity depends on the availability of infrastructure which is more likely to be available in urban than in rural areas. See Vyas and Kumaranayake (2006); Grimm et al (2008); Booysen et al (2008).

This brings us to the third and final proxy measure, which is household income. To impute the household income the regression based approach was used following the procedure of Rusten and

[69] The use of dichotomised variables in PCA in the calculation of wealth indices was conceptualised by Filmer and Pritchett (2001) where the idea was influenced by the use of dummy variables in regression models. The expectation is that the model would have a better fit. However, Kolenikov (2009b) conducted a Monte Carlo simulation and found that the 'Filmer and Pritchett PCA' was outperformed by the 'Ordinal PCA' (variables are grouped according to rank) and the 'Polychoric PCA'.

Johnson (2004); Grimm et al. (2008) and Oster (2009). This approach is based on the assumption that households smooth out their consumption and therefore monthly household expenditure is likely to be equal to monthly income (Grimm et al. 2008; Oster 2009). In applying the regression based approach we use household and income surveys which are nationally representative surveys designed to capture consumption, expenditure and income patterns. We use the survey which are available to us and include 2008 South African National Income Survey (NIDS) and 2008 Tanzania National Panel Survey (TZNPS). Using these surveys, we regress per capita household expenditure on household characteristics common in NIDS, TZNPS and DHS datasets: $\ln y_h^{\mathrm{NIDS}\,\&\,TZNPS} = \beta^{\mathrm{NIDS}\,\&\,TZNPS} X^{\mathrm{NIDS}\,\&\,TZNPS} + e_h$, where X is household characteristics. Using the estimated coefficients we then calculate income: $\widehat{\ln}\, y_h^{\mathrm{DHS}} = \hat{\beta}^{\mathrm{NIDS}\,\&\,TZNPS} X^{\mathrm{DHS}}$. In Appendix 3.F we explain this process in detail. One of the major drawbacks of this proxy is that it is based on the assumption that household smooth out consumption and as such expenditure is likely to be equal to income, if this assumption however does not hold then the proxy is likely to be biased. Additionally, in imputing this proxy we used parameters estimated from external datasets which raises the question of comparability between the DHS datasets and the external datasets. That is, the explanatory variables in the regression are likely to be inconsistent between the DHS datasets and the external dataset due to either measurement error or differences in the definition of the variables. See Sahn and Stifel (2003); Vyas and Kumaranayake (2006) and Grimm et al (2008).

3.4.2.5 Covariates

Standard control covariates are used following the standard practice recorded in sexual behaviour literature and include: educational attainment, knowledge of HIV prevention methods, and knowing someone with HIV/AIDS or someone who has died of AIDS.

3.4.3 Estimation and Identification Strategy

The most influential literature that addresses identification problems includes the work of Manski (1993, 2000, 2003); Moffit (2001); Grodner and Kniesner (2005); Brock and Durlauf (2004, 2007) and Blume *et al.* (2010). The estimation strategy used here follows their work, in addition to empirical works of Grodner and Kniesner (2005); Munshi and Myaux (2006); Etile (2007) and Giorgi *et al.* (2010). Equation (3.9) is the econometric model that measures the relationship between the *risky sex* index, *expected health*, *expected income* and *health uncertainty*. In the equation y_i is the *risky sex* index, $y_i > 0$ if the individual is sexually active or $y_i = 0$ if the individual is not sexually active, $\tilde{h}_i$ is *expected health*, $\mathbf{X}_i$ are individual control variables, $E[m_i]$ is *expected income*, $\gamma_{\mathrm{health}_g}$ is *health uncertainty* in the social group, $\mathbf{X}_g$ is social group characteristics which is included to avoid omitted variable bias (Manski 1993; Moffit 2001), and u_i is the error term, the unit of analysis in this study is an individual. Equation (3.9) is likely to face four estimation challenges. The first is that of sample selection bias, the second is due to unobservable group level characteristics, the third is social group selection bias, and the fourth challenge is simultaneity (Blume *et al.* 2010).

The first challenge dealing with sample selection bias is due to the inclusion in our model of the population that is not sexually active. The bias arises from the fact that the decision to become sexually active or not is non-random as it is likely to be influenced by many factors such as age where the older you are the more likely it is that you will be sexually active. To control for sample selection bias the study follows suggestions by Greene (2002); Wooldridge (2003); Okten (2004); Bushway *et al.* (2007); Garcia *et al.* (2008); Bezabih *et al.* (2010) and Bollinger and Hirsch (2012) and adopt the censored two stage model (CTSM) as this model includes all observations from the two populations (the sexually active and sexually inactive population) and where missing information are given zero values in our outcome variable. In the first stage a PROBIT model for sample inclusion, determining whether you are sexually active or not, is estimated (equation 3.10) and the inverse Mills ratios from this equation are then included in the second stage IVTOBIT model as an additional regressor.

$$y_i | y_i^s = \alpha + \beta_1.\tilde{h}_i + \beta_2.\mathbf{X}_i + \beta_3.E[m_i] + \beta_4.\gamma_{health_g} + \beta_5.\mathbf{X}_g + u_i \tag{3.9}$$

$$y_i^s = \begin{cases} 1 & \text{if} \quad y_i^{s*} = \alpha' w_i + \varepsilon_i > 0 \\ 0 & \qquad\qquad \text{otherwise} \end{cases} \tag{3.10}$$

where y_i^s is dichotomous and takes the value of 1 if individual *i* is sexually active and 0 otherwise; w_i represents the determinants[70] of becoming sexually active and include age, income, gender, education and knowledge of HIV/AIDS and ε_i is the error term. The unit of analysis in this study is an individual. An exclusion restriction is also included which is a dummy variable capturing whether abstinence is acknowledged as an HIV prevention method. The exclusion restriction is meant to reduce the high multicolinearity between the inverse Mills ration and the regressors in the model, such that if unattended would likely produce unreliable coefficients (Bezabih *et al.* 2010)

The second estimation problem concerns our proxy for *health uncertainty* which, when defined at the community level suffers from the effects of geographically related unobservables (Manski 2000; Blume *et al.* 2010). Unobservable group level characteristics are environmental or institutional conditions such as healthcare facilities. Instead, and as previously mentioned social groups are defined based on age, gender and income across countries. The groups exclude any information on place of residence. Since residential information is usually a major source of unobservables, (Blume *et al.* 2010) our social grouping strategy has had the effect of neutralising geographically related unobservables by assigning individuals into *imagined communities*[71] based on age, gender and income. It is important to note that

[70] Bezabih *et al.* (2010) caution against using variables from the outcome model in the selection model (in our case income or health) as this will result in endogeneity in the selection model.

[71] The 'imagined' communities are defined by the individuals that make up the social group, column 1 of table 3.B1 in the appendix 3.B. For example as previously mentioned social group 1 which comprise 59 individuals who were

modern technology advancement which include cell phones, internet and social networks have drastically weakened the role of physical geography in social interaction, making such social group clusters plausible. See Manski (2000), Conley and Topa (2002), Urry (2003) and Xu *et al*., (2010). Table 3.B1 in Appendix 3.B shows the distribution of the original place of residence (country and region) by social groups and the new *imagined communities* that have been formed as a result of the grouping.

While the geographically related unobservables have been ruled out in the current study, we cannot however rule out the likelihood of our social groups being influenced by non-geographically related unobservables emanating from the fact that the groups consist of individuals of the same age, gender, and income. This similarity means that each social group consist of individuals who are alike, and as such these similar individuals are likely to face similar non-geographically related unobservables. An example is the influence of unobservable educational outcome such as ability, or cleverness or self-discipline which is likely to differ along age lines or income lines. Hence it is plausible that individuals may behave the same as a result of these non-geographically related unobservables. See Brock and Durlauf (2004) and Blume *et al.* (2010). The third challenge is in relation to social group selection bias which arises when individuals are sorted (un)intentionally into groups, yielding biased estimates See Angrist and Pischke (2008) on selection bias and randomisation. In the current study selection bias is due to the non-random assignment of individuals into social groups because of clustering these individuals based on similar socioeconomic characteristics. Hence it is difficult to view this clustering as untainted by selection bias.

The fourth and final challenge is simultaneity between *expected health* ($\tilde{h}_i$), and *health uncertainty* (γ_{health_g}), and the outcome variable *risky sex*. In the case of *expected health* ($\tilde{h}_i$) simultaneity occurs when *HIV-life expectancy 2* and *HIV-life expectancy 1* proxies are used in the model, where simultaneity is a result of the proxy being non-HIV life expectancy. There is likely to be simultaneity between *risky sex* and contextual *health uncertainty* because while an increase in uncertainty can induce an increase in risk, by the same token an increase in risk can increase uncertainty mainly because the uncertainty is health related. Furthermore, as previously alluded to *health uncertainty* is measured by the variance of life expectancy, because *HIV-life expectancy 2* and *HIV-life expectancy 1* have a simultaneity relationship with our outcome variable it is likely that the *health uncertainty* proxies derived from these life expectancies will also inherit the simultaneity relationship.

Recall that instead of considering the *physical environment* the study is based on the *social environment*, specifically social groups, under the knowledge that both the *social and physical environment* can produce

originally from the following countries and region, 15 from Lesotho, 12 from Swaziland, 13 from Zimbabwe, 8 from Tanzania, 6 from Kenya and the final 5 from Uganda. These individuals have now been grouped into the social group 1 and 'imaginary' community 1. Furthermore within each country the individuals reside in different regions, for example the 12 individuals from Swaziland are distributed as follows 3 in Hhohho, 1 in Manzini, 5 in Shiselwen and 3 in Lubombo.

contextual effects. See Manski (1993). The key to our instrumental variable (IV) strategy is social grouping. Recall that the social groups consist of '*similar* others' based on same gender, age and income. Each of these social groups contains the average *expected health* and the *health uncertainty* which is calculated as the variance of life expectancy.

In this vein, the *trendsetter IV* is an extension on the work of Grodner and Kniesner (2006); Etile (2007) and Grodner *et al.* (2011). This instrumental variable which introduces *trendsetters* and age variation between social groups is grounded on the premise that older social groups are more likely to set the trend for younger groups in similar socioeconomic (gender and income) neighbourhoods. As a result it is suffice to say that the older social groups *expected health* and *health uncertainty* can be used as IVs for younger social groups whereas the opposite is not true, since:
$\mathrm{cov}\,(\gamma_{\mathrm{health}_{\mathrm{oldg}}}, \gamma_{\mathrm{health}_{\mathrm{youngg}}}) \neq 0, \mathrm{cov}\,(\gamma_{\mathrm{health}_{\mathrm{old\,g}}}, u_{\mathrm{young\,ind}}) = 0$ whereas, $\mathrm{cov}\,(\gamma_{\mathrm{health}_{\mathrm{young}}}, u_{\mathrm{young\,ind}}) \neq 0$. The assertion that older social groups are more likely to be *trendsetters* of younger groups in *similar* socioeconomic neighbourhoods is supported by a series of literature.

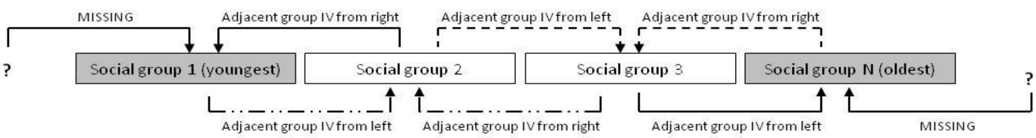
First and foremost studies of Helleringer and Kohler (2005), Anglewicz (2006) and Kohler *et al* (2007) have interestingly observed that the HIV/AIDS risk perception even the levels of worry in a social network have a strong link with the individuals assessment of their own risk and their level of worry. Against this background our argument is that the health related contextual factors of younger social groups are likely to be based on health related trends from the older social groups. This is because the current literature shows that trends are usually initiated into a particular sub-cultures or groups by early adopters or initiators before spreading to other groups in the social system or culture. The early adopters or initiators are the so called *trendsetters* (Pinkerton *et al.* 1995; Bertrand, 2010; Salvini and Vignoli 2011). Consequently, the key insight is that trends do not just appear but rather they emerge in some sub-cultures and diffuses to other sub-groups. See Hendlin *et al.* (2009) and Alexander *et al.*, (2001) for *trendsetters* and smoking; Kelly *et al.* (1992), Kelly *et al.* (1993), Pinkerton *et al.* (1995), Geary *et al.* (2006) and Bertrand (2010)for *trendsetters* and sexual behaviours related to HIV/AIDS; Dragone and Savorelli 2012 on fashion industry and *trendsetters*; Salvini and Vignoli (2011) in marital disruption; Montgonery *et al.*, (1998) and D'Addato *et al.*, (2008) in changes of fertility patterns. As such based on this trendsetting concept the $\mathrm{cov}\,(\gamma_{\mathrm{health}_{\mathrm{oldg}}}, \gamma_{\mathrm{health}_{\mathrm{youngg}}}) \neq 0$.

The *trendsetter IV* (older social groups) are correctly excluded because although the contextual factors of the young and the old social groups will be correlated, the '*similar* others' concept dictates that it is only the contextual factors in your own group that are likely to influence your own individual behaviour. Hence the assertion that $\mathrm{cov}\,(\gamma_{\mathrm{health}_{\mathrm{old\,g}}}, u_{\mathrm{young\,ind}}) = 0$. The limitation of this IV method is that the oldest social group is always excluded[72] as no older social group exists that can be used as IV[73]. Furthermore the IVs make these models exactly identified. We also use the *adjacent social group age bracket IV* based on the work of Grodner and Kniesner (2006); Etile (2007) and Grodner *et al.* (2011) in Appendix 3.I. The Anderson canon correlation test is adopted based on the null hypothesis that the model is underidentified, that is, IVs are irrelevant and are not correlated with endogenous regressors. The study also tests for overidentifying restrictions using the Sargan-Hansen test on the joint null hypothesis that the instruments are valid, that is, the instruments are uncorrelated with the error term and that the instruments are correctly excluded from our equation as regressors. In the case of the non-linear model, IVTOBIT, the Amemiya-Lee Newey statistic test for overidentification which is similar to the Sargan-Hansen tests is used.

Our final estimation model is a selectivity corrected *nonlinear-in-means* model as depicted in equation (3.11). It is worth mentioning that the standard linear-in-means model of social interactions includes as regressors the individual controls, the average outcome of the group (endogenous effect) and the group averages (exogenous effect). However, we do not include the average outcome (average risky sex) as one of the regressors following our theoretical framework, although this does not violate the functional form of the model.

$$y_i|y_i^s = \alpha + \beta_1.\tilde{h}_i + \beta_2.\mathbf{X}_i + \beta_3.E[m_i] + \beta_4.\gamma_{health_g} + \beta_5.\mathbf{X}_g + \beta_6.\lambda_i + u_i \qquad (3.11)$$

Where y_i is the *risky sex* index, $y_i > 0$ if the individual is sexually active or $y_i = 0$ if the individual is not sexually active, $\tilde{h}_i$ is *expected health*, $\mathbf{X}_i$ are individual control variables, $E[m_i]$ is *expected income*, γ_{health_g} is *health uncertainty* in social group g, $\mathbf{X}_g$ is social group characteristics, λ_i is the Mills ratio, and u_i the error term. Our estimation of the selectivity corrected outcome model uses STATA's IVTOBIT model. This is motivated by the fact that the TOBIT model accommodates outcome variables with excess zeros, which is the case for our model where the zeros capture those who are not sexually active. Angrist *et al.* (2008) has found that one gets similar results with OLS estimation as with TOBIT or PROBIT models. As such we augment the IVTOBIT model with the GMM estimation especially given that the linear models in STATA are less constrained and allow for more post-estimation tests.

[72]


[73] We include the '*extended' trend setter of the norm IV* based on trend *setter of the norms IV*. The difference between the two is that the omitted oldest social group is included by instrumenting this group with another oldest social group by different characteristics such as gender. This is also to allow the inclusion of the omitted social group since this IV eliminates the last social group which contains the oldest age group.

3.5 Estimation Results

3.5.1 Regression Results

The analysis in Table 3.1 is based on the *objective life expectancy* calculated using infant mortality, under five mortality and INDEPTH Model life tables and depicts the results of our regression equation (3.11). In ***panel A*** the current child mortality rates (what we refer to as *HIV mortalities*), and apply *pattern 2* from INDEPTH life tables since our current dataset (2000s) is from East and Southern Africa to derive the *HIV-life expectancy 2* used. As such we **use IV** as it is unlikely to be *non-HIV life expectancy*. In ***panel B*** we calculate the life expectancy using *pre-HIV mortalities* from DHS data from the 1980s and *pattern 1* of the INDEPTH model life table, which is based on data from West and Central Africa. Here the *pre-HIV-life expectancy 1* is obtained, and this is likely to give us a *non-HIV life expectancy*. We therefore **do not use IV** when analysing this measure of life expectancy.

The results are based on the second stage CTSM model which involves using the inverse Mills ratio from the first stage Heckman model as an additional regressor in the models. This is so as to eliminate the selection bias in the decision to become sexually active or not, mainly because our sample includes all individuals sexually active or not. The results of the first stage Heckman model are shown in table 3.H1 in Appendix 3.H. Columns 1-2 (Panel *A.1*) and 5-6 (Panel *B.1*) of Table 3.1 are based on the *non-linear-in-means* models, specifically the IVTOBIT model using the censored *risky sex* outcome variable represented by 0 if an individual is not sexually active and > 0 if an individual is sexually active and hence captures the risk associated with sexual behaviour. While columns 3-4 (Panel *A.2*) and 7-8 (Panel *B.2*) use the *linear-in-means* model as an alternative following Angrist and Pischke (2008) who have found that we obtain similar results with OLS or 2SLS estimation as with PROBIT or TOBIT model when the outcome variable is binary or censored. As such the GMM model estimation is used especially given that the linear models in STATA are less constraining and allow for more post-estimation tests like the Sargan-Hansen statistics test for overidentification available in linear models. The models presented in this paper are based on robust standard errors with clustering at the social group level. See Moulton (1990); Williams (2000) and Wooldridge (2003) for a discussion on robust standard errors and clustered data.

Table 3.1 shows that the sample selection bias correction term, the coefficient of the inverse Mills ratio is significant at the 1% level for all models. This indicates that the sample selection bias as a result of including the population that is not sexually active is indeed present in the model and that the use of the CTSM is warranted. Additionally our main regressors, *expected income*, *expected health* and contextual *health uncertainty* are statistically significant at the 1% level and possess the theoretically expected signs. Specifically, our first regressor *expected income* has an ambiguous effect on *risky sex*. This ambiguity persists even after substituting the household income proxy with the wealth index proxy. The results based on the wealth index proxy are shown in table 3.K1 in Appendix 3.K. We note that the results, however, should be interpreted with caution given the drawbacks previously highlighted and in light of our assumptions in section 3.4.2.4.

On the other hand *expected health* has the negative theoretically anticipated sign. In addition and more importantly the variable is significant in determining individual sexual choices, such that as an individual's expected future health improves, they are more likely to revise their sexual behaviours to less risky behaviour. The implication is that an increase in future health increases the incentive to abstain from unsafe sex as predicted by Mannberg's (2010) theory. We find consistent results even after using three different types of *expected health* proxies, namely, *HIV-life expectancy 2* in panel *A*, *pre-HIV-life expectancy 1* in panel *B* and *HIV-life expectancy 1*. Bezabih *et al.* (2010) found similar results among the youth of Cape Town, although the coefficient was not significant. It is worth noting that the *expected health* proxy was subjective life expectancy while our current measure is an objective measure, as such the difference in results could be attributed to differences in the proxy.

The positive *health uncertainty* coefficient supports Mannberg's (2010) theory that the contextual effects of *health uncertainty* is likely to reduce the incentive to abstain from unsafe sex, as the coefficient is statistically significant. Similar results were obtained by Bezabih *et al.*, (2010) where *health uncertainty* was defined at community level to be insignificant.

Another robustness check was to use alternative outcome variables. These alternative variables include number of sex partners and condom use where consistent results are found in all models. Table 3.M1 of Appendix 3.M shows the results based on number of sex partners as the outcome variable, while in table 3.M2 the outcome variable is condom use. The study tests for model identification using the Anderson canon correlation test, where all our models reject the null hypothesis implying that our models are identified, and we show these results in the bottom rows of all our tables. We also test for the overidentifying restrictions in the GMM (IVTOBIT) models using the Sargan-Hansen (Amemiya-Lee Newey) test. Overidentification is only present when the *adjacent social group age brackets IVs* are used. The results are displayed on the last rows of the Tables in Appendix 3.I. Most of the models are weakly identified as they fail to accept the null hypothesis.

Interestingly, the results do not appear to be influenced by the magnitude of sexual risk taking as our findings appear to be similar for both the East African and Southern African countries selected. This is despite of the fact that the Southern African region is regarded as more risky in the existing literature. The results are shown in Table 3.LK1 in Appendix 3.L.

3.5.2 Social Group Falsification Test

A final robustness check is performed to validate our social group composition. To achieve this we construct a *placebo social group* and use this social group to perform a falsification test. This falsification test is conducted using '*placebo treatments*' following Lavy and Schlosser (2007) and is aimed at ensuring that the *social environment* effects found from our proposed social grouping are not the result of measurement errors, such as unobservables. Hence the expectation is that insignificant *social environment* effects will be found when *placebo social groups* are used (Lavy and Schlosser 2007).

Some of the recent studies that have used this falsification test include the research of Giorgi *et al.* (2010) who sought to explain how students select courses based on the influence of other students who took similar courses (where the reference groups consisted of students who took the same courses). Thus for the falsification test Giorgi *et al.* (2010) replaced the groups with placebo groups that were constructed by randomly and artificially allocating the students to hypothetical classes. The placebo social grouping showed no evidence of peer effects, verifying the authors' expectations. In another study Fletcher (2009) analysed how classmates smoking behaviour influenced youth smoking decisions. For the falsification test Fletcher (2009) replaced classmates (peers who smoke) with those in lower or higher grades from the same school, as expected the author found no evidence of peer effects when lower or higher grade classes were used.

Following these studies we apply the falsification test where our *placebo social groups* are assembled by randomly assigning individuals to social groups using ad hoc variables. Recall that our social groups are based on '*similar* others' concept on the assumption that individuals who share the same socioeconomic characteristics are likely to interact. In this vein, the falsification test uses *placebo social groups* which randomly assign individuals into groups using ad hoc '*similar* others' variables. Simply put we are arguing that not any '*similar* others' characteristics will lead to social interaction. These variables include the day the respondent was interviewed, the respondents' month of birth and sample weight information. Our expectation is that there should not be evidence of *social environment* effects based on these *placebo social groups*. Table 3.2 shows the results based on the *placebo social groups* using the *trendsetter* IV, in Appendix 3.J we show the results based on the *adjacent social group age brackets IVs*. As is evident from the table we find no significant effects on the *social environment* variables, confirming that our findings in Table 3.1 are not spurious but are indeed picking up *social environment* effects.

Table 3.1: Second Stage CTSM Risky Sex Results

	Objective life expectancy							
	Panel A *HIV-life expectancy 2* (2000s mortality rates)				Panel B *Pre-HIV-life expectancy 1* (1980s mortality rates)			
	Panel A.1		Panel A.2		Panel B.1		Panel B.2	
	IVTOBIT Nonlinear in means model		GMM Linear in means model		IVTOBIT Nonlinear in means model		GMM Linear in means model	
	Trend setter IV	'Extended' Trend setter IV	Trend setter IV	'Extended' Trend setter IV	Trend setter IV	'Extended' Trend setter IV	Trend setter IV	'Extended' Trend setter IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Expected health	**-0.179*****	**-0.404*****	**-0.240*****	**-0.522*****	**-0.0109*****	**-0.0132*****	**-0.0137*****	**-0.0166*****
	(0.00488)	**(0.0236)**	**(0.00658)**	**(0.0306)**	**(0.000341)**	**(0.000303)**	**(0.000420)**	**(0.000375)**
Health uncertainty	**0.116*****	**0.295*****	**0.157*****	**0.383*****	**0.0143*****	**0.0128*****	**0.0181*****	**0.0162*****
	(0.00414)	**(0.0187)**	**(0.00558)**	**(0.0242)**	**(0.000216)**	**(0.000220)**	**(0.000268)**	**(0.000273)**
Social group characteristics	-1.246***	-3.158***	-1.686***	-4.097***	0.624***	0.478***	0.780***	0.598***
	(0.0462)	(0.204)	(0.0622)	(0.265)	(0.0136)	(0.0138)	(0.0168)	(0.0170)
Expected income	0.0324***	0.0790***	0.0439***	0.103***	-0.0126***	-0.00920***	-0.0157***	-0.0114***
	(0.00118)	(0.00508)	(0.00159)	(0.00658)	(0.000317)	(0.000311)	(0.000390)	(0.000384)
Education, years	0.0342***	0.0470***	0.0460***	0.0613***	0.0169***	0.0164***	0.0214***	0.0208***
	(0.00184)	(0.00437)	(0.00247)	(0.00566)	(0.000998)	(0.000881)	(0.00123)	(0.00109)
Mills ratio	-68.31***	-30.45	-91.56***	-41.14	-90.88***	-86.69***	-116.9***	-110.8***
	(11.43)	(25.84)	(15.39)	(33.47)	(6.395)	(5.615)	(7.928)	(6.999)
HIV prevention, one partner	0.0250**	-0.00694	0.0354**	-0.00634	0.0384***	0.0396***	0.0511***	0.0520***
	(0.0127)	(0.0288)	(0.0171)	(0.0373)	(0.00711)	(0.00629)	(0.00884)	(0.00787)
Healthy looking person, HIV	0.119***	0.219***	0.160***	0.286***	0.0433***	0.0508***	0.0583***	0.0671***
	(0.0127)	(0.0301)	(0.0171)	(0.0390)	(0.00695)	(0.00617)	(0.00865)	(0.00772)
Rural	-0.123***	-0.348***	-0.166***	-0.450***	0.0649***	0.0552***	0.0818***	0.0697***
	(0.00995)	(0.0299)	(0.0134)	(0.0387)	(0.00515)	(0.00460)	(0.00633)	(0.00568)
Southern Africa	1.514***	3.359***	2.033***	4.343***	0.0655***	0.0722***	0.0822***	0.0906***
	(0.0440)	(0.201)	(0.0594)	(0.261)	(0.00468)	(0.00405)	(0.00575)	(0.00501)
Constant	10.86***	10.26***	14.17***	13.12***	10.29***	9.872***	12.94***	12.33***
	(1.164)	(2.609)	(1.568)	(3.379)	(0.652)	(0.573)	(0.808)	(0.714)
Observations	48,430	52,155	48,430	52,155	48,430	52,155	48,430	52,155
Anderson canon correlation statistics (p-value)			0.0000	0.0000			0.0000	0.0000

- Robust Standard errors in parentheses
- *** p<0.01, ** p<0.05, * p<0.1 indicates significance at 1%, 5% and 10%
- Rural reference is Urban and region reference is Eastern Africa
- Instrumented for expected health and health uncertainty in panel A and health uncertainty in panel B
- Excluded instruments: mean expected health of older social group, health uncertainty of older social group

Table 3.2: Second Stage CTSM Risky Sex Results using Placebo Social Groups

	Panel A				Panel B			
	HIV-life expectancy 2 (2000s mortality rates)				*Pre-HIV-life expectancy 1* (1980s mortality rates)			
	Panel A.1 IVTOBIT		**Panel A.2** GMM		**Panel B.1** IVTOBIT		**Panel B.2** GMM	
	Nonlinear in means model		Linear in means model		Nonlinear in means model		Linear in means model	
	Trend setter IV	'Extended' Trend setter IV	Trend setter IV	'Extended' Trend setter IV	Trend setter IV	'Extended' Trend setter IV	Trend setter IV	'Extended' Trend setter IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Expected health	0.158	0.0880	0.213	0.117	-0.196	-0.00990	-0.255	-0.0131
	(0.611)	(0.0972)	(0.817)	(0.128)	(0.300)	(0.0103)	(0.391)	(0.0134)
Health uncertainty[1]	**0.0124**	**-0.0145**	**0.0174**	**-0.0199**	**0.128**	**-0.00176**	**0.167**	**-0.00192**
	(0.0979)	**(0.0340)**	**(0.131)**	**(0.0448)**	**(0.209)**	**(0.00724)**	**(0.273)**	**(0.00941)**
Social group characteristics	-0.102	-0.179	-0.134	-0.242	0.197	0.00323	0.257	0.00470
	(0.165)	(0.282)	(0.220)	(0.371)	(0.320)	(0.0147)	(0.417)	(0.0191)
Income	0.00309	0.00216*	0.00410	0.00282*	0.00549	0.000949***	0.00715	0.00124***
	(0.00813)	(0.00130)	(0.0109)	(0.00171)	(0.00737)	(0.000261)	(0.00962)	(0.000340)
Education, years	0.0238	0.0188***	0.0314	0.0246***	-0.00351	0.0202***	-0.00445	0.0264***
	(0.0157)	(0.00260)	(0.0211)	(0.00343)	(0.0398)	(0.00180)	(0.0520)	(0.00234)
Mills ratio	-252.3	-181.1**	-335.6	-238.4**	17.90	-92.46***	22.52	-121.1***
	(531.4)	(75.47)	(711.0)	(99.41)	(187.3)	(7.471)	(244.5)	(9.751)
HIV prevention, one partner	0.256	0.170	0.343	0.225	-0.129	0.0570***	-0.167	0.0757***
	(0.716)	(0.109)	(0.959)	(0.144)	(0.299)	(0.0130)	(0.390)	(0.0169)
Healthy looking person, HIV	0.337	0.252	0.451	0.333	0.319	0.0748***	0.417	0.0999***
	(0.935)	(0.166)	(1.251)	(0.218)	(0.409)	(0.0138)	(0.534)	(0.0180)
Southern Africa	-1.323	-0.754	-1.789	-1.005	1.737	-0.0187	2.264	-0.0215
	(4.937)	(0.772)	(6.607)	(1.016)	(2.836)	(0.0961)	(3.703)	(0.125)
Constant	18.51	16.17***	24.10	20.97***	1.699	10.47***	1.986	13.41***
	(20.98)	(5.706)	(28.08)	(7.517)	(15.23)	(0.675)	(19.89)	(0.882)
Observations	35,479	52,155	35,479	52,155	35,479	52,155	35,479	52,155

- Robust Standard errors in parentheses
- *** p<0.01, ** p<0.05, * p<0.1 indicates significance at 1%, 5% and 10%
- Definition of instruments follows from table 3.1.
- Region reference is Eastern Africa
- Health uncertainty is not significant in all models. Furthermore, most of the variables in the models are insignificant.

3.6 Conclusion

The aim of this study was to investigate the role of *expected income*, *expected health* and the contextual effects of *health uncertainty* on *risky sex*, an aggregation of risky sexual behaviour. The impetus for this study is the current gap in empirical literature that has assessed the role of contextual factors driving sexual behaviours associated with the risk of HIV/AIDS. To test this relationship we employ a theoretical model formulated by Mannberg (2010). The analysis was made using data from six DHS of selected sub-Saharan African countries, namely Tanzania, Kenya and Uganda in East Africa, and Swaziland, Lesotho, and Zimbabwe from Southern Africa. Fundamentally, the study differs from the conventional sexual behaviour research recorded in the literature and takes the view that to gain a more accurate understanding of this topic, one needs to aggregate individual sexual behaviours. This is because, intuitively, individuals who deem themselves to be practicing *risky sexual behaviour* are more likely to adopt other *protective sexual behaviour* to counteract their overall risk (Bezabih *et al.* 2010). This study therefore aggregated the behaviours into a *risky sex* index and established overall sexual risk using confirmatory factor analysis specifically designed for categorical data (Bezabih *et al.* 2010; Kolenikov 2009a).

In addition, the study includes individuals who are not sexually active in the analysis as we believe that the decision to abstain also constitutes a *protective sexual behaviour* (Bezabih *et al.* 2010). A censored two stage model is used to eliminate sample selection bias (Greene 2002; Wooldridge 2003; Bushway *et al.* 2007; Bezabih *et al.* 2010), where in the first stage a selectivity model for sample inclusion is estimated and the inverse Mills ratios are used as an additional regressor in the second stage model. The sample selection bias is a result of including the individuals who are not sexually active in the analysis and this is because the decision to become sexually active is non-random.

The study uses two different proxies for *expected income*, the first is household income which is obtained through the regression based approach following the work of Oster (2009) and Grimm *et al.* (2008). The second is wealth indices which are calculated following the work of Rusten and Johnson (2004) and Grimm *et al.* (2009). Our departure from their work is that we use confirmatory factor analysis to derive the indices, instead of the principal component analysis due to the bias produced by principal component analysis on categorical data (Kolenikov 2009b; Mannberg 2010).

We follow Bezabih *et al.* (2010) and proxy *expected health* with life expectancy. We use the INDEPTH model life tables to construct life expectancies because DHS data does not contain any information on life expectancy. The proxies which are an extension of Oster (2009) were constructed using alternative mortality rates and INDEPTH life tables patterns and include *HIV-life expectancy 2*, *HIV-life expectancy 1* and *pre-HIV-life expectancy 1*. The motivation for this is that INDEPTH model life tables are likely to produce *HIV life expectancy,* and this will have reverse causality with our outcome variable. A *non-HIV* *life expectancy* will be exogenous in our model, and we believe this to be the case with the *pre-HIV-life expectancy 1*. Our proxy measure for the contextual effects of *health uncertainty* is the variance of life expectancy following Bezabih *et al.* (2010). The difference with our proxy from Bezabih *et al.* (2010) is that the study calculates this at a social group level, while they calculate their proxy at the community level. The reason for this departure is that the place of residence brings with it geographical related unobservables (Blume *et al.* 2010) and presence of unobservables renders the model unidentified. The instruments used to circumvent the other identification problem, follow the work of Grodner and Kniesner (2005); Etile (2007) and Grodner *et al.* (2011). Two different IV strategies are employed motivated by the social interaction literature and social grouping. The first is the *trend setter* instrument and the second is the *adjacent social group age bracket* instrument.

Against this backdrop, the results from the selected DHS data generally reveal a significant negative relationship between *expected health* and *risky sex* and a significant positive relationship between *health uncertainty* and *risky sex*. The results do indeed shed some light on the pivotal role played by health and the underlying health related contextual factors as predicted by Mannberg (2010)'s theory. Interestingly, this pattern was observed for both the East African region and Southern African region, despite of the fact that the sexual behaviour of individuals in the Southern African region is regarded as more risky than that of individuals in the East African region in existing literature on the topic. This observation could be rationalised by the view that health and the associated contextual uncertainty are instrumental in influencing sexual behaviour regardless of the magnitude of sexual risk taking. Furthermore, the sub-Saharan African region in general is a region possessed by health uncertainty, for example uncertainty about the availability and/or affordability of health care including the uncertainty of adequate nutrition.

Our findings are consistent with related studies that have employed data from different sub-Saharan African countries as case studies. For example Oster (2009, 2012) using DHS data found low life expectancy to reduce the incentive to abstain from safe sex. In turn Bezabih *et al.* (2010)'s study on youth in Cape Town, South Africa found future health to increase the incentive to abstain from sexual risk taking, although *health uncertainty* was found to be insignificant. However, it is worth noting that Bezabih *et al.* (2010) used a subjective measure of life expectancy as a proxy for expected health, while our proxy is based on objective measures of life expectancy. We are therefore left to question whether the difference in the results could be explained by the differences in proxies. In addition the results are similar to other studies in sub-Saharan Africa that have identified increased sexual risk taking in war zones areas and among commercial sex workers, for instance in studies by UNAIDS (2002) and Benz (2005). Furthermore, these findings are also consistent with qualitative research on uncertainty and sexual behaviour in sub-Saharan Africa (Varga 1997, 2001).

Overall our findings support UNAIDS '*know your local epidemic*' advocacy for understanding the contextual factors related to HIV/AIDS infection, as health and uncertainty appear to be background factors hindering the change in sexual behaviour. Consequently, this study points to these factors as being pivotal ingredients in the fight against HIV/AIDS, especially given that 80% of new HIV infections currently occur through heterosexual transmission (UNAIDS 2010a). One of the main arguments for this evolution besides medical advancement is the persistent lack of change in sexual behaviour in sub-Saharan Africa (Akwara *et al*. 2003; Maharaj and Cleland 2004; Wellings *et al*. 2006; Kongnyuy and Wiysonge 2007; Morris 2010). Thus it becomes paramount to look at the context within which sexual behaviours can be altered.

This study is not without limitation. The DHS datasets do not contain any information on income or life expectancy. As a result these variables had to be constructed. However, because of the fact that our empirical model uses two constructed regressors under the given assumption and if these assumptions do not hold, these constructed regressors are likely to be poor proxies. As such one of the alternative models used in this paper uses wealth indices that are available in the DHS dataset. Hence in such models the only constructed variable becomes the life expectancy. This alternative model can be found in Table 3.J1 in Appendix 3.J at the end of chapter 3.

Additionally, the social groups cannot be identified in the DHS datasets as such great efforts were placed in identifying the social groups. However, despite these great efforts we cannot deny the fact that theory *open-endness* of social interaction literature makes it difficult to identify these social groups with certainty, and hence makes it easy for such social interaction models to be vulnerable to measurement error. This weakness continues to invade social interaction models. Finally, the instruments depend on social groups which imply that if the assumptions placed on social grouping (that is proper identification of groups and the lack of social group self-selection) do not hold we are likely to be faced with weak or even redundant instruments. We do take comfort in the fact that the post estimation test revealed that the IVs were uncorrelated with the error term and correctly excluded from the equation.

3.7 References

Adetunji, J. and Meekers, D. (2001). Consistency in condom use in the context of HIV/AIDS in Zimbabwe, *Journal of Biosocial Science,* 33: 121-138.

Adimora, A. and Schoenbach, V. (2005). Social context, sexual networks, and racial disparities in rates of sexually transmitted infections, *The Journal of Infectious Diseases*, 191: 115-122.

Airhihenbuwa, C. and Webster, J. (2004). Culture and African contexts of HIV/AIDS prevention, care and support, *Journal of Social Aspects of HIV/AIDS Research Alliance*, 1 (1): 4-13.

Akerlof, G., and Kranton, R. (2005). Identity and the economics of organisation, *The Journal of Economic Perspective*, 19 (1): 9-32.

Akwara, P., Madise, N. and Hinde, A. (2003). Perception of risk of HIV/AIDS and sexual behaviour in Kenya, *Journal of Biosocial Science,* 35: 385-411.

Alexander C., Piazza M., Mekos D. and Valente T. (2001). Peers, schools and adolescent cigarette smoking, Journal of Adolescent Health, 29: 22-30.

Anglewicz P. (2006). Overestimating HIV infection: The construction and accuracy of subjective probabilities of HIV in rural Malawi, Population Association of America, Los Angeles, CA.

Anderson, K. and Beutel, A. (2007). HIV/AIDS prevention knowledge among youth in Cape Town, South Africa, *Journal of Social Sciences*, 3 (3): 143-151.

Angrist, J. and Pischke, J. (2008). Mostly harmless econometrics: An empiricist's companion, Princenton, NJ: Princenton University Press.

Armitag,e C. and Conner, M. (2000). Social cognition models and health behaviour: A structured review, *Psychology and Health*, 15: 173-189.

Aronowitz, T., Rennells, R. and Todd, E. (2005). Heterosexual behaviours in early adolescent African American girls: The role of mother-daughter relationships, *J Fam Nurs*, 11 (2): 122-139.

Avalos, L., Mertens, J., Ward, C., Flisher, A., Bresick, G. and Weisner, C. (2010). Stress, substance use and sexual risk behaviours among primary care patients in Cape Town, South Africa, *AIDS Behaviour*, 14: 359-370.

Bachanas, P., Morris, M. and Lewis-Gess, J. (2002). Predictors of risky sexual behaviour in African American adolescent girls: implications for prevention interventions, *Journal of Paediatric Psychology*, 27 (6): 519 -530.

Benefo, K. (2008). Determinants of Zambian men's extra-marital sex: A multi-level analysis, *Arch Sex Behaviour*, 37: 517-529.

Bertrand J. (2010). Diffusion of Innovations and HIV/AIDS, Journal of Health Communication: International Perspectives, 9 (S1): 113-121.

Bezabih, M., Mannberg, A. and Visser, M. (2010). Sexual risk taking among young adults in Cape Town: Effects of expected health and income, SALDRU Working Paper Number 43.

Bingenheimer, J. and Raudenbush, S. (2004), Statistical and substantive inferences in public health: Issues in the application of multilevel models, *Annual Review of Public Health*, 25: 53-77.

Blume, L., Durlauf, S., Brock, W. and Ionnides, Y. (2010). Identification of social interactions. Assessed 02 February 2011. http://ssrn.com/abstract=1660002.

Bollinger, C. and Hirsch, B. (2012). Is earnings nonresponse ignorable?, *The Review of Economics and Statistics*.

Bongaarts, J. (2006). Late marriage and the HIV epidemic in sub-Saharan Africa, Policy Research Division Working Papers, *The Population Council*, 216: 1-24.

Bound, J., Jaeger, D. and Baker, R. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogeneous explanatory variable is weak, *Journal of the American Statistical Association*, 90 (430): 443-450.

Booysen F., Van Der Berg S. and Burger R. (2008). Using an asset index to assess trends in poverty in seven sub-Saharan Africa countries, World Development, 36 (6): 1113-1130.

Bracher, M., Santow, G. and Watkins, S. (2003). "Moving" and marrying: modelling HIV infection among newly-weds in Malawi, *Demographic Research,* 1 (7): 208-244.

Brock, W. and Durlauf, S. (2004). Chapter 54 interaction-based models, *Handbook of Econometrics*, 5: 3297-3380.

Brock, W. and Durlauf, S. (2007). Identification of binary choice models with social interactions, *Journal of Econometrics*, 140: 52-75.

Buhi, E. and Goodson, P. (2007). Predictors of adolescent sexual behaviour and intention: A theory-guided systematic review, *Journal of Adolescent Health*, 40: 4-21.

Bushway, S., John, B. and Slocum, L. (2007). Is the magic still there? The use of the Heckman two-step correction for selection bias in criminology, *Journal of Quantitative Criminology*, 23 (2): 9024-9040.

Burgard, S. and Lee-Rife, S. (2009). Community characteristics, sexual initiation, and condom use among young Black South Africans, *Journal of Health and Social Behaviour,* 50: 293-309.

Callegari, L., Harper, C., Straten, A., Kamba, M., Chipato, T. and Padian, N. (2008). Consistent condom use in married Zimbabwean women after a condom intervention, *Sexually Transmitted Diseases*, 35 (6): 624-630.

Campbell, C. (1997). Migrancy, masculine identities and AIDS: The psychosocial context of HIV transmission on the South African gold mines, *Social Science Medicine,* 45 (2): 279-281.

Cleland, J., Ali, M. and Shah, I. (2006). Trends in Protective Behaviour among Single vs. Married Young Women in Sub-Saharan Africa: The big picture, *Reproductive Health Matters*, 14 (28): 17-22.

Coates, T., Richter, L. and Caceres, C. (2008). Behavioural strategies to reduce HIV transmission: How to make them work better, *Lancet*, 372: 669-684.

Coleman, E. (2002). Promoting sexual health and responsible sexual behaviour: An introduction, *The Journal of Sex Research*, 39 (1): 3-6.

Courtenay-Quirk, C, Pals, S., Colfax, G., McKirnan D., Gooden L. and Dogan Eroglu D. (2008). Factors associated with sexual risk behaviour among persons living with HIV: Gender and sexual identity group differences, *AIDS Behaviour*, 12: 685-694.

D'Addato A., Vignoli D. and Yavuz S. (2008). Towards smaller family size in Egypt, Morocco and Turkey: Overall change over time or socio-economic compositional effect? Population Review, 47 (1): 21-40.

Deaton, A. (2005). Measuring poverty in a growing world, *Review of Economics and Statistics*, 87 (1): 1-19.

Dembo, R., Belenko, S., Childs, K., Wareham, J. and Schmeidler, J. (2009). Individual and community risk factors and sexually transmitted diseases among arrested youths: A two level analysis, *Journal of Behavioural Medicine*, 32: 303-316.

Desgrees du Lou, A. (2005). The couple and HIV/AIDS in Sub-Saharan Africa: Telling the partner, sexual activity and childbearing, *Population,* 60 (3): 179-198.

DeRose, K. (2009). *The Case for Contextualism: Knowledge, Skepticism and Context*, Volume I, Oxford; New York: Oxford University Press.

Diclemente, R., Crittenden, C., Rose, E., Sales, J., Wingood, G., Crosby, R. and Salazar, L. (2008). Psychosocial predictors of HIV-associated sexual behaviours and the efficacy of prevention interventions in adolescents at-risk for HIV infection: What works and what doesn't work? *Psychosomatic Medicine*, 70: 598-605.

Diez-Roux, A. (1998). Bringing context back to epidemiology: Variables and fallacies in multilevel analysis, *American Journal of Public Health*, 88 (2): 216-222.

Diez-Roux, A. and Aiello A. (2005). Multilevel analysis of infectious diseases, *The Journal of Infectious Diseases*, 191 (1): S25-S33.

Diez-Roux, A. (2010). Neighbourhoods and health, *Annals of the New York Academy of Sciences*, 1186(1): 125-145.

Djamba, Y. (2003). Social capital and premarital sexual activity in Africa: The case of Kinshasa, Democratic Republic of Congo, *Archives of Sexual Behaviour,* 32 (4): 327-337.

Dragone D. and Savorelli L. (2012). Thinness and obesity: a model of food consumption, health concerns and social pressure, Journal of Health Economics, 31: 243-256.

Donald, S. and Newey, W. (2001). Choosing the number of instruments, *Econometrica*, 69 (5): 1161-1191.

Dunkle, K., Stephenson, R., Karita, E., Chomba, E., Kayitenkore, K., Vwalika, C., Greenberg, L. and Allen, S. (2008). New heterosexually transmitted HIV infections in married or cohabiting couples in urban Zambia and Rwanda: An analysis of survey and clinical data, *Lancet,* 371 (9631): 2148-2150.

Durlauf, S. and Ioannides, Y. (2010). Social interactions, *Annual Reviews of Economics*, 2 (45): 451-478.

Elbers, C., Lanjouw, J. and Lanjouw, P. (2003). Micro-level estimation of poverty and inequality, *Econometrica*, 71 (1): 355-364.

Elster, J. (1989). Social norms and economic theory, *The Journal of Economic Perceptive*, 3 (4): 99-117.

Epstein, H. (2010). The mathematics of concurrent partnerships and HIV: A commentary on Lurie and Rosenthal, 2009, *AIDS Behaviour*, 14: 29-30.

Etile, F. (2007). Social norms, ideal body weight and food attitudes, *Health Economics*, 16: 945-966.

Ethier, K., Kershaw, T., Lewis, J., Milan, S., Nicolai, L. and Ickovics, J. (2006). Self-esteem, emotional distress and sexual behaviour among adolescent females: Inter-relationships and temporal effects, *Journal of Adolescent Health*, 38: 268-274.

Fenton, K., Johnson, A. and McManus, S. (2001). Measuring sexual behaviour: Methodological challenges in survey research, *Sexually Transmitted Infections,* 77: 84-92.

Ferguson, A., Pere, M., Morris, C., Ngugi, E. and Moses, S. (2004). Sexual patterning and condom use among a group of HIV vulnerable men in Thika, Kenya, *Sexually Transmitted Infection*, 80: 435-439.

Filmer, D. and Prichett, L. (2001). Estimating wealth effects without expenditure data-or tears: An application to education enrolment in states of India, *Demography*, 38 (1): 115-132.

Finn, A., Franklin, S., Keswell, M., Leibbrandt, M. and Levinsohn, J. (2009). National income dynamic survey: Expenditure, Report on NIDS Wave.

Fletcher, J. (2010). Social interaction and smoking: Evidence using multiple student cohort, instrumental variables and school fixed effects, *Health Economics*, 19: 466-484.

Gabrysch, S., Edwards, T. and Glynn, J. (2008). The role of context: Neighbourhood characteristics strongly influence HIV risk in young women in Ndola, Zambia, *Tropical Medicine and International Health*, 13 (2): 162-170.

Garcia, S., Harou, P., Montagne, C. and Stenger, A. (2009). Models for sample selection bias in contingent valuation: An application to forest biodiversity, *Journal of Forest Economics*, 15: 59-78.

Gausset, Q. (2001). AIDS and cultural practices in Africa: The case of the Tonga (Zambia), *Social Science and Medicine*, 52: 509-518.

Geary C., Burke H., Neupane S., Catelnau L. and Brown J. (2006). Does MTV reach an appropriate audience for HIV prevention message? Evidence from MTV viewership data in Nepal and Brazil, Journal of Health Communication: International Perspective, 11 (7): 665-681.

Geronimus, A. (1996). Black/White Differences in the relationship of maternal age to birthweight: A population-based test of the weathering hypothesis, *Social Science and Medicine*, 42 (4): 589-597.

Geronimus, A., Hicken, M., Keene, D., and Bound, J. (2006). 'Weathering' and age patterns of allostatic load scores among Blacks and Whites in the United States, *American Journal of Public Health*, 96 (5): 826-833.

Gillespie, G., Greener, R., Whitworth, J. and Whiteside, A., (2007). Poverty, HIV and AIDS: Vulnerability and impact in Southern Africa, *AIDS*, 21 (7): S1-S104.

Glick, P. and Sahn, D. (2007). Changes in HIV/AIDS knowledge and testing behaviour in Africa: How much and for whom? *Journal of Population Economy*, 20: 383-422.

Galavotti, C., Cabral, R., Lansky, A., Grimley, D., Riley, G. and Prochaska, J. (1995). Validation of measures of condom and other contraceptive use among women at high risk for HIV infection and unintended pregnancy, *Health Psychology*, 14 (6): 570-578.

Giorgi, G., Pellizari, M. and Redaelli, S. (2010). Identification of social interactions through partially overlapping peer groups, *American Economic Journal: Applied Economics*, 2 (2): 241-275.

Grant, M. (2008). Children's school participation and HIV/AIDS in rural Malawi, *Demographic Research*, 19 (45): 1603-1634.

Greene, W. (2002). *Econometric Analysis*, 5th Edition. Upper Saddle River, New Jersey: Prentice-Hall.

Gregson, S., Todd, J. and Zaba, B. (2009). Sexual behaviour change in countries with generalised hiv epidemics?, Evidence from population cohort studies in sub-Saharan Africa, *Sexually Transmitted Infections,* 85: i1-i2.

Grimm, M., Harttgen, K., Klassen, S. and Misselhorn, M. (2006). A human development index by income groups, Human Development Report 2006, UNDP.

Grimm, M. and Gunther, I. (2007). Growth and poverty in Burkina Faso: A reassessment of the paradox, *Journal of African Economies*, 16 (1): 70-101.

Grimm, M., Harttgen, K., Klassen, S. and Misselhorn, M. (2008). Human development index by income groups, *World Development*, 36 (12): 2527-2546.

Grimm, M., Harttgen, K., Klassen, S., Misselhorn, M., Munxi, T. and Smeeding, T. (2009). Inequality in human development: An empirical assessment of thirty-two countries, Working Paper No. 471.

Grosse, M., Klasen, S. and Spatz, J. (2005). Creating national poverty profiles and growth incidence curves with incomplete income or consumption expenditure data: An application to Bolivia, Institute for Economic Research (IAI) Discussion Paper Number 129, University of Gottingen.

Groenewold, G., Horstman, R. and Bruijn, B. (2004). Gender and the role of men in reproductive health: Applications in studies on HIV sexual risk-behaviour in Zambia, safe motherhood in Nepal, Netherlands Interdisciplinary Demographic Institute.

Grodner, A. and Kniesner, T. (2005). An empirical model of labour supply with social interactions: Econometric issues and tax policy implications, Centre for Policy Research, Working Paper No. 69.

Grodner, A. and Kniesner, T. (2006). Social interactions in labour supply, *Journal of European Economic Association*, 4 (6): 1226-1248.

Grodner, A., Kniesner, T. and Bishop J. (2011). Social interactions in labour markets, Discussion paper series, IZA DP No. 5934.

Gujarati, D. (2004). *Basic Econometrics*, 4th Edition, McGraw-Hill /Irwin Companies, New York.

Haram, L. and Yamba, B. (2009). Dealing with uncertainty in contemporary African lives, Nordiska Afrikainstitutet: Printed in Sweden by GML Print.

Heckman, J. (1979). Sample selection bias as a specification error, *Econometrica*, 47: 153-61.

Helleringer S. and Kohler H. (2005). Social networks, risk perceptions and changing attitudes towards HIV/AIDS: New evidence from a longitudinal study using fixed-effects estimation, *Population Studies*, 59 (3): 265-282.

Hendlin Y., Anderson S. and Glantz S. (2009). 'Acceptable rebellion': marketing hipster aesthetics to sell Camel cigarettes in the US, Tobacco Control, 19: 213-222.

Hennessy, M., Bleakley, A., Fishbein, M. and Jordan, A. (2008). Validating an index of adolescent sexual behaviour using psychosocial theory and social trait correlates, *AIDS Behaviour*, 12: 321-331.

Henry, D. (2005). Social thought and commentary: The legacy of the tank: The violence of peace, *Anthropological Quarterly*, 78 (2): 443-456.

Holschneider, S. and Alexander, C. (2003). Social and psychological influences on HIV preventive behaviours of youth in Haiti, *Journal of Adolescent Health*, 33: 31-40.

Hounton, S., Carabin, H. and Henderson, N. (2005). Towards an understanding of barriers to condom use in rural Benin using the health belief model: A cross sectional survey, *BMC Public Health,* 5 (8): 1-8.

Horton, R. and Pam, D. (2008). Putting prevention at the forefront of HIV/AIDS, *Lancet*, 372: 421-422.

Howe L., Gabrysch S. and Huttly S. (2009). Is the wealth index a proxy for consumption expenditure? A systematic review, Journal of Epidemiology Community Health, 63: 871-880.

Huebner, A. and Howell, L. (2003). Examining the relationship between adolescent sexual risk-taking and perceptions of monitoring, communication, and parenting, *Journal of Adolescent Health*, 33: 71-78.

James, S., Reddy, P.; Ruiter, R., McCauley, A. and van den Borne, B. (2006). The impact of an HIV and AIDS life skills program on secondary school students in KwaZulu-Natal South Africa, *AIDS Education and Prevention,* 18 (4): 281-294.

Jesko, H., Lanjouw, J., Lanjouw, P. and Poggi, J. (2000). Combining census and survey data to trace the spatial dimensions of poverty: A case study of Ecuador, *World Bank Economic Review*, 14 (1): 147-165.

Jipguep, M., Sanders-Philips, K. and Cotton, L. (2004). Another look at HIV in African American Women: The impact of psychosocial and contextual factors, *The Journal of Black Psychology*, 30 (3): 366-385.

Kazianga, H. (2005). HIV/AIDS prevalence and the demand for safe sexual behaviour: Evidence from West Africa, Columbia University.

Kelly J., Larence J., Stevenson Y., Hauth A., Kalichman S., Diaz Y., Brasfield T., Koob J. and Morgan M. (1992). Community AIDS/HIV risk reduction: The effects of endorsements by popular people in three cities, American Journal of Public Health, 82 (11): 1483-1489.

Kelly J., Murphy D., Sikkeme K. and Kalichman S. (1993). Psychological interventions to prevent HIV infection are urgently needed, American Psycologist, 48 (10): 1023-1034.

Kennedy, S., Nolen, S., Applewhite, J., Pan, Z., Shamblen, S. and Vanderhoff, K. (2007). A quantitative study on the condom-use behaviours of eighteen- to twenty-four–year-old urban African American Males, *AIDS Patient Care and STDs*, 21 (5): 306-320.

Kimuna, S. and Djamba, Y. (2005). Wealth and extramarital sex among men in Zambia, *International Family Planning Perspectives*, 31 (2): 83-89.

Kohler H., Behrman J. and Watkins S. (2007). Social networks and HIV/AIDS perceptions, *Demography*, 44 (1): 1-33.

Kolenikov, S. (2009a). Confirmatory factor analysis: confa, *Stata Journal*, 9 (3): 329-373.

Kolenikov, S. and Angeles G. (2009b). Socioeconomic status measurement with discrete proxy variables: Is Principal Component Analysis a Reliable Answer?, *Review of Income and Wealth*, 55 (1): 128-165.

Kongnyuy, E. and Wiysonge, C. (2007). Alcohol use and extramarital sex among men in Cameroon, *BMC International Health and Human Rights*, 7 (6): 1-7.

Lam, D., Ardington, C., Branson, N., Case, A., Leibbrandt, M., Menendez, A., Seekings, J. and Sparks, M. (2008). *The Cape Area Panel Study: A Very Short Introduction to the Integrated Wave 1-2-3-4 Data,* The University of Cape Town.

Lamptey, P., Johnson, J. and Khan, M. (2006). The global challenge of HIV and AIDS, *Population Bulletin*, 61 (1): 1-24.

Larson, A. (1989). Social context of HIV transmission in Africa: Historical and cultural bases of East and Central African sexual relations, *Reviews of Infectious Diseases*, 11 (5): 716-731.

Latkin, C., Williams, C., Wang, J. and Curry, A. (2005). Neighbourhood social disorder as a determinant of drug injection behaviours: A structural equation modelling approach, *Health Psychology*, 24 (1): 96-100.

Latkin, C., Donnell, D., Celentano, D., Aramrattna, A., Liu, T., Vongchak, T., Wiboonnatakul, K., Davis-Vogel, A. and Metzger, D. (2009). Relationships between social norms, social network characteristics, and HIV risk behaviours in Thailand and the U.S, *Health Psychology,* 28 (3): 323-329.

Lavy, V. and Schlosser, A. (2007). Mechanism and impacts of gender peer effects at school, NBER Working Paper No. 13292.

Li, X., Feigelman, S. and Stanton, B. (2000). Perceived parental monitoring and health risk behaviours among urban low-income African-American children and adolescents, *Journal of Adolescent Health*, 27: 43-48.

Logan, T., Cole, J. and Leukefeld, C. (2002). Women, sex, and HIV: Social and contextual factors, meta-analysis of interventions, and implications for practice and research, *Psychological Bulletin*, 128 (6): 851-885.

Lurie, M. and Rosenthal, S. (2010a). Concurrent partnerships as a driver of the HIV epidemic in Sub-Saharan Africa?, The evidence is limited, *AIDS Behaviour*, 14: 17-24.

Lurie, M. and Rosenthal, S. (2010b). The concurrency hypothesis in Sub-Saharan Africa: Convincing empirical evidence is still lacking. Response to Mah and Halperin, Epstein, and Morris, *AIDS Behaviour*, 14: 34-37.

McDonald, J. and Moffit, R. (1980). The uses of Tobit analysis, *The Review of Economics and Statistics*, 62: 318-312.

Macintyre, S., Ellaway, A. and Cummins, S. (2002). Place effects on health: How can we conceptualise, operationalise and measure them? *Social Science and Medicine*, 55: 125-139.

Magnani, R., Seiber, E., Gutierrez, E. and Vereau, D. (2001). Correlates of sexual activity and condom use among secondary-school students in urban, *Studies in Family Planning*, 32 (1): 53-66.

Mah, T. and Halperin, D. (2010a). Concurrent sexual partnerships and the HIV epidemics in Africa: Evidence to move forward, *AIDS Behaviour*, 14: 11-16.

Mah, T. (2010b). Prevalence and correlates of concurrent sexual partnerships among young people in South Africa, *Sexually Transmitted Diseases,* 37 (2): 105-108.

Mah, T. and Halperin D. (2010c). The evidence for the role of concurrent partnerships in Africa's HIV epidemics: A response to Lurie and Rosenthal, *AIDS Behaviour*, 14: 25-28.

Maharaj, P. and Cleland, J. (2004). Condom use within marital and cohabitating partnerships in KwaZulu-Natal, South Africa, *Studies in Family Planning*, 35 (2): 116-124.

Mahdavi P. (2009). But what if someone sees me?: Women, risk and the aftershock of Iran's sexual revolution, Journal of Middle East Women's Studies, 5 (2): 1-22.

Mannberg, A. (2010). Risky sex in a risky world: Sexual behaviour in a HIV/AIDS environment, *Scandinavian Journal of Economics*, 114 (2): 296-322.

Mannberg, A. (2010). Risky sex in a risky world: Sexual behaviour in a HIV/AIDS environment, *Scandinavian Journal of Economics*, 114 (2): 296-322.

Mannberg A. (2012). Risk and rationalization: The role of affect and cognitive dissonance for sexual risk taking, *European Economic Review*, 56: 1325-1337.

Manski, C. (1993). Identification of endogenous social effects: The reflection problem, *The Review of Economic Studies*, 60 (3): 531-542.

Manski, C. (2000). Economic analysis of social interactions, *The Journal of Economic Perspectives*, 14 (3): 115-136.

Manski, C. (2003). Identification problems in the social sciences and everyday life, *Southern Economic Journal*, 70 (1): 11-21.

Manski, C. and Mayshar, J. (2003). Private incentives and social interactions: Fertility puzzles in Israel, *Journal of the European Economic Association*, 1 (1): 181-211.

Mausbach, B., Semple, S., Strathdee, S., Zians, J. and Patterson, T. (2007). Efficacy of a behavioural intervention for increasing safer sex behaviours in HIV-positive MSM Methamphetamine users: Results from the EDGE study, *Drug and Alcohol Dependence*, 87: 249-257.

Mcqueen, D. (1996). The research for theory in health behaviour and health promotion, *Health Promotion International*, 11 (1): 27-32.

Moffit, R. (2001). Policy intervention, low-level equilibrium and social interaction, Social Dynamics, Chapter 3, MIT Press.

Moffitt, R. (2003). Policy interventions, low-level equilibria and social interactions, *Journal of Applied Econometrics,* 18 (5): 45-82.

Montgonery, M. and Casterline, J. (1996). Social learning, social influence and new models of fertility, *Population and Development Review*, 22: 151-175.

Montgonery, M. and Casterline, J. (1998). Social network and the diffusion of fertility control. Assessed 20 May, 2011. http://www.demogr.mpg.de/papers/workshops/010221_paper04.pdf

Morgan, P. and Niraula, B. (1995). Gender inequality and fertility in two Nepali Villages, *Population and Development Review*, 21 (3): 541-561.

Morisky, D., Stein J., Chiao C., Ksobiech K. and Malow R. (2006). Impact of a social influence intervention on condom use and sexually transmitted infections among establishment-based female sex workers in the Philippines: A multilevel analysis, *Health Psychology*, 25 (5): 595-603.

Morris, M. (1997). Sexual network and HIV, *AIDS*, 11 (Supplement A): 209-216.

Morris, M. (2010). Barking up the wrong evidence tree. Comment on Lurie and Rosenthal, "Concurrent partnerships as a driver of the HIV epidemic in Sub-Saharan Africa? The evidence is limited", *AIDS Behaviour*, 14: 31-33.

Moulton R. (1990). An illustration of a pitfall in estimating the effects of aggregated variables on micro units, The Review of Economics and Statistics, 72 (2): 334-338.

Munshi, K. and Myaux, J. (2006). Social norms and fertility transition, *Journal of Development Economics*, 80 (1): 1-38.

Nattrass, N. (2009). Poverty, sex and HIV, *AIDS Behaviour*, 13: 833-840.

Neubourg C. and Vendrik M. (1994), An extended rationality of social norms in labour supply, Journal of Economic Psychology, 15: 93-126.

Noar, S. and Zimmerman, R. (2005). Health behaviour theory and cumulative knowledge regarding health behaviours: Are we moving in the right direction? *Health Education Research*, 20 (3): 275-290.

Noar, S., Chabot, M. and Zimmerman, R. (2008). Applying health behaviour theory to multiple behaviour change: Considerations and approaches, *Preventive Medicine*, 46: 275-280.

Njau, B., Mtweve, S., Barongo, L., Manongi, R., Chugulu, J., Msuya, M, Mwampeta, S., Kiwale, B., Lekule, J. and Jalipa, H. (2006). The influence of peers and other significant persons on sexuality and condom-use among young adults in Northern Tanzania, *African Journal of AIDS Research,* 5 (3): 33-40.

Obermeyer, C. (2005). Refraining research on sexual behaviour and HIV, *Studies in Family Planning*, 36 (1): 1-12.

Okten, C. (2004). Social networks and credit access in Indonesia, *World Development*, 32 (7): 1225-1246.

Oster, E. (2007). Routes of infection: Exports and HIV incidence in Sub-Saharan Africa, University of Chicago and NBER.

Oster, E. (2009). HIV and sexual behaviour change: Why not Africa?, Preliminary Draft, University of Chicago and NBER.

Oster, E. (2012). HIV and sexual behaviour change: Why not Africa?, *Journal of Health Economics*, 31: 35-49.

Painter, J., Borba, C., Hynes, M., Darren, Mays, D. and Glanz, K. (2008). The use of theory in health behaviour research from 2000 to 2005: A Systematic Review, *ann. behav. med.*, 35: 358-362.

Peterman, T., Lindsey, C. and Selik, R. (2005). This place is killing me: A comparison of countries where the incidence of AIDS increased the most and the latest, *Journal of Infectious Diseases*, 191 (1): S123-S126.

Pinkerton S., Holtgrave D., DiFranceisco W., Stevenson Y. and Kelly J. (1998). Cost-effectiveness of a community level HIV risk reduction intervention, American Journal of Public Health, 88: 1239-1242.

Ragnarsson, A., Townsend, L., Thorson, A., Chopra, M. and Ekström, A. (2009). Social networks and concurrent sexual relationships: A qualitative study among men in an urban South African community, *AIDS Care*, 21 (10): 1253-1258.

Ragnarsson, A., Townsen, L., Ekström, A., Chopra, M. and Thorson, A. (2010). The construction of an idealised urban masculinity among men with concurrent sexual partners in a South African township, *Global Health Action*, 3 (5092): 1-7.

Reniers, G. and Watkins, S. (2009). Polygamy and the spread of HIV in Sub-Saharan Africa: A case of Benign concurrency, California Centre for Population Research, On-Line Working Paper Series.

Rossem, R., Meekers, D. and Akinyemi, Z. (2001). Consistent condom use with different types of partners: Evidence from two Nigerian surveys, *AIDS Education and Prevention*, 13 (3): 252-267.

Rossier, C., and Bernadi, L. (2009). Social interaction effects on fertility: Intentions and behaviours, *European Journal of Population*, 25: 467-485.

Rosenzweig, M. and Wolpin, K. (2000). Natural 'natural experiments', *Journal of Economic Literature*, 38: 827-874.

Rutsein, S. and Johnson, K. (2004). The DHS wealth index, *DHS Comparative Reports No. 6*, Calverton, Maryland: ORC Macro.

Sahn, D. and Stifel, D. (2000). Poverty comparison over time and across countries in Africa, *World Development*, 28 (12): 2123-2155.

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

Salvini S. and Vignoli D. (2011). Things change: Women's and men's marital disruption dynamics in Italy during a time of social transformation, 1970-2003, Demography, 24 (5): 145-174.

Smith, R. and Blundell, W. (1986). An exogeneity test for a simultaneous equation Tobit model with an application to labor supply, *Econometrica*, 54: 679-686.

Somer,s C. and Paulson, S. (2000). Students' perceptions of parent-adolescent closeness and communication about sexuality: Relations with sexual knowledge, attitudes, and behaviours, *Journal of Adolescent Health*, 23 (5): 629-644.

Staiger, D. and Stock, J. (1997). Instrumental variables regression with weak instruments, *Econometrica*, 65 (3): 557-586.

Stillwaggon, E. (2002). HIV/AIDS in Africa: Fertile terrain, *Journal of Development Studies*, 38 (6): 1-22.

Takyi, B. (2003). Religion and women's health in Ghana: Insights into HIV/AIDS preventive and protective behaviour, *Social Science and Medicine*, 56: 1221-1234.

UNAIDS, (1999). Sexual behavioural change for HIV: Where have theories taken us?, UNAIDS Best Practice Collection, Geneva, Switzerland.

UNAIDS, (2007). UNAIDS expert consultation on behaviour change, 25-26 September 2006, Joint United Nations Programme on HIV/AIDS, International Conference Centre, Geneva, Switzerland.

UNAIDS, (2008). 2008 report on global AIDS epidemic - Preventing new HIV infections: The key to reversing the epidemic, Geneva, Switzerland.

UNAIDS, (2009). AIDS epidemic update, Joint UNAIDS and WHO, Geneva, Switzerland.

UNAIDS, (2010). Joint action for results: UNAIDS Outcome Framework, 2009-2011, Joint United Nations Programme on HIV/AIDS Geneva, Switzerland.

UNAIDS, (2010b). Global report: UNAIDS Report on the Global AIDS Epidemic 2010, Joint United Nations Programme on HIV/AIDS, Geneva, Switzerland.

UNDP, (2003). Millennium development goals: A Compact Among Nations to end Human Poverty, Human Development Report 2003, United Nations Development Programme.

Varga, C. (1997). The condom conundrum: Barriers to condom use among commercial sex workers in Durban, South Africa, *African Journal of Reproductive Health*, 1 (1): 74-88.

Varga, C. (2001). Coping with HIV/AIDS in Durban's commercial sex industry, *AIDS Care*, 13 (3): 351-365.

Vancouver, J., Thompson, C., Tischner, E. and Putka, D. (2002). Two studies examining the negative effect of self-efficacy on performance, *Journal of Applied Psychology*, 87 (3): 506-516.

Vendrick, M. (2003). Dynamics of a household norm in female labour supply, *Journal of Economic Dynamics and Control*, 27, 823-841.

Vyas S. and Kumaranayake L. (2006). Constructing socio-economic status indices: how to use principal components analysis, Health Policy and Planning, 21 (6): 459-468.

Visser, M., Makin, J. and Lehobye, K. (2006), Stigmatizing attitudes of the community towards people living with HIV/AIDS. *Journal of Community and Applied Social Psychology*, 16: 42-58.

Voeten, H., Egesah, O. and Habbema, D. (2004). Sexual behaviour is more risky in rural than in urban areas among young women in Nyanza, Kenya, *Sexually Transmitted Diseases,* 31 (8): 481-487.

Weinstein, N. (1993). Testing four competing theories of health-protective behaviour, *Health Psychology*, 12 (4): 324-333.

Weir, S., Tate J., Zhusupov, B. and Boerma, J. (2004). Where the action is: Monitoring local trends in sexual behaviour, *Sexually Transmitted Infections,* 80: ii63-ii68.

Wellings, K., Collumbien, M., Slaymaker, E., Singh, S., Hodges, Z., Patel, D. and Bajos, N. (2006). Sexual behaviour in context: A global perspective, *Lancet,* 368: 1706-1728.

Williams R. (2000). A note on robust variance estimation for clustered-correlated data, *Biometrics*, 56: 645-646.

Wilson, D. and Halperin, D. (2008). "Know your epidemic, know your response": A useful approach, if we get it right, *Lancet*, 372: 423-426.

Wingood G. and DiClemente R. (2000). Application of the theory of gender and power to examine HIV related exposure, risk factors and effective interventions for women, *Health Education and Behaviour*, 27 (5): 539-565.

Wojcicki J. and Malala J. (2001). Condom use, power and HIV/AIDS risk: Sex workers bargain for survival in Hillbrow/Joubert Park/Berea, Johannesburg, *Social Science and Medicine,* 53 (1): 99-121.

World Health Organisation (WHO), (2008). Priority interventions: HIV/AIDS prevention, treatment and care in the health sector, Geneva, Switzerland.

World Health Organisation (WHO), (2010). World health statistics 2010, Geneva, Switzerland.

Wooldridge, J., (2002). *Econometric Analysis of Cross Section and Panel Data*, Cambridge, the MIT Press.

Wooldridge, J., (2003). *Introductory Econometrics: A Modern Approach*, 2nd Edition, Thomson South-Western.

Wooldridge J. (2003). Clustered-sample methods in applied econometrics, *The American Economic Review*, 93 (2): 133-138.

Appendix 3.A: Characteristics of the DHS Data

Table 3.A1: Overall Characteristics of the Selected DHS

	Kenya	Lesotho	Swaziland	Tanzania	Uganda	Zimbabwe	Total	East	South
Observation (n)	11,773	9,892	9,143	12,964	11,034	16,082	70,888	35,771	35,117
Observation (%)	16.6	14.0	12.9	19.0	15.6	22.7	100.0	50.5	49.5
• Age (mean)	28.5	28.6	27.3	28.2	28.2	27.8	28.1	28.4	27.9
• Gender (% women)	69.6	71.7	54.5	79.7	77.3	55.4	67.8	75.6	59.8
• Sexually active population (%)	83.2	82.8	77.1	82.9	84.4	76.8	81.1	83.5	78.6
• Intention, postpone sex- marriage (%)			90.0	92.6	84.7	89.1	89.3	89.2	89.4
• Education									
no education	13.5	7.3	8.2	22.0	17.4	3.1	11.8	17.8	5.6
primary	53.3	58.9	33.5	64.0	59.1	31.6	49.5	59.0	39.8
secondary	25.0	32.2	49.9	11.8	18.9	61.2	34.1	18.3	50.1
higher	8.21	1.7	8.48	2.16	4.58	4.08	4.73	4.9	4.55
• Marital status									
never married	34.4	38.0	56.5	28.2	26.9	36.7	36.0	29.8	42.2
married	53.3	49.1	28.6	56.4	49.7	50.7	49.0	53.3	44.5
living together	3.9	0.8	8.1	6.6	11.9	2.0	5.3	7.4	3.2
widowed	3.1	7.0	3.6	2.0	3.4	4.7	3.9	2.8	5.1
divorced	1.6	0.7	0.6	4.0	1.0	3.1	2.0	2.3	1.8
not living together	3.7	4.4	2.7	2.8	7.1	2.7	3.8	4.4	3.2
• Age at first marriage (mean)	20.2	19.7	22.9	18.9	18.3	20.3	19.7	19.1	20.6
• Age of first intercourse (mean)	16.9	18.1	17.9	17.1	16.5	18.5	17.5	16.8	18.2
• No. of partners in 12 month (%)									
mean	0.9	1.1	1.0	1.0	1.0	0.9	1.0	1.0	1.0
no partners	15.5	14.5	14.6	10.8	13.5	15.3	14.0	13.2	14.9
1 partner	79.1	71.4	76.1	81.4	79.4	79.2	78.1	80.0	76.1
2 and above partners	5.4	14.1	9.3	7.8	7.1	5.5	7.9	6.8	9.0
• No. of partners lifetime (%)									
mean		6.3	4.1	6.6	3.1	3.4	3.9	3.7	4.0
1 partner		15.3	26.8	18.6	39.3	45.3	35.7	35.5	35.8
2-10 partners		74.2	67.5	68.6	57.8	50.6	59.2	59.8	58.8
11-20 partners		6.7	4.2	7.7	2.1	2.7	3.5	3.1	3.7
21 and above partners		3.7	1.5	5.1	0.9	1.4	1.7	1.6	1.7
• Condom use, 'last partner'(%)	9.3	18.1	42.5	9.6	9.9	15.6	16.0	9.6	23.7
• Relationship, 'last partner' (%)									
spouse/cohabiting partner	79.2	65.0	53.6	81.6	83.4	81.4	75.7	81.4	69.4
boyfriend/fiancé	15.8	32.4	42.9	14.9	14.5	17.2	21.4	15.1	28.3
casual acquaintance	5.0	2.7	3.6	3.6	2.1	1.4	3.0	3.6	2.3
• Time (days) knew 'last partner'	679.8	882.9	92.4	153.8	526.5	42.1	387.9	443.7	350.5
• Age of 'last partner'									
younger	0.4	1.3	37.7	4.6	3.0	3.3	9.2	3.2	10.3
about the same age	45.7	32.5	5.5	11.7	7.4	3.3	7.9	16.3	6.3
less than 10 years older	49.6	45.7	51.3	71.9	81.3	74.5	67.8	71.5	67.1
10 or more years older	3.9	10.0	5.4	9.9	7.6	18.9	14.1	7.9	15.3
older, don't know difference	0.4	10.5	0.1	1.9	0.7	0.0	1.0	1.2	1.0
• Alcohol with 'last partner'									
did not consume alcohol			92.2		76.1	90.0	79.9	67.4	86.4
neither drunk, consumed alc.		6.8	0.4	35.1	7.5	0.7	4.2	10.6	0.9
respondent drunk only		22.2	3.0	8.8	2.6	4.1	4.1	3.3	4.6
partner drunk only		55.9	3.7	45.4	10.8	4.9	9.7	14.8	7.0
both drunk		15.1	0.8	10.8	2.9	0.3	2.1	3.8	1.2
• Knows PLWHA (% yes)	74.0	24.9	50.8	29.4	65.2	29.8	45.8	56.8	34.2
• Could get a condom (% yes)	77.2	74.6	87.0	92.2	63.4	89.0	81.9	78.8	84.8
• Heard of STD (% yes)	98.7	93.5	99.7	99.2	99.2	98.8	98.3	99.1	97.5
• Heard of HIV (% yes)	98.6	92.4	99.6	99.0	99.1	98.3	97.9	98.9	97.0
• AIDS: abstain (% yes)	93.5	82.4	94.2	90.6	88.2	84.6	88.6	90.7	86.6
• AIDS: use condom (% yes)	72.9	79.9	89.7	75.4	72.3	78.9	77.8	73.6	82.0
• AIDS: one sex partner (% yes)	94.2	85.7	92.5	90.0	91.0	84.0	89.2	91.6	86.7
• Get AIDS: mosquito (% yes)	15.4	36.1	18.9	13.6	27.3	18.8	21.0	18.6	23.5
• Get AIDS: sharing food (% yes)	10.4	27.9	11.5	12.2	12.5	12.6	14.3	11.8	17.4

Appendix 3.B: Characteristics of the Social Groups

Table 3.B1: Distribution of the social groups across Regions and Countries

	Lesotho							Swaziland				Zimbabwe						Tanzania													Kenya						Uganda						
Social Group	Maseru	Mohale's H.	Quthing	Qasha's N.	Mokhotlong	Thaba-tseka	• • •	Hhohho	Manzini	Shiselwen	Lubombo	Mashonala	Matebeleland	Masvingo	Harare	Bulawayo	• • •	Dodoma	Arusha	Kilimanja	Tanga	Morogoro	Iringa	Mbeya	Singida	Tabora	Kigoma	Shinyanga	Mwanza	• • •	Nairobi	Coast	Eastern	Nyanza	Rift valley	• • •	Central 1	Kampala	Eastern	North	West nile	• • •	Total
1	3	2	2	1	0	1		3	1	5	3	2	2	0	1	0		0	1	0	0	0	0	0	0	0	1	2	0		0	1	0	0	1		2	0	1	0	1		59
2	1	1	1	0	3	2		3	2	2	3	0	0	2	4	1		1	0	0	1	1	0	0	0	0	0	1	1		0	1	3	0	1		1	0	1	3	0		62
3	1	1	1	0	0	0		4	5	1	4	2	1	1	0	0		1	0	0	0	0	0	0	0	0	0	0	0		0	0	3	0	1		1	0	0	0	0		50
4	1	1	3	1	3	0		2	7	4	5	8	3	10	1	1		1	0	1	1	1	0	1	1	0	1	0	0		1	0	6	9	2		1	2	1	6	4		160
5	2	1	4	0	1	2		7	5	5	4	2	4	5	3	0		1	1	2	0	2	1	1	0	1	1	0	1		4	1	3	1	2		2	3	1	3	1		146
6	5	1	0	1	3	0		4	7	4	2	2	1	1	3	1		1	2	1	0	0	0	0	0	0	2	0	1		0	2	0	2	3		1	0	1	1	2		85
7	3	0	0	0	0	0		6	4	2	6	8	4	7	6	5		0	0	1	0	2	0	1	1	0	0	0	0		0	2	3	7	3		0	0	4	2	1		144
8	1	1	1	2	2	1		5	5	2	1	9	0	2	3	3		0	1	0	0	1	0	0	0	0	0	0	0		0	2	3	2	4		2	1	1	4	3		123
9	10	10	5	6	14	10		5	5	2	1	0	0	0	0	0		2	3	4	1	1	4	2	0	0	2	0	1		4	2	5	12	6		3	5	1	4	4		225
10	1	1	0	1	2	1		7	4	6	5	1	3	0	0	0		0	0	1	0	1	0	0	2	1	0	1	0		0	0	2	0	2		0	0	1	2	1		79
11	3	0	1	2	0	0		1	3	1	2	5	0	6	11	4		0	0	0	0	0	0	0	0	0	0	0	0		2	2	5	1	2		3	2	2	2	1		117
12	8	12	3	9	18	8		2	5	1	2	0	0	0	0	0		3	0	2	1	0	2	0	4	0	3	0	1		7	6	3	5	3		4	7	5	4	3		219
13	3	3	0	2	0	1		1	2	3	3	3	2	4	15	7		0	0	2	0	0	0	0	0	0	0	0	0		5	0	2	1	3		2	0	1	4	0		113
14	4	9	8	7	6	7		13	8	5	7	0	0	0	0	0		3	1	0	1	1	2	0	1	3	2	3	4		1	7	5	1	6		4	5	5	9	6		232
15	2	2	0	2	4	2		2	4	2	4	6	4	3	1	0		0	0	0	0	0	0	1	0	0	1	0	0		0	1	4	1	2		0	0	0	0	0		95
16	0	1	0	1	1	0		7	2	4	2	2	2	6	2	0		2	0	0	0	0	0	0	0	0	0	0	0		0	2	1	2	5		1	0	0	1	0		98
17	13	1	2	3	22	13		3	4	2	3	0	0	0	0	0		0	2	0	0	0	0	0	1	1	1	0	2		7	3	6	8	1		4	7	5	1	1		181
18	2	1	0	0	0	0		3	10	2	6	2	3	8	13	16		0	0	0	0	0	0	0	0	0	0	0	0		0	2	0	2	1		2	2	0	0	0		100
19	10	5	6	5	25	22		6	9	1	10	0	0	0	0	0		0	2	0	0	0	2	0	0	1	0	0	0		8	4	1	5	5		8	6	1	2	0		184
20	3	0	0	0	0	0		5	6	2	6	2	0	5	11	17		0	0	0	0	0	0	0	0	0	0	0	0		6	0	1	0	2		0	1	0	0	0		86
21	3	3	0	4	2	2		9	34	27	20	4	3	3	0	2		0	0	0	0	1	1	1	0	6	0	2	0		0	3	1	4	2		0	0	3	3	0		175
22	2	3	3	4	1	3		11	7	12	10	6	6	5	0	0		0	0	1	0	1	0	0	1	3	1	0	2		0	0	1	1	1		2	1	2	2	0		127
23	7	8	14	14	5	2		7	15	14	14	0	0	0	0	0		1	6	3	1	2	0	2	2	4	5	1	4		1	8	5	7	5		2	2	3	8	2		270
24	3	4	0	1	0	3		5	9	10	11	6	5	3	0	0		0	1	0	0	0	1	0	2	1	0	2	1		0	0	2	0	3		0	0	1	4	2		124
25	2	9	1	1	2	1		4	13	3	7	7	5	5	3	0		4	1	1	0	0	0	1	1	0	0	2	1		2	1	4	2	8		1	1	1	5	1		142
26	9	4	7	10	10	3		8	8	13	4	0	0	0	0	0		1	1	1	0	1	1	0	2	4	0	1	1		1	4	4	3	5		0	5	3	10	5		213
27	3	0	1	0	1	1		5	7	7	5	6	0	5	0	0		3	1	0	0	0	0	1	0	0	0	1	0		0	1	2	4	3		2	1	2	1	1		103
28	3	1	0	1	3	2		5	3	3	5	2	3	2	5	0		1	1	2	0	0	1	0	2	1	3	0	1		1	2	4	0	5		1	1	1	1	0		98
29	0	0	2	1	1	0		10	5	9	3	3	1	8	1	0		0	0	0	0	1	1	1	0	1	0	1	0		2	1	3	1	4		0	0	2	1	2		112
•																																											
•																																											
•																																											

Appendix 3.C: Construction of the Risky Sex Index using CFA

We use the STATA CFA programme to construct our risky sex index. We hypothesise that there is an unobserved factor, the *risky sex* index, which is related to our sexual behaviour variables (type of relationship, number of sex partners and age of sexual debut) as formulated in equation (3.C1).

$$y_{ij} = \mu_j + \sum_{k=1}^{m} \lambda_{jk}\xi_{ik} + \delta_{ij}, \quad j = 1, \ldots . p \qquad (3.C1)$$

Where; y_j, $j = 1, .. p$ are the observed variables (sexual behaviour variables); ξ_k, $k = 1, .. m$ is the number of factors, in our case 1 which represents risky sex, λ_{jk}, are the factor loading/regression coefficients and δ_j is the measurement error. The CFA model usually assumes multivariate normality and asymptotic robustness of the observed variables, implying that the log likelihood of the observation is:

$$\ln L\left(Y, \Sigma(\theta)\right) = -\sum_{i=1}^{n}\left[\frac{p}{2}\ln 2\pi + \frac{1}{2}\ln \Sigma(\theta) + \frac{1}{2}(y_i - \mu)'\Sigma^{-1}(\theta)(y_i - \mu)\right]$$

$$= -\frac{np}{2}\ln 2\pi + \frac{n}{2}\ln \Sigma(\theta) - \frac{1}{2}\operatorname{tr}\Sigma^{-1}(\theta)S \qquad (3.C2)$$

Where *S* is the maximum likelihood estimate of the covariance matrix, θ are the means. The maximum likelihood estimators (MLEs), $\hat{\theta}$, can be obtained by maximising equation (3.C2).

The asymptotic variance-covariance matrix in this case is the inverse of the observed information matrix/ negative hessian matrix. However, when the normality assumption is violated, as in the case of our observed sexual behaviour variables as they are categorical in nature, the quasi (MLE) still maintains some desirable properties: they are asymptotically normal and the inverse information matrix gives consistent estimates of the parameter variance (if the model is correctly specified and the error terms are independent). However, when the asymptotic robustness is violated then the variance-covariance matrix will be inconsistently estimated by the observed information matrix (Kolenikov, 2010).

An alternative method that has been proposed for estimating the variance-covariance matrix that will ensure asymptotic robustness in the case of non-normal (categorical) data is the Satorra-Bertler 'robust' standard error[74]. We follow Kolenikov, (2010), for this discussion. The maximum likelihood ratio statistics test when the assumptions are not violated is depicted in equation (3.C2) which has asymptotic χ^2 distribution and the degree of freedom is equal to the number of over-identifying model conditions $q = p^* - t$. However when the observed variables violate the assumption the (quasi-) likelihood ratio test statistic comprises a mixture of $\chi 2$ as depicted in equation (3.C3), implying that the (quasi-) likelihood ratio statistics have a non-standard distribution based on the sum of weighted χ^2 variables.

[74] CFA was conducted using STATA, we first had to install the CFA programme which we did by typing: **findit confa**

$$T = -2\left[\ln L\ (\mathbf{Y}, \Sigma\,\hat{\theta}) - \ln L\ (\mathbf{Y}, S)\right] \overset{d}{\rightarrow} \chi^2_q \qquad (3.C3)$$

$$T_u \overset{d}{\rightarrow} \sum_{j=1}^{df_u} \alpha_j X_j, \qquad X_j \sim i.i.d.\ \chi^2_1 \qquad (3.C4)$$

Where; α_j = eigenvalues of the matrix UΓ and $U = V - V\,\Delta(\Delta'V)^{-1}\,\Delta'V$. Hence Satora and Bertler (1994) conceptualised the Satterthwate-type corrections, which include the T_{sc} statistic that corrects the scale of the distribution and the T_{adj} statistic which corrects both the scale and degree of freedom. This is followed by an additional correction to the T statistic proposed by Yuan and Bentler (1997) represented by T_2 in equation (3.C7) (Kolenikov 2010).

$$T_{SC} = \frac{T}{\hat{c}},\ \text{referred to}\ \chi^2_{df_u},\ \text{where}\ \hat{U}\ \text{is U evaluated at}\ \theta\ \text{and}\ \hat{c} = \frac{1}{df_u}\,\mathrm{tr}\left[\hat{U}\hat{\Gamma}_N\right] \qquad (3.C5)$$

$$T_{adj} = \frac{\hat{d}}{\hat{c}}T,\ \text{referred to}\ \chi^2_{\hat{d}},\ \text{where d might be non} - \text{interger and}\ \hat{d} = \frac{\left(\mathrm{tr}\left[\hat{U}\hat{\Gamma}_N\right]\right)^2}{\mathrm{tr}\left[\left(\hat{U}\hat{\Omega}_N{}^2\right)\right]} \qquad (3.C6)$$

$$T_2 = T/\left(1 + \frac{T}{N}\right),\ \text{refered to}\ \chi^2\ \text{with}\ df_u\ \text{degree of freedom} \qquad (3.C7)$$

We performed the CFA with Satorra-Bertler 'robust' standard errors[75] on our observed sexual behaviour variables. Root mean squared error of approximation (RMSEA) which measures the absolute fit was conducted to evaluate the model fit, where RMSEA values of 0.05 and less and a confidence interval covering this range indicate a good fit. Our RMSEA values fall in this range therefore indicating a good fit. We also conducted the Tucker-lewis non-normed fit index (TLI) test and Bentler's comparative fit index (CFI) test where indices close to 0 show a poor fit and those close to 1 a good fit. Our TLI and CFI were close to 1 indicating a good fit. Once we had established that the model was a good fit we proceeded to obtain the factor scores. To do this we used the regression method[76], which obtains the estimates of the predicted factor scores by minimising the sum of the squared deviation of the factors from the true values. Table 3.C1 shows the output of the confirmatory factor analysis.

[75] Apart from Satorra-Bertler standard errors confa in STATA also provides the Huber Sandwich standard error test

[76] confa in STATA also offers the Bartlelt method as an alternative factor scoring method.

Table 3.C1: Risky Sex CFA Output

```
log likelihood = -152500.26                              Number of obs = 49070
------------------------------------------------------------------------------
             |              Satorra-Bentler
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Means        |
  age_firsex |    2.61561   .0049897   524.20   0.000     2.605831     2.62539
    rel_type |   1.271164    .002289   555.34   0.000     1.266677     1.27565
sex12_quater |   2.183534   .0026065   837.72   0.000     2.178425    2.188642
Loadings     |
ris~04052011 |
  age_firsex |          1          .        .       .            .           .
    rel_type |   2.374513    .286979     8.27   0.000     1.812045    2.936982
sex12_quater |   1.313658   .0861864    15.24   0.000     1.144736     1.48258
Factor cov.  |
risk~0405201 |   .0175215    .002512     6.98   0.000     .0125981    .0224448
Var[error]   |
  age_firsex |   1.204194   .0050729   237.38   0.000     1.194252    1.214137
    rel_type |    .158308   .0124246    12.74   0.000     .1339562    .1826598
sex12_quater |   .3031461   .0053907    56.24   0.000     .2925806    .3137116
-------------+----------------------------------------------------------------
R2           |
  age_firsex |     0.0143
    rel_type |     0.3842
sex12_quater |     0.0907
------------------------------------------------------------------------------
No degrees of freedom to perform the goodness of fit test
Test vs independence: LR = 2037.063    ; Prob[chi2( 3) > LR] = 0.0000
```

Table 3.C2: Distribution of Sexual Behaviour by Risky Sex Index Quartile

Risky sex quartiles: DHS	Age of sexual debut	Number of partners	Relationship type	Condom use	Kenya (%)	Tanzania (%)	Uganda (%)	Lesotho (%)	Swaziland (%)	Zimbabwe (%)
1	20.357	1.000	1.033	0.077	32.7	29.8	23.6	32.7	26.6	46.7
2	16.489	1.000	1.050	0.058	17.9	23.1	23.5	15.0	13.7	18.6
3	14.754	1.238	1.057	0.051	28.7	28.8	36.4	17.6	13.6	16.2
4	17.026	1.271	3.121	0.474	20.7	18.4	16.6	34.8	46.1	18.6

Relationship type: 1=spouse, 2=live-in-partner, 3=partner not live in, 4=casual acquaintance
Condom use: 0=used condom, 1=did not use condom

Appendix 3.D: Constructing the Expected Future Health Proxy

The appendix gives a description of how life expectancy table 3.D1, 3.D2 and 3.D3 was derived.

Table 3.D1: Life Expectancy, 2000s (calculated from post HIV/AIDS era mortalities & pattern 2)

Age	Kenya		Tanzania		Uganda		Swaziland		Zimbabwe		Lesotho	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
0	55.5	58.6	48.9	55.3	46.2	49.1	58.8	62.0	65.4	65.3	61.5	65.2
1	59.4	61.6	53.0	58.9	49.7	52.8	62.1	65.6	67.5	67.4	64.5	67.2
5	58.3	60.2	52.4	58.0	49.2	52.3	60.7	66.0	65.4	66.8	64.2	65.2
10	54.3	56.0	48.6	53.9	45.5	48.5	56.5	61.9	61.1	62.6	60.1	60.8
15	49.8	51.4	44.2	49.4	41.1	44.0	52.0	57.4	56.4	58.0	55.5	56.1
20	45.3	47.0	39.8	45.0	36.8	39.8	47.5	52.9	51.8	53.4	50.9	51.6
25	41.2	43.1	35.9	41.3	32.9	36.3	43.3	48.6	47.4	49.0	46.5	47.4
30	37.8	40.1	32.8	38.6	29.9	34.0	39.7	44.4	43.5	44.7	42.2	44.0
35	34.7	37.4	30.0	36.2	27.3	32.1	36.4	40.3	39.9	40.4	38.2	40.8
40	31.9	34.5	27.6	33.5	25.1	29.9	33.3	36.0	36.4	36.1	34.2	37.5
45	29.1	31.2	25.3	30.4	23.0	27.2	30.4	31.9	33.1	31.8	30.5	33.9
50	26.5	27.8	23.2	27.2	21.1	24.4	27.5	27.8	29.8	27.7	26.8	30.2
55	23.7	24.2	20.7	23.7	18.9	21.2	24.5	24.0	26.5	23.8	23.3	26.3
60	20.6	20.7	18.0	20.3	16.3	18.1	21.3	20.6	22.9	20.2	20.1	22.6
65	17.9	17.3	15.6	17.1	14.2	15.2	18.4	17.3	19.8	16.8	17.2	19.0
70	15.4	14.2	13.6	14.1	12.4	12.6	15.8	14.6	16.9	14.1	14.8	15.7
75	13.6	11.7	12.2	11.8	11.2	10.6	13.8	12.7	14.5	12.1	12.7	12.9
80	12.4	9.2	11.5	9.6	10.9	8.7	12.4	11.4	12.8	11.0	11.9	10.3
85	11.8	7.8	11.9	8.7	11.9	8.2	11.9	11.4	12.0	11.2	12.7	9.0
90	5.1	4.8	4.8	4.8	4.5	4.5	5.2	5.2	5.4	4.9	5.2	5.2
95	3.0	2.9	2.9	2.9	2.8	2.8	3.0	3.0	3.0	2.9	3.0	3.0
100	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 3.D2: Life Expectancy, 2000s (calculated from pre HIV/AIDS era mortalities & pattern 1)

Age	Kenya		Tanzania		Uganda		Swaziland		Zimbabwe		Lesotho	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
0	55.0	58.9	48.5	55.7	45.7	49.7	58.5	61.9	65.0	65.2	61.9	65.3
1	59.3	62.0	53.0	59.5	49.7	53.7	62.1	65.3	67.5	67.2	64.5	67.4
5	60.3	62.6	54.8	60.8	51.8	56.0	62.4	63.9	66.7	65.2	62.7	66.9
10	56.4	58.7	51.3	57.0	48.4	52.5	58.4	59.7	62.4	60.8	58.5	62.6
15	52.0	54.2	47.0	52.6	44.2	48.2	53.9	55.1	57.8	56.1	53.9	58.0
20	47.6	49.8	42.7	48.3	39.9	44.1	49.4	50.6	53.2	51.5	49.3	53.5
25	43.3	45.5	38.5	44.1	35.8	40.1	45.0	46.7	48.6	47.4	45.0	49.0
30	39.2	41.4	34.7	40.1	32.1	36.3	40.9	43.5	44.3	44.0	41.2	44.7
35	35.4	37.4	31.2	36.3	28.7	32.7	36.9	40.6	40.1	40.8	37.7	40.5
40	31.8	33.2	27.8	32.2	25.5	28.8	33.2	37.4	36.1	37.4	34.4	36.1
45	28.3	29.1	24.7	28.2	22.5	25.1	29.6	34.0	32.2	33.8	31.3	31.9
50	25.0	25.2	21.7	24.4	19.7	21.5	26.1	30.4	28.5	30.1	28.2	27.7
55	21.9	21.5	18.9	20.9	17.1	18.3	22.8	26.6	24.9	26.3	25.0	23.8
60	19.0	18.2	16.4	17.8	14.8	15.5	19.8	22.9	21.5	22.5	21.6	20.2
65	16.4	15.1	14.2	14.9	12.8	12.9	17.0	19.4	18.4	18.9	18.5	16.9
70	14.5	12.6	12.7	12.6	11.5	11.0	14.9	16.1	16.0	15.6	15.7	14.1
75	12.8	10.9	11.5	11.1	10.5	9.8	13.1	13.4	13.8	12.9	13.4	12.1
80	12.8	9.8	12.1	10.4	11.6	9.6	12.9	10.7	13.2	10.3	11.6	11.1
85	14.7	9.9	15.2	11.2	15.5	11.0	14.6	9.2	14.4	8.9	10.6	11.4
90	5.1	4.8	4.8	4.8	4.5	4.5	5.2	5.2	5.4	4.9	5.2	5.2
95	3.0	2.9	2.9	2.9	2.8	2.8	3.0	3.0	3.0	2.9	3.0	3.0
100	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 3.D3: Life Expectancy, 1980s (calculated from *pre HIV/AIDS era mortalities & pattern 1*)

	Kenya		Tanzania		Uganda		Zimbabwe		Botswana	
	Male	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**
0	55.4	75.5	41.5	47.7	40.6	57.0	52.3	75.5	58.5	57.7
1	57.8	79.7	47.5	54.5	48.2	59.0	55.1	79.7	59.8	65.2
5	57.9	78.6	50.9	57.9	52.2	59.3	55.8	78.6	59.0	67.0
10	53.9	74.2	47.7	54.5	49.1	55.4	52.0	74.2	54.8	63.1
15	49.5	69.5	43.7	50.3	45.1	50.9	47.7	69.5	50.3	58.7
20	45.0	64.8	39.5	46.2	40.9	46.5	43.2	64.8	45.7	54.4
25	40.6	60.2	35.5	42.3	36.9	42.3	38.9	60.2	41.2	50.1
30	36.5	55.7	31.9	38.6	33.4	38.2	34.9	55.7	37.0	46.1
35	32.6	51.2	28.7	35.0	30.2	34.3	31.2	51.2	33.0	42.1
40	29.0	46.6	25.7	31.1	27.1	30.2	27.7	46.6	29.2	37.9
45	25.5	42.1	22.9	27.4	24.3	26.2	24.4	42.1	25.5	33.8
50	22.3	37.6	20.2	23.8	21.6	22.3	21.3	37.6	22.1	29.8
55	19.2	33.3	17.6	20.5	19.0	18.8	18.4	33.3	18.9	26.1
60	16.4	29.2	15.5	17.7	16.7	15.8	15.8	29.2	16.0	22.7
65	14.0	25.2	13.5	15.0	14.6	12.9	13.5	25.2	13.5	19.5
70	12.3	21.5	12.2	12.9	13.2	10.8	12.0	21.5	11.8	16.7
75	11.0	18.1	11.1	11.5	12.0	9.5	10.8	18.1	10.5	14.5
80	11.7	15.1	11.9	10.8	12.4	9.3	11.7	15.1	11.4	12.8
85	15.4	12.5	15.3	11.3	14.9	11.0	15.5	12.5	15.8	11.8
90	4.6	7.0	4.7	5.0	5.0	4.3	4.5	7.0	4.4	5.9
95	2.8	3.6	2.9	3.0	3.0	2.8	2.8	3.6	2.8	3.2
100	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

The life expectancy tables 3.D1, 3.D2 and 3.D3 are calculated using the INDEPTH MLT. The INDEPTH tables are based on the *Brass relationship life table system* (BRLTS). The BRLTS[77] assumes that any two mortality patterns can be related through the relationship depicted in equation (3.D1):

$$Y_x = \alpha + \beta Y_{sx} \equiv \text{logit}\,(l_x) = \alpha + \beta\text{logit}\,(l_{sx}) \quad (3.D1)$$

Where l_x and l_{sx} are the survivorship probability of the two mortality patterns, with l_{sx} being the *standard survivorship pattern* (INDEPTH 2004). Given the above relationship one can determine a complete life table of any population from a *standard life table*. INDEPTH tables have generated *standard life tables* for SSA by gender. These tables have been further classified into *pattern 1* and *pattern 2*. *Pattern 1* is based on populations of West and Central Africa, while *pattern 2* depicts those of East and Southern Africa. Table (3.D4) outlines the infant and under five mortality rates that were used to calculate life tables for each country and thereafter life expectancy. The infant and under-five mortality are lower in the 1980s than they are currently, except for Tanzania and Uganda.

[77] A relational model of mortality consists of a life table representing the mortality conditions of a standard population, also referred to as a standard life table, along with a mathematical function, which expresses the relationship between the standard mortality schedule and that of any other population (INDEPTH 2004, p. 6).

Table 3.D4: DHS Infant and Under-five Mortality by Gender and Country

	Gender	Infant mortality ($_1q_0$)	Under-five mortality ($_5q_0$)
Post HIV/AIDS Era			
• Kenya (2003)	Male	84	122
	Female	67	103
• Lesotho (2004)	Male	89	109
	Female	78	95
• Swaziland (2006)	Male	80	108
	Female	73	103
• Tanzania (2004)	Male	83	135
	Female	82	130
• Uganda (2006)	Male	98	165
	Female	74	132
• Zimbabwe (2005)	Male	51	71
	Female	48	68
Pre HIV/AIDS Era			
• Kenya (1989)	Male	63	97
	Female	54	86
• Botswana (1988)	Male	46	64
	Female	31	47
• Tanzania (1991)	Male	104	160
	Female	95	147
• Uganda (1988)	Male	111	196
	Female	101	178
• Zimbabwe (1988)	Male	63	91
	Female	50	79

To outline the procedure that was used to calculate 3.D1, 3.D2 and 3.D3 we use Kenyan women (2003) as an example and follow the INDEPTH (2004) methodology.

Using infant (67) and under-five mortality (103) we first calculate the *estimated survival function* at age zero and age five by subtracting the mortality rates from one, naming them l_1 and l_5 respectively; $l_1 = 1 - \frac{67}{1000} = 0.9330$ and $l_2 = 1 - \frac{103}{1000} = 0.8970$. We then obtain the *standard survival function* from the *standard life table* at age zero (l_{s1}) and age five (l_{s5}) from INDEPTH tables; l_{s1}= 0.9224 and l_{s5}=0.8766, we then calculate α and β using the following formula:

$$\beta = \frac{\text{logit}\,(l_5) - \text{logit}\,(l_1)}{\text{logit}\,(l_{s5}) - \text{logit}\,(l_{s1})} = 0.9118 \approx 0.9 \text{ and } \alpha = \text{logit}\,(l_1) - \frac{[\text{logit}\,(l_{s1})][\,\text{logit}\,(l_5)] - \text{logit}\,(l_1)}{\text{logit}\,(l_{s5}) - \text{logit}\,(l_{s1})} = -0.1883 \approx -0.2$$

Thereafter we use α and β to select the relevant *estimated survival function* (l_x) from INDEPTH tables, which we depict in column 5 in table (3.D5). The values in this column (l_x) although derived from the INDEPTH tables, can be calculated from the Brass formula in equation (3.D2) which is derived from equation (3.D1).

$$l_x = \frac{1}{1 + e^{2\alpha + \beta \ln\left(\frac{1 - l_{sx}}{l_{sx}}\right)}} \quad (3.D2)$$

Because we only have empirical values of l_1 and l_5, (infant and under-five mortality rate), and not mortality rates of each age group, we need to factor this into our l_x, (the values in column 5). The

INDEPTH tables has developed *correctional factor* (γ_x) depicted in equation (3.D3) to minimise the error caused by the difference between observed l_x, and the estimated l_x as a result of using only l_1 and l_5:

$$\gamma_x = \frac{1}{l_x\left[1+e^{2\alpha+\beta\ln\left(\frac{1-l_{sx}}{l_{sx}}\right)}\right]} \tag{3.D3}$$

We therefore select γ_x corresponding to $\beta = 0.8$ from the INDEPTH tables (column 6 of table 3.D5), we divide this with l_x in column 5, to give us column 7, which in essence is a product of equation (3.D4), we then calculate the logit of column 7, and put the logit output in column 8 of table 3.D5.

$$l_x = \frac{1}{\gamma_x\left[1+e^{2\alpha+\beta\ln\left(\frac{1-l_{sx}}{l_{sx}}\right)}\right]} \tag{3.D4}$$

Next we calculate the *standard survival function* obtained from the *standard life table* as shown in column 9 of table 3.D5. We then calculate the logit of the *standard survival function* in column 10, and finally we regress logit(l_x/ γ_x) on logit(l_{sx}) to obtain new values of $\hat{\alpha} = -0.3$ and $\hat{\beta} = 1.0$. With these new values we select the *new estimated survival function* ($\acute{l}_x$) for Kenyan women from INDEPTH tables and depict these in column 11 of table 3.D5.

Table 3.D5: Survival Function for Kenya (2003) Women

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Country	**Gender**	**Year**	**Age**	l_x	γ_x	l_x/γ_x	**Logit** (l_x/γ_x)	l_{sx}	**Logit** (l_{sx})	$\acute{l}_x$
Kenya	Female	2003	0	1.00000	1	1.00000		1.0000		1.00000
Kenya	Female	2003	1	0.93263	1	0.93263	-1.313904	0.9224	-1.2377	0.93556
Kenya	Female	2003	5	0.89702	1	0.89702	-1.082272	0.8766	-0.9803	0.89666
Kenya	Female	2003	10	0.88560	1	0.88560	-1.023282	0.8617	-0.9147	0.88386
Kenya	Female	2003	15	0.87963	1	0.87963	-0.994466	0.8539	-0.8828	0.87713
Kenya	Female	2003	20	0.87101	1	0.87101	-0.954959	0.8426	-0.8389	0.86735
Kenya	Female	2003	25	0.85182	1	0.85182	-0.874474	0.8174	-0.7494	0.84538
Kenya	Female	2003	30	0.81723	1	0.81723	-0.748846	0.7720	-0.6098	0.80528
Kenya	Female	2003	35	0.77769	1	0.77769	-0.626128	0.7205	-0.4735	0.75895
Kenya	Female	2003	40	0.74196	1	0.74196	-0.52809	0.6746	-0.3645	0.71689
Kenya	Female	2003	45	0.71117	1	0.71117	-0.450537	0.6357	-0.2784	0.68065
Kenya	Female	2003	50	0.68006	1	0.68006	-0.377024	0.5971	-0.1967	0.64414
Kenya	Female	2003	55	0.65146	1	0.65146	-0.312731	0.5623	-0.1253	0.61076
Kenya	Female	2003	60	0.61834	1.1	0.56213	-0.1249	0.5229	-0.0458	0.57240
Kenya	Female	2003	65	0.57766	1.1	0.52515	-0.050333	0.4759	0.04824	0.52586
Kenya	Female	2003	70	0.52384	1.1	0.47622	0.0476	0.4162	0.1692	0.46546
Kenya	Female	2003	75	0.44883	1.1	0.40803	0.186063	0.3379	0.33633	0.38399
Kenya	Female	2003	80	0.36830	1.1	0.33482	0.343236	0.2604	0.52195	0.30072
Kenya	Female	2003	85	0.25831	1.2	0.21526	0.646758	0.1657	0.80821	0.19522

Thereafter we derive the complete life table for Kenyan women, from $\acute{l}_x$ in column 11 of table 3.D5 as depicted by table 3.D6, where *n* in column 5 is age interval, $\acute{l}_x$ is number of survivors in age *x*; ${}_nd_x = \acute{l}_x - \acute{l}_{x+n}$ is the number of deaths that occurred between ages *x* and x_n. Column 7 shows ${}_na_x$, the percentage of age interval x and x_n who died, ${}_nm_x = \frac{{}_nd_x}{nl_{x+n}+n_na_x\,{}_nd_x}$ is the death rate; ${}_nL_x = {}_nl_{x+n} + n\ {}_na_x\ {}_nd_x$ is number of person-years lived between age interval x and x_n, T_x is number of person-years lived after age x, which is an accumulation of ${}_nL_x$ from age 100 to x, and $e_x = \frac{T_x}{l_x}$ is life expectancy at age x (INDEPTH, 2004).

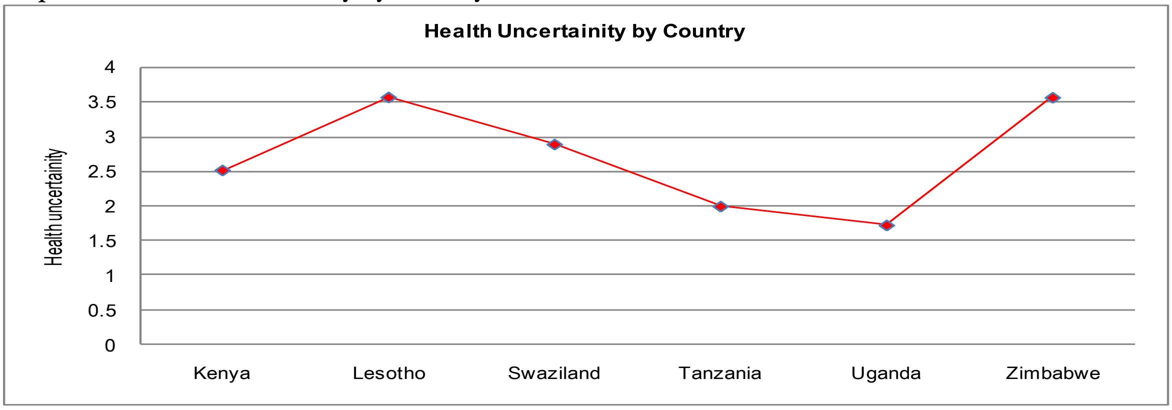
Table 3.D6: Complete Life Table for Kenya (2003): Women

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Country	**Gender**	**Year**	**Age**	**n**	l_x	${}_nd_x$	${}_na_x$	${}_nm_x$	${}_nL_x$	T_x	e_x
Kenya	Female	2003	0	1	100000	10689	0.35	0.11487107	93052	4564785	45.6
Kenya	Female	2003	1	4	89311	6723	0.34	0.019802926	339495	4471733	50.1
Kenya	Female	2003	5	5	82588	2173	0.5	0.005332417	407508	4132238	50.0
Kenya	Female	2003	10	5	80415	1132	0.5	0.002835352	399245	3724730	46.3
Kenya	Female	2003	15	5	79283	1631	0.5	0.004157135	392338	3325485	41.9
Kenya	Female	2003	20	5	77652	3599	0.5	0.00948947	379263	2933148	37.8
Kenya	Female	2003	25	5	74053	6318	0.5	0.017823793	354470	2553885	34.5
Kenya	Female	2003	30	5	67735	6870	0.5	0.021368585	321500	2199415	32.5
Kenya	Female	2003	35	5	60865	5833	0.5	0.020131669	289743	1877915	30.9
Kenya	Female	2003	40	5	55032	4721	0.5	0.017926203	263358	1588173	28.9
Kenya	Female	2003	45	5	50311	4483	0.5	0.01865216	240348	1324815	26.3
Kenya	Female	2003	50	5	45828	3869	0.5	0.017629034	219468	1084468	23.7
Kenya	Female	2003	55	5	41959	4185	0.5	0.020995071	199333	865000	20.6
Kenya	Female	2003	60	5	37774	4728	0.5	0.026704321	177050	665668	17.6
Kenya	Female	2003	65	5	33046	5602	0.5	0.03704414	151225	488618	14.8
Kenya	Female	2003	70	5	27444	6692	0.5	0.055539879	120490	337393	12.3
Kenya	Female	2003	75	5	20752	5926	0.5	0.066625443	88945	216903	10.5
Kenya	Female	2003	80	5	14826	6340	0.5	0.108785175	58280	127958	8.6
Kenya	Female	2003	85	5	8486	1536	0.5	0.039804552	38590	69678	8.2
Kenya	Female	2003	90	5	6950	4365	0.5	0.183120352	23837	31088	4.5
Kenya	Female	2003	95	5	2585	2269	0.5	0.312994548	7251	7251	2.8
Kenya	Female	2003	100		315	315.4			0	0	0.0

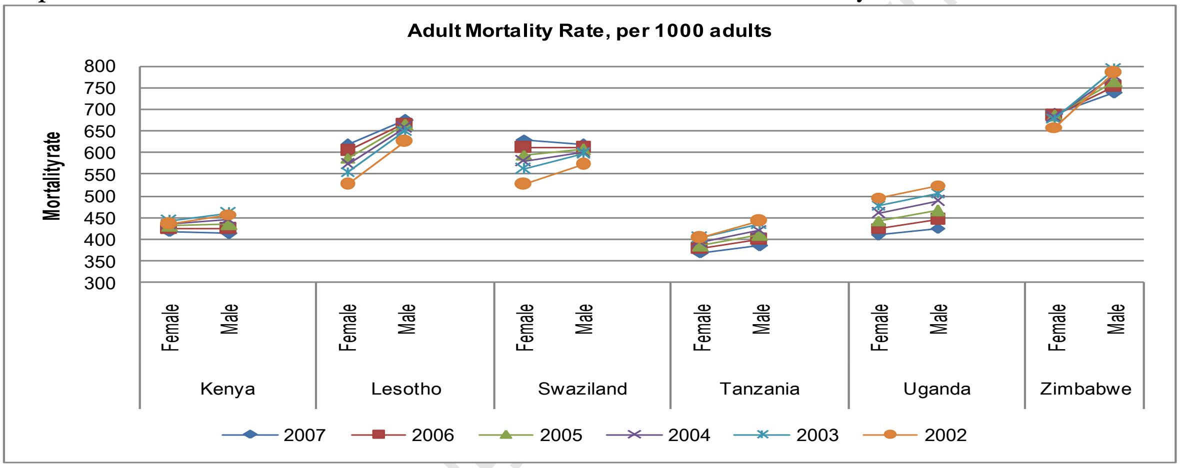
The procedure is repeated for each gender in each country to give the complete life table in 3.D1, table 3.D2 and table 3.D3. We then create a five year interval of the age variable in our dataset and match the age interval with the life expectancy in tables using country and gender.

Appendix 3.E: Validating the Health Uncertainty Proxy

Graph 3.E1: Health Uncertainty by Country



Graph 3.E2: World Bank Health Indicators: Male and Female Adult Mortality



Appendix 3.F: Imputing Household Income using Regression Methods

This Appendix explains the regression based approach that was used to derive the household income. The regression based approach is based on the assumption that households smooth out their consumption therefore monthly household expenditure should equal monthly income (Grim and Gunther 2007; Oster 2009). We apply the regression based approach using the Living Standards Measurement Surveys (LSMS) following Grim and Gunther (2007) and Oster (2009). LSMS surveys are nationally representative household surveys specifically designed to capture consumption, expenditure and income patterns for developing nations[78]. The LSMS surveys which were available to us include the 2008 South African National Income Dynamics Study (NIDS) and 2008 Tanzania National Panel Survey (TZNPS). Using these surveys, we regress per capita household expenditure on a set of household characteristics common to NIDS, TZNPS and DHS, that is, $\ln y_h^{\mathrm{NIDS}\,\&\,TZNPS} = \beta^{\mathrm{NIDS}\,\&\,TZNPS} X^{\mathrm{NIDS}\,\&\,TZNPS} + e_h$, where X is household characteristics common to DHS, NIDS and TZNPS. We then use the estimated coefficients to impute income, $\widehat{\ln} y_h^{\mathrm{DHS}} = \hat{\beta}^{\mathrm{NIDS}\,\&\,TZNPS} X^{\mathrm{DHS}}$. In this Appendix we explain the procedure in detail.

We begin by exploring the total expenditure information in both the NIDS and TZNPS. This total expenditure includes expenditure on food and rentals. In the NIDS data *total monthly food expenditure* is a summation of the amount *spent* on food, the *gift* amount, amount given as *payment*, and the amount *produced* by the household. Any missing information on any of these categories was imputed using either regression methods or the cell median method (Finn *et al.*, 2009). Unlike NIDS, TZNPS does not have data on *total monthly food expenditure*, we therefore need to calculate this using the available data. To ensure uniformity between NIDS and TZNPS datasets, we perform the following procedures on the TZNPS data; we first add quantity purchased, quantity produced and the gift quantity and name this *total quantity consumed* per food item. Thereafter we derive the monetary value of each food item which is the cost of one gram/millilitre of each food item. Then we calculate the average monetary value for each food item[79], as depicted in Table 3.F1 at the end of this Appendix.

We then multiplied the price of one gram/millilitre of each food item with the total quantity consumed to obtain expenditure per food item. After this we summed 56 of the 59 food items omitting the 3 which comprised expenditure on alcoholic beverages to obtain *total weekly food*

[78] The motivation for using LSMS surveys is that they cover the whole of SSA, they are compatible with DHS data; and are reliable and of a high quality.

[79] Of the 59 food items, 11 had a standard deviation of 10, while the rest had on average a standard deviation of 1. Food item 22 is an aggregation of honey, syrups, jams, marmalade, jellies and canned fruits, hence the high deviation of 83.8 is to be expected. The same holds for food items 21, 23, 24, 42, 43, 53, 55 and 58. However, what is puzzling is food item 2 which is rice husk with a deviation of 36; food item 45, fresh milk with a standard deviation of 165; and food item 48 cooking oil with a standard deviation of 713. One can only assume that the large variety of cooking oils, from liquid to solid variants from various manufactures/brands is driving this high deviation, the same can be said for milk and rice products.

expenditure. We imputed the 44 missing values using regression based on the following regressors: number of household members, schooling of head of household dummy, household has electricity dummy, household has TV dummy, roofing dummy, occupation dummy, and region dummy. The model gave an R^2 of 0.33, with all variables being statistically significant. This information is captured in Table 3.F2 at the end of Appendix 3.F. We later multiplied the *total weekly food expenditure* by 4 to derive the *total monthly food expenditure*.

Like the food expenditure section, the NIDS data includes *monthly rent* information[80]. From the TZNPS data we used the available data to derive a *monthly rent* variable. The rental section of TZNPS asked the following questions: "*What is the household tenure status of the residence*?", where households choose from the following responses: owner occupied, employer provided, rented, free and nomads. Those who indicated that the house was rented were asked to mention the rent amount they pay each month. We therefore need to determine the rents for the remaining households who do not pay rentals. We do this by regressing log monthly rent on the following regressors: number of rooms, number of rooms squared, the head of household has education dummy, household has electricity dummy, household has flush toilet dummy, household has piped water, and region dummy. The model is depicted in Table 3.F3 at the end of Appendix 3.F. We then added *total monthly food expenditure* and *monthly rent* to get *total household expenditure*. We divided this by the number of household members to obtain the *total household expenditure per capita*, our variable of interest.

Next we identified the *household characteristics* common in the three datasets (DHS, NIDS and TZNPS). These common characteristics include six durable goods ownership variables namely, ownership of radio, television, motor vehicle and tractors, motorcycle and bicycle; four housing attributes namely type of roof, type of wall, type of toilet facility and whether there is electricity in the house; and four variables from the household roster, namely number of household members, age of head of household, gender of head of household, occupational and educational attainment of head of household. The three datasets were merged keeping *total household expenditure* and *household characteristics* data, we include the GDP[81] and GDP per capita of the relevant countries in our merged dataset. We convert the expenditure values to US$[82] for 2008, since our data is for 2008, for South Africa the conversion rate was 8.0 while for Tanzania it was 1196.

Using NIDS and TZNPS household data we regress *total household expenditure per capita* on the *household characteristics* previously outlined. The output of the regression is depicted in Table 3.F4 at the end of this Appendix. The regression output shows that household income per capita as expected is on

[80] The monthly rental is derived from households who live in rented houses; households who live in houses which they own; and households who live in "*free*" houses. In cases where an individual indicated that the house was owned or free, a rent amount is calculated and any further missing information is imputed using regression methods.

[81] Oster (2009) included GDP and found that this improved the models fit

[82] Source: World Bank, International Comparison Program database, www.data.worldbank.org/indicators

average higher for households where the head of household is male and has a higher education qualification. Additionally possessing durable goods such as radio, TV, or car, are significant indicators of higher household income. Per capita income is also higher for households with better housing conditions, for example having electricity and flushing toilets. We then use the estimated coefficients to impute income.

Table 3.F1: Price of 1 Gram/Millilitre per Food Item

	Food Item	Observation	Mean Price	Std. Dev.	Minimum	Maximum
1	Rice (paddy)	9	0.7	0.4	0.2	1.3
2	Rice (husked)	1687	2.1	36.5	0.0	1500.0
3	Maize (green, cob)	100	0.7	0.6	0.0	5.0
4	Maize (grain)	229	0.5	0.3	0.0	2.0
5	Maize (flour)	1679	0.7	0.5	0.0	9.7
6	Millet and sorghum (grain)	39	0.9	0.5	0.2	2.5
7	Millet and sorghum (flour)	200	1.0	0.6	0.0	3.0
8	Wheat, barley grain and other cereals	75	1.1	0.6	0.1	5.0
9	Bread	831	1.5	0.9	0.0	15.0
10	Buns, cakes and biscuits	1036	1.9	1.7	0.0	25.0
11	Macaroni, spaghetti	208	1.7	0.7	0.5	4.8
12	Other cereal products	89	1.3	0.8	0.4	5.0
13	Cassava fresh	539	0.5	0.5	0.0	5.0
14	Cassava dry/flour	179	0.4	0.2	0.0	1.3
15	Sweet potatoes	410	0.4	0.3	0.0	2.4
16	Yams/cocoyam	90	0.6	0.4	0.1	2.0
17	Irish potatoes	801	0.7	0.3	0.1	2.9
18	Cooking bananas, plantains	445	0.7	2.6	0.1	55.0
19	Other starches	27	1.1	0.6	0.2	2.5
20	Sugar	2371	1.3	0.4	0.0	14.0
21	Sweets	195	11.5	12.6	0.7	75.0
22	Honey, syrups, jams, marmalade, jellies, canned fruits	248	12.5	83.8	0.1	1000.0
23	Peas, beans, lentils and other pulses	1628	2.4	49.5	0.1	2000.0
24	Groundnuts in shell/shelled	446	1.3	1.0	0.2	14.3
25	Coconuts (mature/immature)	1147	1.0	1.8	0.0	50.0
26	Cashew, almonds and other nuts	9	2.2	2.8	0.3	7.0
27	Seeds and products from nuts/seeds (*excl.* cooking oil)	24	1.5	1.9	0.3	10.0
28	Onions, tomatoes, carrots and green pepper, other viungo	2569	1.1	0.9	0.0	20.0
29	Spinach, cabbage and other green vegetables	1401	0.9	1.1	0.0	25.0
30	Canned, dried and wild vegetables	30	1.0	0.8	0.2	2.7
31	Ripe bananas	488	1.0	0.8	0.1	8.8
32	Citrus fruits (oranges, lemon, tangerines, etc.)	624	0.7	0.6	0.0	10.0
33	Mangoes, avocadoes and other fruits	406	0.8	0.6	0.1	8.6
34	Sugarcane	244	0.5	0.6	0.0	5.0
35	Goat meat	225	3.3	3.4	0.5	50.0
36	Beef including minced sausage	1053	3.6	1.1	0.2	7.2
37	Pork including sausages and bacon	165	2.9	1.2	0.3	10.0
38	Chicken and other poultry	192	4.0	2.0	0.3	11.4
39	Wild birds and insects	8	2.8	1.4	0.7	5.0
40	Other domestic/wild meat products	14	1.4	1.0	0.3	4.0
41	Eggs	351	4.2	2.0	0.1	30.0
42	Fresh fish and seafood (including dagaa)	1114	2.6	18.0	0.0	600.0
43	Dried/salted/canned fish and seafood (incl. dagaa)	1071	8.6	183.3	0.1	6000.0
44	Package fish	4	3.2	1.7	1.0	5.0
45	Fresh milk	619	9.3	164.9	0.1	4000.0
46	Milk products (like cream, cheese, yoghurt etc)	121	1.0	1.6	0.0	15.0
47	Canned milk/milk powder	39	8.1	6.8	0.4	20.0
48	Cooking oil	511	37.4	713.0	0.0	16000.0
49	Butter, margarine, ghee and other fat products	168	4.8	5.0	0.4	40.0
50	Salt	3044	0.8	0.9	0.0	17.5
51	Other spices	317	3.3	6.4	0.3	100.0
52	Tea dry	2033	7.4	6.7	0.0	150.0
53	Coffee and cocoa	52	14.0	16.3	0.5	70.0
54	Other raw materials for drinks	24	0.6	1.4	0.0	6.7
55	Bottled/canned soft drinks (soda, juice, water)	703	2.1	19.2	0.0	500.0
56	Prepared tea, coffee	20	1.1	0.7	0.2	2.6
57	Bottled beer	81	0.9	1.2	0.0	3.4
58	Local brews	264	1.5	13.7	0.1	200.0
59	Wine and spirits	4	6.2	6.8	0.2	12.5

Table 3.F2: Regression used for NIPS Food Expenditure Imputation

Dependent variables: Food expenditure	(1) OLS
Number of Household members	0.125*** (0.00435)
Head of household has education[1]	0.196*** (0.0304)
Household has no electricity[2]	-0.0854 (0.0523)
Household has no television[3]	-0.346*** (0.0533)
Household has no metal sheet roofing[4]	-0.190*** (0.0285)
Dodoma	-0.500*** (0.0990)
Kilimanjaro	-0.182* (0.0948)
Tanga	-0.215** (0.0926)
Morogoro	-0.474*** (0.0935)
Pwani	-0.215** (0.108)
Daresalaam	-0.301*** (0.0749)
Lindi	-0.590*** (0.0876)
Mtwara	-0.610*** (0.0828)
Ruvuma	-0.596*** (0.0897)
Iringa	-0.444*** (0.0904)
Mbeya	-0.390*** (0.0871)
Singida	-0.528*** (0.114)
Tabora	-0.537*** (0.0945)
Rukwa	-0.428*** (0.0992)
Kigoma	-0.657*** (0.0948)
Shinyanga	-0.212** (0.0894)
Kagera	-0.368*** (0.0913)
Mwanza	-0.574*** (0.0911)
Mara	-0.550*** (0.113)
Manyara	-0.237** (0.102)
Kaskazini Unguja	-0.452*** (0.105)
Kusini Unguja	-0.421*** (0.137)
Magharibu Unguja	-0.478*** (0.0842)
Kaskazini Pemba	-0.670*** (0.101)
Kusini Pemba	-0.639*** (0.0994)
Constant	10.94*** (0.0826)
Observations	3,143
R-squared	0.337

Standard errors in parentheses
● *** p<0.01, ** p<0.05, * p<0.1 indicates significance at 1%, 5% and 10%
● Dummy Reference: [1] head has no education • [2]household has electricity • [3]household has television, • [4]house has metal roofing and • region dummy reference is Arusha

Table 3.F3: Regression used for NIPS Rent Imputation

Dependent variables: Rent expenditure	(1) OLS
Number of rooms	0.572***
	(0.0934)
Number of rooms2	-0.0489***
	(0.0158)
House has no flush toilet[1]	-0.329***
	(0.0960)
House has no electricity[2]	-0.529***
	(0.0676)
House has a stand pipe[3]	-0.286***
	(0.108)
House gets water outside compound[4]	-0.427***
	(0.102)
Head of household has no education[5]	-0.391***
	(0.131)
Dodoma	5.143***
	(0.630)
Arusha	5.886***
	(0.468)
Kilimanjaro	5.092***
	(0.534)
Tanga	5.118***
	(0.510)
Morogoro	5.137***
	(0.474)
Pwani	5.197***
	(0.578)
Daresalaam	6.120***
	(0.452)
Lindi	5.127***
	(0.510)
Mtwara	5.092***
	(0.471)
Ruvuma	4.608***
	(0.578)
Iringa	5.305***
	(0.490)
Mbeya	5.088***
	(0.491)
Singida	4.837***
	(0.507)
Tabora	4.727***
	(0.478)
Rukwa	5.279***
	(0.477)
Kigoma	5.102***
	(0.483)
Shinyanga	5.543***
	(0.475)
Kagera	5.460***
	(0.519)
Mwanza	5.428***
	(0.469)
Mara	5.923***
	(0.480)
Manyara	5.593***
	(0.519)
Magharibu Unguja	5.382***
	(0.466)
Kusini Pemba	5.041***
	(0.643)
Constant	3.770***
	(0.493)
Observations	506
R-squared	0.636

Standard errors in parentheses
● *** p<0.01, ** p<0.05, * p<0.1 indicates significance at 1%, 5% and 10%
Dummy Reference: [1]household has flush toilet • [2]household has electricity • [3]household has piped water • [4]head of household has an education and • [5]Kaskazini Unguja

Table 3.F4: Regression used for Per Capita Income (US$) Imputation

Dependent variables: Per Capita Household Income	(1) (OLS)
GDP per capita	-0.000172***

	(3.81e-06)
Household size	-0.135***
	(0.00265)
Age of head	-0.000797
	(0.00249)
Age of head2	3.88e-05
	(2.40e-05)
Head women	-0.0861***
	(0.0144)
Head has no education	-0.616***
	(0.0360)
Head has primary education	-0.535***
	(0.0331)
Head has tertiary education	-0.325***
	(0.0310)
Household possess a radio	0.123***
	(0.0147)
Household possess a TV	0.0965***
	(0.0203)
Household possess a fridge	0.171***
	(0.0199)
Household possess a bicycle	0.147***
	(0.0200)
Household possess a motorcycle	0.214***
	(0.0455)
Household possess a car	0.379***
	(0.0234)
Household has electricity	-0.00494
	(0.0211)
Household has metal sheet (GCI) roofing	-0.0360**
	(0.0146)
Household has brick walls	0.0244
	(0.0158)
Household has flush toilet	0.141***
	(0.0187)
Constant	4.802***
	(0.0674)
Observations	9,795
R-squared	0.421
Predicted Income in DHS household (mean)	44.72
Predicted Income in DHS household (standard deviation)	26.83

Standard errors in parentheses

- *** p<0.01, ** p<0.05, * p<0.1 indicates significance at 1%, 5% and 10%

Appendix 3.G: Imputing Household Income by Wealth Index

The household wealth index is calculated using DHS household characteristics common to all six countries. These include: type of toilet, whether there is electricity in the household, presence of radio, television and bicycle, ownership of motorcycle or car, whether a telephone was available and type of floor material used to build the house.

The wealth index is calculated using Confirmatory Factor Analysis (CFA) with Satorra-Bertler "robust" standard errors, since this is specifically designed for categorical data as in the case of DHS household data. All ten variables had positive factor loading. Additionally, although the RMSEA were slightly higher than expected (0.08), the Tucker-lewis non-normed fit index (TLI) and Bentler's comparative fit index (CFI) were 0.9 and 0.9 respectively indicating a good fit. The output for the CFA is shown in Table 3.G1 at the end of Appendix 3.G.

Table 3.G1: CFA for Our Wealth Index

log likelihood = -78179.112 Number of obs = 41022

	Coef.	Satorra-Bentler Std. Err.	z	P>|z|	[95% Conf.	Interval]
Means						
toilet	2.056969	.0029081	707.32	0.000	2.05127	2.062669
electricity	1.787626	.0020193	885.27	0.000	1.783668	1.791584
radio	1.383989	.0024013	576.35	0.000	1.379283	1.388696
television	1.814709	.0019183	945.99	0.000	1.810949	1.818469
fridge	1.887304	.0015613	1208.82	0.000	1.884244	1.890364
motorcycle	1.984545	.000609	3258.48	0.000	1.983351	1.985739
car	1.943664	.0011384	1707.38	0.000	1.941433	1.945896
mainroof	1.529813	.0024643	620.80	0.000	1.524983	1.534643
Loadings						
wealthindex						
toilet	1	.	.	.	.	.
electricity	.8934538	.0065334	136.75	0.000	.8806486	.906259
radio	.4617158	.0054255	85.10	0.000	.4510821	.4723495
television	.8240739	.0069621	118.36	0.000	.8104283	.8377195
fridge	.6326114	.0069062	91.60	0.000	.6190755	.6461473
motorcycle	.0365848	.0025666	14.25	0.000	.0315544	.0416153
car	.288657	.0060631	47.61	0.000	.2767735	.3005405
mainroof	.0894097	.0073106	12.23	0.000	.0750813	.1037382
Factor cov.						
wealthinde~x	.140934	.0020922	67.36	0.000	.1368333	.1450346
Var[error]						
toilet	.2059985	.0014873	138.50	0.000	.2030834	.2089136
electricity	.0547693	.0008717	62.83	0.000	.0530608	.0564778
radio	.206497	.000864	239.01	0.000	.2048036	.2081904
television	.0552502	.0009301	59.40	0.000	.0534273	.0570731
fridge	.043594	.000611	71.35	0.000	.0423965	.0447914
motorcycle	.0150276	.0005686	26.43	0.000	.0139132	.0161421
car	.0414189	.0006579	62.96	0.000	.0401295	.0427083
mainroof	.2479845	.0006617	374.78	0.000	.2466877	.2492814
R2						
toilet	0.4062					
electricity	0.6726					
radio	0.1270					
television	0.6340					
fridge	0.5640					
motorcycle	0.0124					
car	0.2209					
mainroof	0.0045					

Goodness of fit test: LR = 8353.384 ; Prob[chi2(20) > LR] = 0.0000
Test vs independence: LR = 7.8e+04 ; Prob[chi2(28) > LR] = 0.0000

Satorra-Bentler Tsc = 6207.297 ; Prob[chi2(20) > Tsc] = 0.0000
Satorra-Bentler Tadj = 4551.246 ; Prob[chi2(14.7) > Tadj] = 0.0000
Yuan-Bentler T2 = 6940.149 ; Prob[chi2(20) > T2] = 0.0000

Appendix 3.H: Results of the First Stage CTSM

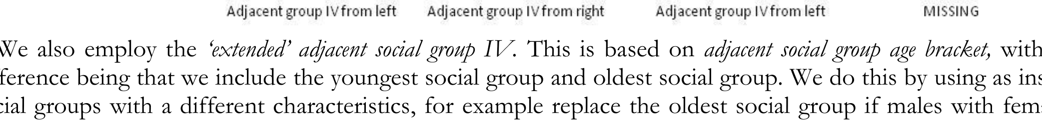
Table 3.H1: First Stage - Heckman Model

	(1)
Dependent variable: Had Sex	DHS
Age	0.117***
	(0.0227)
Age squared	-0.00158***
	(0.000364)
Gender	0.0101
	(0.0757)
Income	-9.82e-05
	(0.000985)
Education	-0.00663
	(0.0105)
HIV prevention, one partner	0.0171
	(0.126)
Lesotho	0.0101
	(0.132)
Swaziland	-0.0350
	(0.129)
Tanzania	-0.0286
	(0.122)
Uganda	-0.0125
	(0.127)
Zimbabwe	-0.0439
	(0.114)
Constant	-0.997
	(0.905)
Select	
HIV prevention, condom	-0.0500
	(0.118)
Constant	3.053***
	(0.113)
Mills	
lambda	-8.832
	(194.9)
Observations	65,293

Standard errors in parentheses
● *** p<0.01, ** p<0.05, * p<0.1 indicates significance at 1%, 5% and 10%

Appendix 3.I: Results using adjacent social group age bracket IV

The key to our IV strategy is social grouping. Recall that our social groups consist of individuals of the same gender, age and income. Based on these social grouping characteristics the *adjacent social group age bracket IV* is adopted from the works of Grodner and Kniesner (2006); Etile (2007) and Grodner *et al.*, (2011). As such the *adjacent social group age bracket IV* assumption is that the $\gamma_{\text{health}_{21-25}}$ (*health uncertainty* of a 21-25 age group) for example will be correlated with that of the adjacent age group to the left, $\gamma_{\text{health}_{15-20}}$ and to the right, $\gamma_{\text{health}_{26-30}}$ because of similar socioeconomic neighbourhoods, however the adjacent social $\gamma_{\text{health}_{15-20}}$ and $\gamma_{\text{health}_{26-30}}$ are uncorrelated with unobservables of individuals in social group $\gamma_{\text{health}_{21-25}}$ (for example unobservables of a 21 year old) because the individuals in age group 21-25 are not members of the adjacent groups (Grodner and Kniesner 2006; Etile 2007; Grodner *et al.* 2011). Therefore, adjacent groups to the left and to the right of the social group can be used as IVs. The advantage of the IVs is that they allow for overidentification of the models, with the only limitation being that the IVs exclude the youngest social group and the oldest social group[83] from analysis as these groups lack IVs[84].

[83] 

[84] We also employ the *'extended' adjacent social group IV*. This is based on *adjacent social group age bracket,* with the only difference being that we include the youngest social group and oldest social group. We do this by using as instruments social groups with a different characteristics, for example replace the oldest social group if males with female oldest social group. This is so as to include the total sample since the adjacent social group IV omits the first age group and the last age group.

Table 3.I1: Second Stage CTSM Risky Sex Results

	Objective life expectancy															
	Panel A *HIV-life expectancy 2* (2000s mortality rates)								**Panel B** *Pre-HIV-life expectancy 1* (1980s mortality rates)							
	Panel A.1				**Panel A.2**				**Panel B.1**				**Panel B.2**			
Dependent variable Risky sex index	IVTOBIT Nonlinear in means model				GMM Linear in means model				IVTOBIT Nonlinear in means model				GMM Linear in means model			
Type of instrument	Adjacent group IV[1]	'Extended' Adjacent group IV[1]	Trend setter IV[2]	'Extended' Trend setter IV[2]	Adjacent group IV[1]	'Extended' Adjacent group IV[1]	Trend setter IV[2]	'Extended' Trend setter IV[2]	Adjacent group IV[1]	'Extended' Adjacent group IV[1]	Trend setter IV[2]	'Extended' Trend setter IV[2]	Adjacent group IV[1]	'Extended' Adjacent group IV[1]	Trend setter IV[2]	'Extended' Trend setter IV[2]
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Expected health	**-0.0783*** **	**-2.450*** **	**-0.179*** **	**-0.404*** **	**-0.0945*** **	**-3.657*** **	**-0.240*** **	**-0.522*** **	**-0.00711*** **	**-0.00678*** **	**-0.0109*** **	**-0.0132*** **	**-0.00848*** **	**-0.00894*** **	**-0.0137*** **	**-0.0166*** **
	(0.00305)	**(0.907)**	**(0.00488)**	**(0.0236)**	**(0.00366)**	**(1.351)**	**(0.00658)**	**(0.0306)**	**(0.000313)**	**(0.000418)**	**(0.000341)**	**(0.000303)**	**(0.000371)**	**(0.000516)**	**(0.000420)**	**(0.000375)**
Health uncertainty	**0.0411*** **	**2.004*** **	**0.116*** **	**0.295*** **	**0.0501*** **	**2.999*** **	**0.157*** **	**0.383*** **	**0.0123*** **	**0.0181*** **	**0.0143*** **	**0.0128*** **	**0.0148*** **	**0.0224*** **	**0.0181*** **	**0.0162*** **
	(0.00249)	**(0.750)**	**(0.00414)**	**(0.0187)**	**(0.00298)**	**(1.118)**	**(0.00558)**	**(0.0242)**	**(0.000213)**	**(0.000314)**	**(0.000216)**	**(0.000220)**	**(0.000254)**	**(0.000388)**	**(0.000268)**	**(0.000273)**
Social group characteristics	-0.482***	-21.73***	-1.246***	-3.158***	-0.587***	-32.50***	-1.686***	-4.097***	0.558***	0.862***	0.624***	0.478***	0.666***	1.054***	0.780***	0.598***
	(0.0284)	(8.137)	(0.0462)	(0.204)	(0.0340)	(12.12)	(0.0622)	(0.265)	(0.0127)	(0.0205)	(0.0136)	(0.0138)	(0.0151)	(0.0252)	(0.0168)	(0.0170)
Expected income	0.0121***	0.542***	0.0324***	0.0790***	0.0148***	0.811***	0.0439***	0.103***	-0.0113***	-0.0176***	-0.0126***	-0.00920***	-0.0134***	-0.0215***	-0.0157***	-0.0114***
	(0.000709)	(0.203)	(0.00118)	(0.00508)	(0.000849)	(0.302)	(0.00159)	(0.00658)	(0.000285)	(0.000459)	(0.000317)	(0.000311)	(0.000340)	(0.000564)	(0.000390)	(0.000384)
Education, years	0.618***	0.207***	0.0342***	0.0470***	0.0267***	0.307***	0.0460***	0.0613***	0.0127***	0.0147***	0.0169***	0.0164***	0.0152***	0.0188***	0.0214***	0.0208***
	(0.0265)	(0.0751)	(0.00184)	(0.00437)	(0.00126)	(0.112)	(0.00247)	(0.00566)	(0.000980)	(0.00109)	(0.000998)	(0.000881)	(0.00117)	(0.00135)	(0.00123)	(0.00109)
Mills ratio	0.0222***	205.9	-68.31***	-30.45	-85.38***	317.9	-91.56***	-41.14	-72.24***	-85.83***	-90.88***	-86.69***	-85.73***	-109.9***	-116.9***	-110.8***
	(0.00105)	(202.9)	(11.43)	(25.84)	(7.791)	(302.2)	(15.39)	(33.47)	(6.329)	(6.929)	(6.395)	(5.615)	(7.545)	(8.634)	(7.928)	(6.999)
HIV prevention, one partner	-71.97***	-0.125	0.0250**	-0.00694	0.0260***	-0.188	0.0354**	-0.00634	0.0301***	0.0356***	0.0384***	0.0396***	0.0364***	0.0472***	0.0511***	0.0520***
	(6.487)	(0.200)	(0.0127)	(0.0288)	(0.00881)	(0.298)	(0.0171)	(0.0373)	(0.00714)	(0.00776)	(0.00711)	(0.00629)	(0.00853)	(0.00970)	(0.00884)	(0.00787)
Healthy looking person, HIV	0.0214***	1.432***	0.119***	0.219***	0.0627***	2.146***	0.160***	0.286***	0.00727	0.0191**	0.0433***	0.0508***	0.00845	0.0299***	0.0583***	0.0671***
	(0.00733)	(0.547)	(0.0127)	(0.0301)	(0.00906)	(0.814)	(0.0171)	(0.0390)	(0.00714)	(0.00767)	(0.00695)	(0.00617)	(0.00851)	(0.00960)	(0.00865)	(0.00772)
Rural	0.0523***	-2.265***	-0.123***	-0.348***	-0.0480***	-3.387***	-0.166***	-0.450***	0.0665***	0.0877***	0.0649***	0.0552***	0.0802***	0.108***	0.0818***	0.0697***
	(0.00754)	(0.859)	(0.00995)	(0.0299)	(0.00686)	(1.279)	(0.0134)	(0.0387)	(0.00497)	(0.00575)	(0.00515)	(0.00460)	(0.00591)	(0.00712)	(0.00633)	(0.00568)
Southern Africa	-0.0402***	20.79***	1.514***	3.359***	0.746***	31.04***	2.033***	4.343***	0.0478***	0.0601***	0.0655***	0.0722***	0.0570***	0.0763***	0.0822***	0.0906***
	(0.00573)	(7.721)	(0.0440)	(0.201)	(0.0317)	(11.50)	(0.0594)	(0.261)	(0.00452)	(0.00501)	(0.00468)	(0.00405)	(0.00537)	(0.00620)	(0.00575)	(0.00501)
Constant	9.785***	11.73	10.86***	10.26***	11.44***	15.76	14.17***	13.12***	8.357***	9.668***	10.29***	9.872***	9.725***	12.10***	12.94***	12.33***
	(0.662)	(17.28)	(1.164)	(2.609)	(0.795)	(25.73)	(1.568)	(3.379)	(0.645)	(0.706)	(0.652)	(0.573)	(0.769)	(0.880)	(0.808)	(0.714)
Observations	46,774	52,155	48,430	52,155	42,369	52,155	48,430	52,155	42,369	52,155	48,430	52,155	42,369	52,155	48,430	52,155
Anderson canon correlation statistics (p-value)					0.0000	0.0284	0.0000	0.0000					0.0000	0.0000	0.0000	0.0000
Sargan-Hansen (p-value)					0.0000	0.6381			0.0000	0.0000			0.0000	0.0000		
Amemiya-Lee Newey statistic (p-value)	0.0000	0.6385														

- Robust Standard errors in parentheses
- *** p<0.01, ** p<0.05, * p<0.1 indicates significance at 1%, 5% and 10%
- Definition of instruments follows from table 3.1.
- Region reference is Eastern Africa
- Health uncertainty is not significant in all models. Furthermore, most of the variables in the models are insignificant.

Appendix 3.J: Placebo social group using adjacent social group age bracket IV

Table 3.J1: Second Stage CTSM Risky Sex Results based on placebo social groups

	Objective life expectancy															
	Panel A *HIV-life expectancy 2* (2000s mortality rates)								Panel B *Pre-HIV-life expectancy 1* (1980s mortality rates)							
Dependent variable	Panel A.1				Panel A.2				Panel B.1				Panel B.2			
Risky sex index	IVTOBIT Nonlinear in means model				GMM Linear in means model				IVTOBIT Nonlinear in means model				GMM Linear in means model			
Type of instrument	Adjacent group IV	'Extended' Adjacent group IV	Trend setter IV	'Extended' Trend setter IV	Adjacent group IV	'Extended' Adjacent group IV	Trend setter IV	'Extended' Trend setter IV	Adjacent group IV	'Extended' Adjacent group IV	Trend setter IV	'Extended' Trend setter IV	Adjacent group IV	'Extended' Adjacent group IV	Trend setter IV	'Extended' Trend setter IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Expected health	-0.0330	0.0889	0.158	0.0880	-0.0442	0.119	0.213	0.117	-0.0123***	-0.0124***	-0.196	-0.00990	-0.0160***	-0.0160***	-0.255	-0.0131
	(0.0676)	(0.0943)	(0.611)	(0.0972)	(0.0887)	(0.124)	(0.817)	(0.128)	(0.000269)	(0.000156)	(0.300)	(0.0103)	(0.000356)	(0.000206)	(0.391)	(0.0134)
Health uncertainty[1]	**-0.000167**	**-0.00188**	**0.0124**	**-0.0145**	**-0.000254**	**-0.00272**	**0.0174**	**-0.0199**	**0.000879**	**0.000455**	**0.128**	**-0.00176**	**0.00124**	**0.000564**	**0.167**	**-0.00192**
	(0.00628)	**(0.00821)**	**(0.0979)**	**(0.0340)**	**(0.00825)**	**(0.0108)**	**(0.131)**	**(0.0448)**	**(0.00142)**	**(0.000578)**	**(0.209)**	**(0.00724)**	**(0.00185)**	**(0.000754)**	**(0.273)**	**(0.00941)**
Social group characteristics	0.0317	-0.0987	-0.102	-0.179	0.0425	-0.132	-0.134	-0.242	0.00856	-0.000305	0.197	0.00323	0.0110	-0.000360	0.257	0.00470
	(0.0265)	(0.103)	(0.165)	(0.282)	(0.0347)	(0.136)	(0.220)	(0.371)	(0.00785)	(0.00631)	(0.320)	(0.0147)	(0.0103)	(0.00824)	(0.417)	(0.0191)
Income	0.000431	0.00217*	0.00309	0.00216*	0.000521	0.00284*	0.00410	0.00282*	0.000894***	0.00100***	0.00549	0.000949***	0.00114***	0.00127***	0.00715	0.00124***
	(0.00101)	(0.00126)	(0.00813)	(0.00130)	(0.00133)	(0.00166)	(0.0109)	(0.00171)	(0.000111)	(6.43e-05)	(0.00737)	(0.000261)	(0.000145)	(8.39e-05)	(0.00962)	(0.000340)
Education, years	0.0200***	0.0194***	0.0238	0.0188***	0.0264***	0.0255***	0.0314	0.0246***	0.0213***	0.0198***	-0.00351	0.0202***	0.0281***	0.0260***	-0.00445	0.0264***
	(0.00388)	(0.00198)	(0.0157)	(0.00260)	(0.00510)	(0.00261)	(0.0211)	(0.00343)	(0.00145)	(0.000837)	(0.0398)	(0.00180)	(0.00190)	(0.00110)	(0.0520)	(0.00234)
Mills ratio	-80.99	-182.3**	-252.3	-181.1**	-104.5	-240.0**	-335.6	-238.4**	-85.30***	-91.43***	17.90	-92.46***	-111.8***	-119.7***	22.52	-121.1***
	(60.72)	(74.15)	(531.4)	(75.47)	(79.74)	(97.65)	(711.0)	(99.41)	(9.258)	(5.351)	(187.3)	(7.471)	(12.21)	(7.038)	(244.5)	(9.751)
HIV prevention, one partner	0.0238	0.171	0.256	0.170	0.0299	0.227	0.343	0.225	0.0344***	0.0543***	-0.129	0.0570***	0.0456***	0.0726***	-0.167	0.0757***
	(0.0509)	(0.106)	(0.716)	(0.109)	(0.0669)	(0.140)	(0.959)	(0.144)	(0.0103)	(0.00599)	(0.299)	(0.0130)	(0.0137)	(0.00791)	(0.390)	(0.0169)
Healthy looking person, HIV	0.0385	0.252	0.337	0.252	0.0473	0.333	0.451	0.333	0.0666***	0.0781***	0.319	0.0748***	0.0873***	0.103***	0.417	0.0999***
	(0.111)	(0.158)	(0.935)	(0.166)	(0.145)	(0.208)	(1.251)	(0.218)	(0.0101)	(0.00577)	(0.409)	(0.0138)	(0.0133)	(0.00763)	(0.534)	(0.0180)
Southern Africa	0.209	-0.766	-1.323	-0.754	0.280	-1.022	-1.789	-1.005	0.00138	0.00481	1.737	-0.0187	0.00120	0.00564	2.264	-0.0215
	(0.543)	(0.758)	(4.937)	(0.772)	(0.713)	(0.999)	(6.607)	(1.016)	(0.00636)	(0.00367)	(2.836)	(0.0961)	(0.00832)	(0.00478)	(3.703)	(0.125)
Constant	10.20***	15.36***	18.51	16.17***	12.95***	19.86***	24.10	20.97***	9.682***	10.34***	1.699	10.47***	12.37***	13.23***	1.986	13.41***
	(2.935)	(3.873)	(20.98)	(5.706)	(3.855)	(5.101)	(28.08)	(7.517)	(0.966)	(0.549)	(15.23)	(0.675)	(1.274)	(0.722)	(19.89)	(0.882)
Observations	17,794	52,155	35,479	52,155	17,794	52,155	35,479	52,155	17,794	52,155	35,479	52,155	17,794	52,155	35,479	52,155

- Robust Standard errors in parentheses
- *** p<0.01, ** p<0.05, * p<0.1 indicates significance at 1%, 5% and 10%
- Definition of instruments follows from table 3.1.
- Region reference is Eastern Africa
- Health uncertainty is not significant in all models. Furthermore, most of the variables in the models are insignificant.

Appendix 3.K: Regression Output Using an Alternative Expected Income Proxy

Table 3.K1: Wealth Index as the Expected Income Proxy

	Panel A 2000s mortality rates				Panel B 1980s mortality rates			
	IVTOBIT CTSM 2nd stage Nonlinear in means mode Risky sex index		GMM CTSM 2nd stage Linear in means model Risky sex index		IVTOBIT CTSM 2nd stage Nonlinear in means model Risky sex index		GMM CTSM 2nd stage Linear in means model Risky sex index	
	Adjacent group IV	Trend setter IV	Adjacent group IV	Trend setter IV	Adjacent group IV	Trend setter IV	Adjacent group IV	Trend setter IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Expected health	**-0.0545*****	**-0.535*****	**-0.0645*****	**-0.835*****	**-0.00244*****	**-0.00472*****	**-0.00288*****	**-0.00583*****
	(0.00184)	**(0.0722)**	**(0.00218)**	**(0.114)**	**(0.000526)**	**(0.000434)**	**(0.000618)**	**(0.000555)**
Health uncertainty	**0.0186*****	**0.350*****	**0.0223*****	**0.553*****	**0.0156*****	**0.0165*****	**0.0185*****	**0.0212*****
	(0.00131)	**(0.0512)**	**(0.00155)**	**(0.0808)**	**(0.000428)**	**(0.000230)**	**(0.000504)**	**(0.000298)**
Social group characteristics	-0.313***	-4.895***	-0.375***	-7.732***	0.798***	0.844***	0.940***	1.078***
	(0.0186)	(0.707)	(0.0220)	(1.116)	(0.0267)	(0.0133)	(0.0314)	(0.0171)
Expected income[1]	-0.0926***	-3.155***	-0.110***	-5.012***	0.491***	0.557***	0.580***	0.717***
	(0.0139)	(0.479)	(0.0164)	(0.756)	(0.0170)	(0.0125)	(0.0201)	(0.0161)
Education, years	0.0131***	0.272***	0.0159***	0.431***	-0.0359***	-0.0342***	-0.0421***	-0.0431***
	(0.00121)	(0.0396)	(0.00144)	(0.0624)	(0.00141)	(0.00110)	(0.00166)	(0.00141)
Mills ratio	-81.47***	-84.10	-95.46***	-112.0	-71.88***	-86.30***	-84.11***	-112.6***
	(6.099)	(51.95)	(7.223)	(81.87)	(8.876)	(8.758)	(10.45)	(11.25)
HIV prevention, one partner	0.0190***	-0.0756	0.0224***	-0.127	0.0450***	0.0432***	0.0532***	0.0564***
	(0.00709)	(0.0614)	(0.00840)	(0.0967)	(0.0103)	(0.0100)	(0.0121)	(0.0129)
Healthy looking person, HIV	0.0273***	0.0826	0.0316***	0.125	-0.00612	0.0111	-0.00794	0.0139
	(0.00703)	(0.0587)	(0.00831)	(0.0925)	(0.0103)	(0.00981)	(0.0121)	(0.0126)
Rural	-0.137***	-1.421***	-0.162***	-2.226***	0.187***	0.185***	0.221***	0.235***
	(0.00750)	(0.201)	(0.00886)	(0.317)	(0.0111)	(0.00898)	(0.0131)	(0.0115)
Southern Africa	0.405***	4.364***	0.479***	6.821***	0.0962***	0.100***	0.113***	0.126***
	(0.0153)	(0.597)	(0.0181)	(0.943)	(0.00712)	(0.00707)	(0.00838)	(0.00906)
Constant	10.93***	22.75***	12.65***	32.85***	7.669***	9.145***	8.795***	11.63***
	(0.624)	(5.548)	(0.739)	(8.745)	(0.906)	(0.893)	(1.067)	(1.147)
Observations	41,769	47,202	41,769	47,202	41,769	47,202	41,769	47,202

Standard errors in parentheses

- *** p<0.01, ** p<0.05, * p<0.1 indicates significance at 1%, 5% and 10%

[1]CFA wealth index is the expected income proxy in this regression

Appendix 3.L: Regression Output for Eastern Africa and Southern Africa

Table 3.L1: A Comparison between East and Southern Africa

	Panel A East Africa								Panel B Southern Africa							
	Panel A1 2000s mortality rates				Panel A2 1980s mortality rates				Panel B1 2000s mortality rates				Panel B2 1980s mortality rates			
	GMM CTSM 2nd stage Nonlinear in means mode Risky sex index		IVTOBIT CTSM 2nd stage Linear in means model Risky sex index		GMM CTSM 2nd stage Nonlinear in means model Risky sex index		IVTOBIT CTSM 2nd stage Linear in means model Risky sex index		GMM CTSM 2nd stage Nonlinear in means mode Risky sex index		IVTOBIT CTSM 2nd stage Linear in means model Risky sex index		GMM CTSM 2nd stage Nonlinear in means model Risky sex index		IVTOBIT CTSM 2nd stage Linear in means model Risky sex index	
	Adjacent group IV	Trend setter IV	Adjacent group IV	Trend setter IV	Adjacent group IV	Trend setter IV	Adjacent group IV	Trend setter IV	Adjacent group IV	Trend setter IV	Adjacent group IV	Trend setter IV	Adjacent group IV	Trend setter IV	Adjacent group IV	Trend setter IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Expected health	**-0.0766*** **	**-0.238*** **	**-0.0893*** **	**-0.304*** **	**-0.0150*** **	**-0.0311*** **	**-0.0173*** **	**-0.0382*** **	**-0.0763*** **	**-0.148*** **	**-0.0957*** **	**-0.206*** **	**-0.0140*** **	**-0.0695*** **	**-0.0175*** **	**-0.0937*** **
	(0.00595)	**(0.0125)**	**(0.00687)**	**(0.0160)**	**(0.000566)**	**(0.000697)**	**(0.000650)**	**(0.000835)**	**(0.00291)**	**(0.00370)**	**(0.00363)**	**(0.00522)**	**(0.00110)**	**(0.00147)**	**(0.00135)**	**(0.00203)**
Health uncertainty	**0.0326*** **	**0.137*** **	**0.0383*** **	**0.176*** **	**0.00844*** **	**0.00677*** **	**0.00984*** **	**0.00831*** **	**0.0447*** **	**0.103*** **	**0.0569*** **	**0.147*** **	**0.0137*** **	**0.00417*** **	**0.0170*** **	**0.00566*** **
	(0.00407)	**(0.00884)**	**(0.00470)**	**(0.0113)**	**(0.000260)**	**(0.000283)**	**(0.000300)**	**(0.000338)**	**(0.00268)**	**(0.00358)**	**(0.00332)**	**(0.00502)**	**(0.000332)**	**(0.000424)**	**(0.000413)**	**(0.000582)**
Social group characteristics	-0.461***	-1.708***	-0.543***	-2.185***	0.306***	0.0237	0.354***	0.0207	-0.495***	-1.081***	-0.629***	-1.515***	0.440***	-1.003***	0.542***	-1.366***
	(0.0527)	(0.110)	(0.0609)	(0.141)	(0.0181)	(0.0218)	(0.0208)	(0.0259)	(0.0300)	(0.0401)	(0.0372)	(0.0562)	(0.0318)	(0.0424)	(0.0392)	(0.0582)
Expected income[1]	0.0110***	0.0415***	0.0130***	0.0531***	-0.00607***	0.000312	-0.00702***	0.000611	0.0108***	0.0248***	0.0137***	0.0349***	-0.00845***	0.0194***	-0.0104***	0.0264***
	(0.00123)	(0.00265)	(0.00142)	(0.00338)	(0.000397)	(0.000493)	(0.000456)	(0.000586)	(0.000683)	(0.000939)	(0.000847)	(0.00132)	(0.000606)	(0.000835)	(0.000746)	(0.00114)
Education, years	0.0137***	0.00424	0.0160***	0.00526	0.0161***	0.0179***	0.0187***	0.0222***	0.00295*	-0.000825	0.00357*	-0.000332	0.0116***	0.00707***	0.0144***	0.0104***
	(0.00124)	(0.00265)	(0.00144)	(0.00338)	(0.00105)	(0.00107)	(0.00121)	(0.00128)	(0.00156)	(0.00223)	(0.00194)	(0.00311)	(0.00156)	(0.00193)	(0.00193)	(0.00264)
Mills ratio	-113.9***	-177.7***	-131.2***	-228.0***	-86.47***	-125.7***	-99.49***	-155.6***	-70.51***	-92.15***	-87.64***	-128.9***	-42.53***	-64.23***	-52.36***	-89.22***
	(7.812)	(16.31)	(9.063)	(20.86)	(6.676)	(6.714)	(7.704)	(8.085)	(9.485)	(13.71)	(11.81)	(19.19)	(9.588)	(11.88)	(11.90)	(16.39)
HIV prevention, one partner	0.0735***	0.0971***	0.0865***	0.130***	0.0517***	0.0722***	0.0616***	0.0948***	0.00310	0.00863	0.00432	0.0129	-0.00483	-0.00115	-0.00564	-0.000240
	(0.0105)	(0.0214)	(0.0122)	(0.0274)	(0.00931)	(0.00905)	(0.0108)	(0.0110)	(0.00889)	(0.0128)	(0.0110)	(0.0179)	(0.00902)	(0.0112)	(0.0112)	(0.0154)
Healthy looking person, HIV	0.0123	0.0597***	0.0142	0.0759***	-0.0152*	0.0331***	-0.0182**	0.0411***	0.0508***	0.0958***	0.0633***	0.137***	0.0134	0.0946***	0.0175	0.133***
	(0.00908)	(0.0183)	(0.0105)	(0.0235)	(0.00797)	(0.00774)	(0.00919)	(0.00934)	(0.0101)	(0.0141)	(0.0126)	(0.0197)	(0.0102)	(0.0123)	(0.0127)	(0.0170)
Rural	0.000383	-0.0299**	0.00115	-0.0379**	0.0156***	0.00849	0.0188***	0.0113*	-0.0608***	-0.142***	-0.0763***	-0.199***	0.106***	-0.169***	0.131***	-0.229***
	(0.00624)	(0.0133)	(0.00722)	(0.0169)	(0.00541)	(0.00556)	(0.00622)	(0.00661)	(0.00809)	(0.0115)	(0.0100)	(0.0161)	(0.00928)	(0.0118)	(0.0115)	(0.0162)
Tanzania[1]	-0.228***	-0.729***	-0.266***	-0.931***	-0.212***	-0.455***	-0.245***	-0.558***								
	(0.0198)	(0.0419)	(0.0229)	(0.0536)	(0.0102)	(0.0120)	(0.0117)	(0.0144)								
Uganda[1]	-0.520***	-1.810***	-0.607***	-2.310***	-0.148***	-0.390***	-0.171***	-0.477***								
	(0.0471)	(0.101)	(0.0545)	(0.129)	(0.00994)	(0.0117)	(0.0114)	(0.0139)								
Lesotho[2]									-0.210***	-0.459***	-0.266***	-0.639***	-0.0285*	-0.695***	-0.0423**	-0.950***
									(0.0147)	(0.0201)	(0.0182)	(0.0281)	(0.0161)	(0.0204)	(0.0199)	(0.0281)
Swaziland[2]									-0.196***	-0.412***	-0.248***	-0.581***	0.108***	-0.509***	0.128***	-0.696***
									(0.0118)	(0.0158)	(0.0146)	(0.0221)	(0.0147)	(0.0191)	(0.0181)	(0.0262)
Constant	14.45***	24.43***	16.51***	30.99***	10.24***	14.78***	11.63***	18.04***	10.36***	14.30***	12.64***	19.52***	5.534***	10.05***	6.585***	13.46***
	(0.838)	(1.742)	(0.972)	(2.228)	(0.683)	(0.688)	(0.788)	(0.829)	(0.975)	(1.405)	(1.213)	(1.967)	(0.979)	(1.213)	(1.215)	(1.673)
Observations	21,023	23,777	21,023	23,777	21,023	23,777	21,023	23,777	21,346	24,653	21,346	24,653	21,346	24,653	21,346	24,653

Standard errors in parentheses

- *** p<0.01, ** p<0.05, * p<0.1 indicates significance at 1%, 5% and 10%

[1] The reference country is Kenya, [2] The reference country is Zimbabwe

Appendix 3.M: Results using an Alternative Outcome Variable

Table 3.M1: Sex Partners Results

	Objective life expectancy															
	2000s mortality rates								1980s mortality rates							
	IVTOBIT				GMM				IVTOBIT				GMM			
	CTSM 2nd stage				CTSM 2nd stage				CTSM 2nd stage				CTSM 2nd stage			
	Linear in means model				Nonlinear in means model				Linear in means model				Nonlinear in means model			
	Risky sex index				Risky sex index				Risky sex index				Risky sex index			
	Adjacent group IV	'Naïve' Adjacent group IV	Trend setter IV	'Naïve' Trend setter IV	Adjacent group IV	'Naïve' Adjacent group IV	Trend setter IV	'Naïve' Trend setter IV	Adjacent group IV	'Naïve' Adjacent group IV	Trend setter IV	'Naïve' Trend setter IV	Adjacent group IV	'Naïve' Adjacent group IV	Trend setter IV	'Naïve' Trend setter IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Expected health	**-0.0382*** **	**-0.553*** **	**-0.0841*** **	**-0.274*** **	**-0.0497*** **	**-0.782*** **	**-0.131*** **	**-0.384*** **	**-0.00280*** **	**-0.00571*** **	**-0.00558*** **	**-0.00692*** **	**-0.00381*** **	**-0.00732*** **	**-0.00844*** **	**-0.00999*** **
	(0.00260)	**(0.0925)**	**(0.00266)**	**(0.0169)**	**(0.00305)**	**(0.124)**	**(0.00376)**	**(0.0236)**	**(0.000270)**	**(0.000336)**	**(0.000318)**	**(0.000280)**	**(0.000316)**	**(0.000406)**	**(0.000393)**	**(0.000342)**
Health uncertainty	**0.0202*** **	**0.453*** **	**0.0484*** **	**0.201*** **	**0.0261*** **	**0.642*** **	**0.0797*** **	**0.283*** **	**0.00666*** **	**0.0111*** **	**0.0105*** **	**0.00986*** **	**0.00864*** **	**0.0145*** **	**0.0137*** **	**0.0126*** **
	(0.00214)	**(0.0768)**	**(0.00229)**	**(0.0134)**	**(0.00251)**	**(0.103)**	**(0.00322)**	**(0.0187)**	**(0.000191)**	**(0.000277)**	**(0.000211)**	**(0.000221)**	**(0.000224)**	**(0.000335)**	**(0.000262)**	**(0.000269)**
Social group characteristics	-0.225***	-4.938***	-0.480***	-2.108***	-0.293***	-6.996***	-0.805***	-2.982***	0.314***	0.495***	0.498***	0.415***	0.405***	0.651***	0.629***	0.510***
	(0.0244)	(0.843)	(0.0255)	(0.146)	(0.0285)	(1.130)	(0.0358)	(0.203)	(0.0113)	(0.0176)	(0.0131)	(0.0134)	(0.0133)	(0.0212)	(0.0162)	(0.0163)
Income	0.00575***	0.122***	0.0125***	0.0519***	0.00748***	0.173***	0.0209***	0.0734***	-0.00649***	-0.0104***	-0.0106***	-0.00861***	-0.00837***	-0.0137***	-0.0133***	-0.0105***
	(0.000602)	(0.0207)	(0.000647)	(0.00357)	(0.000704)	(0.0278)	(0.000909)	(0.00497)	(0.000261)	(0.000404)	(0.000312)	(0.000311)	(0.000306)	(0.000487)	(0.000384)	(0.000378)
Education, years	-0.00717***	0.0600***	0.000424	0.0323***	-0.00756***	0.0872***	0.00413***	0.0464***	-0.0111***	-0.00608***	-0.00761***	-0.00592***	-0.0126***	-0.00678***	-0.00856***	-0.00643***
	(0.000656)	(0.0124)	(0.000758)	(0.00295)	(0.000768)	(0.0166)	(0.00106)	(0.00410)	(0.000547)	(0.000567)	(0.000629)	(0.000550)	(0.000640)	(0.000686)	(0.000771)	(0.000667)
Mills ratio	-80.90***	-72.36**	-91.54***	-76.04***	-94.06***	-88.67**	-116.1***	-92.91***	-75.49***	-82.80***	-92.42***	-83.27***	-86.69***	-99.06***	-114.3***	-101.6***
	(5.285)	(32.99)	(6.467)	(15.34)	(6.206)	(44.22)	(9.085)	(21.34)	(5.262)	(5.298)	(5.827)	(5.143)	(6.173)	(6.467)	(7.220)	(6.288)
HIV prevention, one partner	0.0191***	-0.00918	0.0210***	-0.00954	0.0227***	-0.0146	0.0270***	-0.0143	0.0242***	0.0282***	0.0311***	0.0288***	0.0293***	0.0355***	0.0413***	0.0374***
	(0.00595)	(0.0377)	(0.00718)	(0.0173)	(0.00699)	(0.0505)	(0.0101)	(0.0241)	(0.00591)	(0.00593)	(0.00646)	(0.00575)	(0.00693)	(0.00725)	(0.00803)	(0.00705)
Healthy looking person, HIV	0.0144**	0.199***	0.0286***	0.0610***	0.0188***	0.284***	0.0444***	0.0883***	-0.00351	0.00554	0.00905	0.0113**	-0.00450	0.00650	0.0153**	0.0181***
	(0.00578)	(0.0442)	(0.00688)	(0.0166)	(0.00678)	(0.0593)	(0.00967)	(0.0231)	(0.00573)	(0.00571)	(0.00619)	(0.00549)	(0.00671)	(0.00698)	(0.00767)	(0.00672)
Rural	-0.00831*	-0.405***	-0.0336***	-0.180***	-0.0121**	-0.575***	-0.0587***	-0.255***	0.0396***	0.0536***	0.0519***	0.0473***	0.0497***	0.0691***	0.0655***	0.0583***
	(0.00471)	(0.0759)	(0.00567)	(0.0179)	(0.00552)	(0.102)	(0.00794)	(0.0249)	(0.00435)	(0.00463)	(0.00490)	(0.00442)	(0.00510)	(0.00562)	(0.00604)	(0.00538)
Southern Africa	0.278***	4.432***	0.650***	2.133***	0.364***	6.274***	1.031***	2.994***	0.00540	0.0157***	0.0113**	0.0185***	0.0106**	0.0229***	0.0218***	0.0302***
	(0.0219)	(0.750)	(0.0235)	(0.137)	(0.0257)	(1.006)	(0.0331)	(0.190)	(0.00388)	(0.00397)	(0.00438)	(0.00383)	(0.00454)	(0.00482)	(0.00539)	(0.00466)
Constant	10.03***	15.29***	11.97***	12.98***	11.61***	19.79***	15.25***	16.32***	8.763***	9.524***	10.54***	9.592***	9.925***	11.20***	12.82***	11.50***
	(0.541)	(3.434)	(0.660)	(1.569)	(0.635)	(4.604)	(0.928)	(2.184)	(0.537)	(0.541)	(0.594)	(0.525)	(0.630)	(0.660)	(0.737)	(0.642)
Observations	54,467	67,735	61,160	67,735	54,467	67,735	61,160	67,735	54,467	67,735	61,160	67,735	54,467	67,735	61,160	67,735

Standard errors in parentheses

- *** p<0.01, ** p<0.05, * p<0.1 indicates significance at 1%, 5% and 10%

Table 3.M2: DHS Condom Use Results

	Objective life expectancy															
	2000s mortality rates								1980s mortality rates							
	IVTOBIT CTSM 2nd stage Linear in means model Risky sex index				GMM CTSM 2nd stage Nonlinear in means model Risky sex index				IVTOBIT CTSM 2nd stage Linear in means model Risky sex index				GMM CTSM 2nd stage Nonlinear in means model Risky sex index			
	Adjacent group IV	'Naïve' Adjacent group IV	Trend setter IV	'Naïve' Trend setter IV	Adjacent group IV	'Naïve' Adjacent group IV	Trend setter IV	'Naïve' Trend setter IV	Adjacent group IV	'Naïve' Adjacent group IV	Trend setter IV	'Naïve' Trend setter IV	Adjacent group IV	'Naïve' Adjacent group IV	Trend setter IV	'Naïve' Trend setter IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Expected health	**-0.186*** **	**-1.428*** **	**-0.302*** **	**-0.879*** **	**-0.252*** **	**-2.310*** **	**-0.477*** **	**-1.325*** **	**-0.0111*** **	**-0.00593*** **	**-0.0200*** **	**-0.0193*** **	**-0.0158*** **	**-0.0105*** **	**-0.0294*** **	**-0.0288*** **
	(0.00706)	**(0.236)**	**(0.00761)**	**(0.0545)**	**(0.00955)**	**(0.359)**	**(0.0119)**	**(0.0819)**	**(0.000608)**	**(0.000781)**	**(0.000597)**	**(0.000554)**	**(0.000820)**	**(0.00111)**	**(0.000839)**	**(0.000793)**
Health uncertainty	**0.113*** **	**1.153*** **	**0.198*** **	**0.662*** **	**0.154*** **	**1.878*** **	**0.319*** **	**1.001*** **	**0.0267*** **	**0.0370*** **	**0.0261*** **	**0.0243*** **	**0.0360*** **	**0.0519*** **	**0.0367*** **	**0.0347*** **
	(0.00581)	**(0.196)**	**(0.00654)**	**(0.0432)**	**(0.00786)**	**(0.299)**	**(0.0102)**	**(0.0649)**	**(0.000430)**	**(0.000644)**	**(0.000396)**	**(0.000436)**	**(0.000588)**	**(0.000919)**	**(0.000564)**	**(0.000625)**
Social group characteristics	-1.415***	-12.70***	-2.271***	-7.217***	-1.908***	-20.64***	-3.579***	-10.87***	1.044***	1.663***	0.936***	0.796***	1.405***	2.308***	1.291***	1.121***
	(0.0662)	(2.151)	(0.0728)	(0.469)	(0.0895)	(3.277)	(0.113)	(0.705)	(0.0254)	(0.0410)	(0.0246)	(0.0265)	(0.0345)	(0.0581)	(0.0344)	(0.0376)
Income	0.0359***	0.314***	0.0585***	0.178***	0.0484***	0.510***	0.0923***	0.268***	-0.0207***	-0.0350***	-0.0186***	-0.0155***	-0.0279***	-0.0485***	-0.0255***	-0.0218***
	(0.00163)	(0.0529)	(0.00185)	(0.0115)	(0.00221)	(0.0806)	(0.00288)	(0.0173)	(0.000588)	(0.000940)	(0.000585)	(0.000614)	(0.000795)	(0.00133)	(0.000817)	(0.000870)
Education, years	0.00470***	0.173***	0.0196***	0.116***	0.00833***	0.282***	0.0365***	0.178***	-0.0150***	-0.00545***	-0.00919***	-0.00350***	-0.0175***	-0.00389**	-0.00898***	-0.00119
	(0.00178)	(0.0317)	(0.00217)	(0.00949)	(0.00240)	(0.0482)	(0.00335)	(0.0143)	(0.00123)	(0.00132)	(0.00118)	(0.00109)	(0.00166)	(0.00188)	(0.00164)	(0.00154)
Mills ratio	-97.78***	-69.55	-98.37***	-79.21	-157.5***	-132.5	-176.1***	-145.3*	-75.78***	-99.49***	-97.52***	-104.4***	-125.7***	-169.9***	-163.2***	-175.1***
	(14.36)	(84.19)	(18.48)	(49.33)	(19.42)	(128.2)	(28.68)	(74.16)	(11.84)	(12.33)	(10.94)	(10.17)	(16.08)	(17.72)	(15.49)	(14.61)
HIV prevention, one partner	0.0254	-0.0649	0.0244	-0.0586	0.0416*	-0.0905	0.0432	-0.0776	0.0489***	0.0487***	0.0555***	0.0546***	0.0722***	0.0785***	0.0868***	0.0864***
	(0.0162)	(0.0962)	(0.0205)	(0.0557)	(0.0219)	(0.147)	(0.0319)	(0.0838)	(0.0133)	(0.0138)	(0.0121)	(0.0114)	(0.0181)	(0.0199)	(0.0173)	(0.0165)
Healthy looking person, HIV	0.111***	0.508***	0.140***	0.252***	0.160***	0.852***	0.223***	0.393***	0.0281**	0.0115	0.0650***	0.0745***	0.0479***	0.0371*	0.108***	0.123***
	(0.0157)	(0.113)	(0.0197)	(0.0533)	(0.0212)	(0.172)	(0.0306)	(0.0801)	(0.0129)	(0.0133)	(0.0116)	(0.0109)	(0.0175)	(0.0192)	(0.0165)	(0.0157)
Rural	-0.000428	-0.959***	-0.0762***	-0.519***	-0.0176	-1.595***	-0.157***	-0.810***	0.218***	0.261***	0.197***	0.193***	0.273***	0.345***	0.252***	0.251***
	(0.0128)	(0.194)	(0.0162)	(0.0577)	(0.0173)	(0.295)	(0.0251)	(0.0867)	(0.00980)	(0.0108)	(0.00920)	(0.00874)	(0.0133)	(0.0154)	(0.0129)	(0.0125)
Southern Africa	1.262***	11.27***	2.294***	6.757***	1.726***	18.30***	3.663***	10.21***	-0.105***	-0.123***	-0.0630***	-0.0888***	-0.119***	-0.152***	-0.0656***	-0.103***
	(0.0596)	(1.915)	(0.0671)	(0.440)	(0.0806)	(2.917)	(0.105)	(0.661)	(0.00873)	(0.00923)	(0.00822)	(0.00757)	(0.0118)	(0.0132)	(0.0115)	(0.0108)
Constant	14.76***	26.92***	16.52***	22.41***	21.88***	44.13***	27.06***	35.43***	9.327***	11.56***	11.72***	12.28***	14.35***	18.60***	18.42***	19.42***
	(1.469)	(8.765)	(1.886)	(5.048)	(1.987)	(13.35)	(2.929)	(7.589)	(1.208)	(1.258)	(1.115)	(1.038)	(1.640)	(1.809)	(1.580)	(1.492)
Observations	54,467	67,735	61,160	67,735	54,467	67,735	61,160	67,735	54,467	67,735	61,160	67,735	54,467	67,735	61,160	67,735

Standard errors in parentheses
- *** p<0.01, ** p<0.05, * p<0.1 indicates significance at 1%, 5% and 10%

Chapter 4
The Role of Community Social Disorganisation in Contact and Contact Related Crime

Abstract

This study investigates how community social disorganisation, *family disruption*, *residential instability*, *community disadvantage*, *lack of basic services* and *unemployment*, can induce individuals to resort to *crimes of social behaviour,* namely *contact crime and contact related crime* in Cape Town, South Africa. The study uses a fixed effects model on a three year panel of seventy seven communities. To obtain the results the study needed to first circumvent four estimation challenges namely, excessive collinearity, measurement error as a result of reporting bias, omitted variable bias, and the influence of community unobservables. We find that the community social disorganisation attributes to be positive and significant in predicting *contact crime* and *contact related crime*.

Keywords: Community social disorganisation, Contact crime, Contact related crime

4.1 Introduction

We study the impact of social disorganisation on crime amongst communities in Cape Town which consist of a unique assortment of heterogeneous communities still largely segregated along racial lines. South Africa, and Cape Town in particular, has complex and unique socially disorganised communities, an aftermath of the apartheid past and the current high prevalence of HIV/AIDS (Emmett 2003; Demombynes and Ozler 2005; Breetzke 2010; Budlender and Lund 2011; Schatz *et al.* 2011). A poignant quote by Silber and Geffen (2009) provides us with some insight into the intimate relationship between social disorganisation and crime in Khayelitsha, one of Cape Town's largest informal settlements:

...Khayelitsha, where 38% of inhabitants live in formal structures, 20% live with access to water, 65% with flush toilet, and a significant number without electricity, unemployment is over 50%…Women are frequently beaten and raped walking to the toilet or fetching water from taps not more than 50 metres from their homes, children are routinely injured or killed by cars that hurtle through their backyards, illness and death are common, houses are frequently lost to fire and flooding and contact crime is ever present. These burdens pervade every crevice of a township that is overwhelmingly black and poor…The inadequate provision of these and other basic amenities… serves to exacerbate the broader deficit and security and results in several constitutional rights being routinely violated... (Silber and Geffen 2009: p. 37).

A mounting body of literature argues that the prevailing socially disorganised communities are pivotal ingredients in creating a conducive environment for crimes to flourish[85] (Blackmore 2003; Emmett 2003; Louw 2007; Landau and Misago 2009; Whyte 2010; Lau *et al.* 2010; Daniels and Adams 2010; Breetzke 2010). However, there exists a paucity of empirical literature to substantiate the effects of socially disorganised communities on crime in developing countries in general, and in sub-Saharan Africa in particular (Ward 2007[86]; Breetzke 2010; SAIRR[87] 2010; Daniels and Adams 2010).

Consequently, this study analyses the effects of community social disorganisation, specifically *family disruption*, *residential instability*, *community disadvantage*, *lack of basic services* and *unemployment* on *contact crime* and *contact related crime* in the Cape Town Metropolitan area of South Africa using a three year panel data. This is achieved by extending *Becker's theory* to include social structures following the work of Fajnzylber *et al.* (2002) and Haddad and Moghadam (2011). This theme is closely related to the empirical work of Sampson *et al.* (1989); Hannon *et al.* (1998); Gorman *et al.* (2001); Fajnzylber *et al.* (2002)[88]; Breetzke (2010); Miyoshi (2011) and Haddad and Moghadam (2011). As such we

[85] Gangsterism in Western Cape is linked to forced removals of Coloured families and their dispersal all over the Cape Flats. Standing and other researchers have argued that informal social control that communities had over youth was lost when established Cape Town communities were disbanded under the Group Areas Act. Crime and felony-related conflict became more prominent in Coloured communities after they were relocated to the Cape Flats (Daniels and Adams 2010: p. 47).

[86] While these factors have been identified in a number of US studies as being related to outcomes probably linked to social organisation of communities, their relevance in other contexts has yet to be established (Ward 2007: p. 79).

[87] South African Institute of Race Relation (SAIRR).

[88] It is worth noting that the study by Fajnzylber *et al.* (2002) is based on panel data where countries are the units of analysis.

acknowledge these studies and the contribution they make to empirical literature on criminal behaviour.

The study contributes to the current literature in that most of the existing empirical studies on social disorganisation concentrate on international experience and little research exists on sub-Saharan Africa (Ward 2007; SAIRR 2010; Daniels and Adams 2010). Our literature review identifies one study by Breetzke (2010) which was conducted in the Tswane region of South Africa. In addition, most of the aforementioned empirical studies are based on cross-sectional analysis. Yet it is well established that one of the major shortcomings of such cross sectional studies is the bias resulting from lack of controlling for community unobservables (Fajnzylber *et al.* 2002; Blackmore 2003). Hence the study extends the current literature by using panel data analysis affording us the ability to control for community unobservables. In addition we also control for measurement error and omitted variable bias which is often a challenge that is ignored in empirical criminological literature (Fajnzylber *et al.* 2002).

In this study the community is the unit of analysis. In total there are 77 communities over a period of 3 years covering the years 2001/2002, 2004/2005 and 2005/2006, implying that our total sample size is 231 communities. The panel is a combination of the Cape Area Panel Study (CAPS) data and Cape Town crime data. CAPS is a panel study of young adults in Metropolitan Cape Town, while the Cape Town crime data is crime statistics from the South African police service. As previously stated to derive the results the study needed to first circumvent four estimation challenges namely, excessive collinearity among the regressors, measurement error as a result of reporting bias, omitted variable bias, and the influence of community unobservables. To control for unobservable community heterogeneity the study adopts a fixed effects model of the panel data.

To curb the excessive collinearity we construct indices of the related variables, as proposed by among others Greene (2002). Our proxy measure for *family disruption* is the frequency of interaction between the youth and both parents, where the lower the frequency of interaction with both parents the higher the *family disruption*. Our proxy measure for *unemployment* is if the head of the household is unemployed. Lastly the proxy measure for *residential instability* is the percentage of individuals that have recently moved. The types of crime that constitute *contact crimes* include murder, attempted murder, rape, assault, and all forms of robbery. While the *contact related crimes* include malicious damage to property, arson, public violence and crimen injuria.

Two approaches are used to control for measurement error. The first is the ‘naive approach’ where we constrain the analysis to include only those crime statistics that are not prone to measurement error. This approach involves eliminating rape and robbery from the *contact crime* outcome variable. The second approach follows Demombynes and Ozler (2005) and adjusts the rape and robbery statistics with the probability of reporting crime derived from the 1998 Victims of Crime Survey.

The last estimation challenge is omitted variable bias as a result of the mobile nature of some crimes. This type of error is often ignored in the empirical literature of crime. The examples of such economic crimes include house robbery, bank robbery and cash in transit heists. The *mobile* nature of these crimes makes it difficult to establish the relationship with social or economic characteristics as these crimes are not area specific. To militate against this bias we follow the 'naive approach' and exclude crime statistics that are prone to omitted variable bias. In our case this involves eliminating robbery from our list of *contact crimes*.

Additionally, all the regressors are interacted with *majority racial population group* in each community. This is a dummy variable consisting of Black, Coloured and White population groups residing in the greater Cape Town area. This is in order to capture the effects of the differences in communities based on racial composition, an effect of the apartheid past where communities differ by racially determined geographical areas (Emmett 2003; Demombynes and Ozler 2005; Hipp 2007; Breetzke 2008; Breetzke 2010). In general the results reveal that the social factors are positive and significant in predicting both *contact crimes* and *contact related crimes*. In addition the results reveal that the majority Black communities are the most affected by social disorganisation which is consistent with the literature (Emmett 2003; Breetzke 2010; SAIRR 2010; Daniels and Adams 2010).

The remainder of this paper is organised as follows: section 4.2 provides the details of existing empirical literature related to the current theme of this study including our contribution to the existing literature. Thereafter section 4.3 provides a background to and motivation for the study. In section 4.4 we outline the theoretical framework of criminal behaviour that will guide our empirical model. Section 4.5 introduces the data and defines our outcome variable and the main regressors used as proxies for community social disorganisation. This is followed by the empirical framework which provides details of our estimation model, the likely estimation challenges of our model and our strategy to overcome these challenges. Section 4.6 describes the empirical results of our econometric modelling. We conclude in section 4.7 with a discussion of our findings.

4.2 Review of Previous Empirical Research

The overall theme of this paper relates to the previous empirical literature on socioeconomic structures that predict criminal behaviour. This include Sampson and Groves (1989); Hannon and Defronzo (1998); Gorman *et al.* (2001); Fajnzylber *et al.* (2002); Breetzke (2010); Miyoshi (2011) and Haddad and Moghadam (2011). Sampson and Groves (1989) which is perhaps one of the most cited empirical criminological studies utilises cross sectional data to investigate *social disorganisation theory* in 238 localities in the United Kingdom. This theory is founded on the premise that factors such as ethnic diversity, poverty, population mobility, and family disruption increase the complexity of a community's social organisation and ultimately lead to high levels of crime (Sampson and Groves 1989; Kawachi *et al.* 1999; Demombynes and Ozler 2005; Ward 2007; Breetzke 2010; Daniels and

Adams 2010). In line with this theory Sampson and Groves (1989) find evidence that communities with higher levels of family disruptions, ethnic heterogeneity and urbanisation do indeed have higher crime rates.

Along similar lines Hannon and Defronzo (1998) adopts *anomie/strain theory* and uses cross sectional data from 406 counties in the United States of America (USA) to examine the effects of resource deprivation, unemployment, residential stability and public assistance on county crime rates. *Strain theory* argues that delinquency occurs when individuals turn to illegitimate channels after they are not able to achieve their goals through legitimate channels (Agnew 1985). Consistent with *strain theory* Hannon and Defronzo (1998) find that resource deprivation is less significant in determining crime in areas with higher levels of public assistance. Conversely Gorman *et al.* (2001) analysed the effects of neighbourhood structure, specifically poverty levels, racial concentration, residential stability, educational attainment and alcohol outlets (for example clubs and hotels) on violent crime in 98 blocks in the USA and found that areas with a high concentration of alcohol outlets had a higher incident of violent crime.

On the other hand, Fajnzylber *et al.* (2002) extended *Becker's theory* to include social and demographic factors that determine crime in a multinational study. According to *Becker's theory* the decision to commit crime is rational and based on the expected costs and benefits of committing a crime. The expected cost in this case is the charge/sentencing given to criminals that are apprehended, while the expected benefit is the difference between the wealth from the crime and the opportunity cost of the crime (Fajnzylber *et al.* 2002; Miyoshi *et al.* 20110). Fajnzylber *et al.* (2002) use panel data of 45 Western industrialised countries, Latin America and the Caribbean, Eastern Europe and Central Asia, the Middle East and North Africa and Asia (excluding sub-Saharan Africa) and find that previous crime rates, GDP growth rate, income distribution, presence of policing, drugs, urbanisation, religion, proportion of young males, and average years of education influence murder and robbery.

Closely aligned with Sampson *et al.* (1989) and in a more recent application of *social disorganisation theory* is an empirical study by Breetzke (2010) who measured the effects of family disruption, unemployment and deprivation on crime in 354 communities in Tshwane, South Africa. Breetzke (2010) using data from the 2001 South African population census and South African Police Services (SAPS) crime data finds that whether the father is deceased or estranged is positively significant in determining crime, whereas whether a household is female headed is insignificant in predicting crime. In addition, contrary to the authors' expectation, unemployment was found to be negatively related to crime. In another recent empirical investigation Miyoshi (2011) provide evidence that is consistent with *Becker's theory* using panel data from Japanese provinces. The finding from their study indicates that policing, employment and education are significant in predicting crime. Finally, Haddad and Moghadam (2011) also estimate an extension to *Becker's theory* in Iran using panel data on provinces.

The authors find that economic variables play a key role in predicting the outcome variables, burglary and threats.

As previously mentioned this study contributes to the current literature in that most of the existing empirical research ignores the aforementioned estimation challenges. Additionally the current empirical studies on social disorganisation focus on international experience (Ward 2007; SAIRR 2010; Daniels and Adams 2010). Among the aforementioned studies the empirical study by Breetzke (2010) is the only one conducted in South Africa in the Tswane area. We would like to acknowledge other empirical studies on crime in South Africa which have also been reviewed in detail by Breetzke (2010). Gilfillan (1999) assessed countrywide socio-economic determinants of crime, while Brown (2001) applied pairwise correlation to determine the nationwide socioeconomic variables that predict crime. Blackmore (2003) used a nationwide panel to determine the effects of socio-economic factors on crime in South Africa. We also acknowledge other empirical studies on crime that have used similar panel data methods based on international experience. The list is however unlikely to be exhaustive; Cornwell and Trumbull (1994) for example use panel data on the counties of North Carolina and find labour markets and the criminal justice system to be important deterring factors of crime. Along similar lines, Fajnzylber *et al.*, (2002) used panel data of 45 Western industrialised countries, Latin America and the Caribbean, Eastern Europe and Central Asia, the Middle East and North Africa and Asia (excluding sub-Saharan Africa) and find that economic variables, crime deterrence, demographic, cultural as well as illegal drugs influence murder and robbery. In another recent empirical investigation Miyoshi (2011) provide evidence that is consistent with Becker's theory using panel data of Japanese provinces and find that policing, employment and education are significant in predicting crime. Finally, Haddad and Moghadam (2011) also estimate an extension to Becker's theory in Iran using panel data of provinces and find that economic variables play a key role in predicting burglary and threats.

4.3 Motivation for Research Problem in the South African Context

The current international literature shows that community social disorganisation leads to criminal behaviours (Sampson and Groves 1989; Kawachi *et al.* 1999), we study this assertion in the context of communities in Cape Town, South Africa. Figure 4.1 shows the number of crimes committed in the Metropolitan area of Cape Town (our study area) by type of crime from the period 2001/2002 to 2005/2006. As is evident, the majority of the crimes are *contact crimes,* which usually occur as a result of social behaviour (SAPS 2010), hence mirroring the condition of existing social structures. The study therefore investigates whether community social disorganisation has any role to play in *contact crimes* and *contact related crimes*.

Figure 4.1: Total Number of Crimes by the Five Major Categories in Cape Town Metropolitan Area

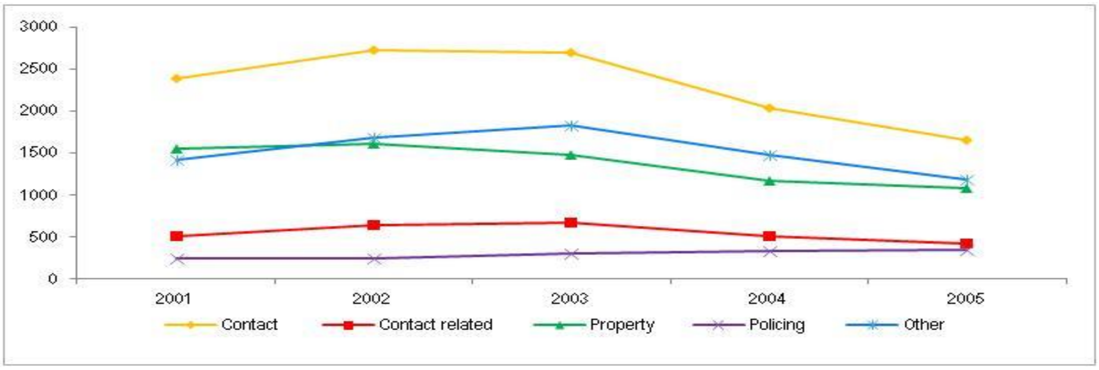
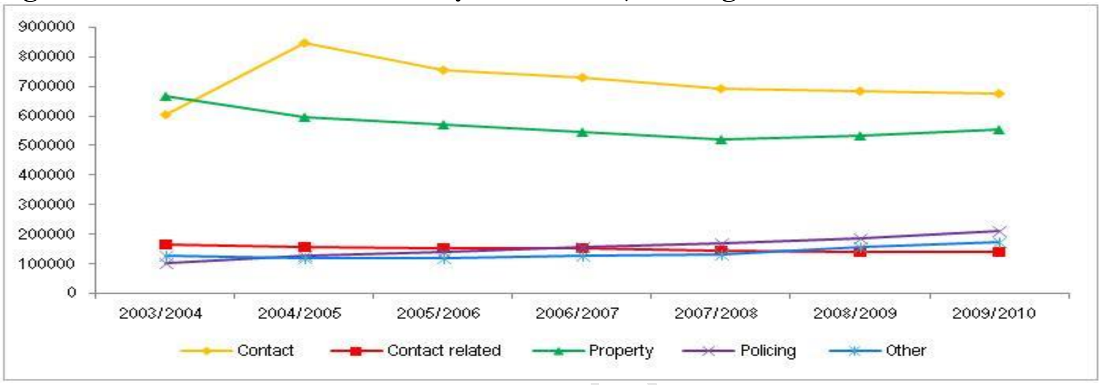


Figure 4.2: Total Number of Crime by the Five Major Categories in South Africa



Source: SAPS (2010)

Evidently and interestingly, this pattern is reflected throughout South Africa as depicted by figure 4.2 which shows *contact crimes* to be the most prevalent[89].

This section gives a brief overview of the roots of social disorganisation in South African communities and highlights the current state of South African communities in relation to social disorganisation. The section mainly follows the works of Emmett (2003); Breetzke (2010); South African Institute of Race Relations (SAIRR) (2010); Budlender and Lund (2011) and Schatz *et al.* (2011). The literature points to two main intertwined sources of social disorganisation namely South Africa's apartheid past and the HIV/AIDS epidemic. We highlight below the role of these issues in social disorganisation in South African communities.

The consensual view asserts that the apartheid era scarred South Africa socially (Mare 2001; Emmett 2003; Breetzke 2010; Lau *et al.* 2010; Daniels and Adams 2010; John 2010; Schatz *et al.* 2011) mainly because the political context was favourable in facilitating the breakdown of the formal and informal

[89] Of the 2.1 million crimes reported in the 2009/2010 period 31.9% were *contact crimes*; 26.1% were *property crimes*, 25.5% were *other crimes*, while 10.0% were *policing crimes* and the remaining 6.5% were *contact related crimes* (SAPS 2010).

social networks and associational ties inherent in the social structure of South African communities[90] (Emmett 2003; Ward 2007; Breetzke 2010; Daniels and Adams 2010; Lau, 2010; Mathews *et al.* 2011).

More broadly, the apartheid system was characterised by for example racial segregation policies, the migrant labour systems and hostels, pass laws and the Group Areas Act, the Bantu Education Act, school boycotts and popular uprisings. Consequently, this sparked a breakdown of traditional family structures in South Africa and increased separations within households, single parenthood, and isolation of nuclear families from extended families (Emmett 2003; Daniels and Adams 2010; Budlender and Lund 2011; Schatz *et al.* 2011). Further to this the segregation policies and a series of forced removals engendered the spatial separation of residents from their communities (Mare 2001; Emmett 2003; Breetzke 2010). The historical segregation policies also caused division between and within racial groups, cultivating mistrust, fear and breakdown in local ties and friendships[91] (Emmett 2003; Vincent 2008; John 2010; Ward and Bakhuis 2010).

The migrant labour system in itself cultivated a *non-traditional family* in South Africa (Emmett 2003; Whyte 2010; Budlender and Lund 2011). By the late 1980s it was estimated that 40% (approximately 2 million) of Black workers were migrant labourers, which is the highest ever recorded[92] (Emmett 2003). This paved the way for a culture of absentee fathers and family desertion as migrant workers had multiple partners and secondary families[93] (Emmett 2003; Whyte 2010; Budlender and Lund 2011; Mathews *et al.* 2011). The result is increased poverty among women (the majority of the migrant workers were male) inevitably forcing women to migrate as well in search of employment. The residual effects of the migrant labour system were a change in composition in the South African household and parents grappling for control over their children (Emmett 2003; Mathews *et al.* 2011).

[90] Social and medical scientists have long documented the social disintegration of black communities in South Africa under the assault of colonialism, apartheid and racial oppression. Among the symptoms of disintegration, with increasing divorce rates, separations, single parenthood and teenage pregnancy: breakdown of authority of parents and teachers, high unemployment rates, low performance in all spheres of life, including school, high crime and violence rates at all levels of social interaction (Emmet 2003: p. 5). Structures of apartheid which served to fragment, dislocate and alienate the individual from a sense of community, belonging, identity, and association with language and cultural groups in part resulted in 'fractures in *humanness*', as manifested in the high prevalence of violence in South Africa (Lau 2010: p.7). South African townships are deprived and poor, plagued by social disorganisation caused by apartheid policies such as forced removals, which contributed to family breakdown and the rapid formation of street gangs (Mathews *et al.* 2011: p. 961).

[91] There is a sense that the physical violence of the past remains as long-term socio-psychic violence and manifests in social division, suspicion and lack of trust (John 2010: p. 59).

[92] There is no other country where such a system has existed for so long and has trapped so large a proportion of the labour force in a dehumanising structure (Wilson *et al.* 1989 cited in Emmett 2003: p. 5).

[93] For Blacks and Coloureds in South Africa, apartheid has had a devastating impact on family life. Studies on fatherhood have shown that the migrant labour system particularly impacted on black men's availability for their families, resulting in fathers abandoning and neglecting their children. With child-rearing practices gendered and largely considered as womens' domain... however, many children are raised not only without a father, but also without a biological mother, with grandmothers or aunts taking on the parenting role… Given this context children turn to others in their social environment for affirmation,…including peers which leads to exposure to gang culture with its attendant violence (Mathews *et al.* 2011: p. 961).

Additionally, the migrant labour system created a favourable environment in the country for HIV/AIDS to flourish. Needless to say the situation further fuelled an increase in single headed and child headed households as many perished from the AIDS epidemic (Emmett 2003; Whyte 2010; Ragnarsson *et al.* 2010; Schatz *et al.* 2011; Kenyon and Zondo 2011). Consequently, South Africa currently has one of the highest HIV/AIDS prevalence in the world (UNAIDS 2009; Mah 2010a; Mah and Halperin 2010b; UNAIDS 2010).

It has therefore become the 'norm' to have single female headed households and absentee fathers in South Africa[94] (SAIRR 2011; Schatz *et al.* 2011; Mathews *et al.* 2011). Currently, the country faces disturbingly high rates of *family disruption*[95] (Whyte 2010; Schatz *et al.* 2011; Budlender and Lund 2011; SAIRR 2011). Further to this, there is extensive practice of *formal and informal polygamy* and multiple concurrent partnerships in South Africa (Mah 2010a) which also affects household composition. In a recent study by SAIRR on 'First Steps to Healing the South African Family', the main findings were that increasingly South African children are growing up with only one parent. The same pattern is observed in the Cape Area Panel Study (CAPS) household data, where we find in wave 1 (2002) that 71% of households were female headed, in wave 3 (2005) this increased to 77%, and in wave 4 (2006) to 81%. Additionally, more than 50% of these household heads are either divorced or widowed or never married (never married forms the highest frequency among the single heads of households). Table 4.B1 of Appendix 4.B summarises the current state of South African families.

Amidst these there has been a consistent *lack of basic services* for most of the country but particularly in impoverished communities (Emmett 2003; Breetzke 2010; Alexander 2010; Lau *et al.* 2010). A number of studies have suggested that crime in South Africa cannot be eliminated without provision of basic services (Silber and Geffen 2009; Hough 2009; Vromans *et al.* 2011). The vestiges of apartheid have influenced this vast differences in community resources amongst which the most controversial is basic services[96] (Goldin 2010). This is because South African communities were typified by racial division with Black and Coloured groups facing desolate conditions and having the least resources allocated towards their communities.

These effects have spilled over so that even at present the Black communities are the most impoverished and lacking in basic services (Emmett 2003; Breetzke 2010; Goldin 2010). Recent riots

[94] The latest available data about fathers in South Africa shows that the proportion of fathers who are living but absent increased between 1996 and 2009, from 42% to 48%. Boys growing up in households with absentee fathers are more likely to display 'hyper-masculine' behaviour, including aggression. These findings correspond with research from United States, where it was found that absence of fathers when children were growing up was associated with poor educational outcomes, anti-social behaviour, delinquency, and disrupted employment in later life (SAIRR 2010: p. 4).

[95] It is stated that the South African household composition is incongruent with the Western nuclear household (Emmett 2003; Schatz *et al.* 2011; Budlender and Lund 2011).

[96] The historical patterns of access to water and other areas of public service delivery in South Africa have been markedly skewed. Despite the reversal of the regime and the fact that South Africa is a middle-income country, there are a significant number of people who are water-poor (Goldin 2010: p. 195).

during the 2011 South African local government elections re-emphasised this *lack of basic services* which thus far the South African government has failed to provide especially to informal settlements inhabited mainly by the Black population[97]. This is coupled by high mobility within the country. The impetus for *residential instability* is the search for employment opportunity as *unemployment* and poverty remain staggering high (Breetzke 2008). South Africa has one of the highest unemployment rates globally (Blackmore 2003; Klasen and Woolard 2008; SAIRR 2010; Davies and Thurlow 2010; Tregenna 2011), mainly concentrated among the Black population (Klasen and Woolard 2008).

4.4 Theoretical Framework

This section presents a simple theoretical framework for criminal behaviour that supports our hypothesis. The theoretical framework follows Fajnzylber et al. (2002) and Haddad and Moghadam (2011) who extended Becker's theory (Becker 1968) which is grounded in the rational utility maximising framework. Becker's (1968) influential paper 'Crime and Punishment: An Economic Approach' initiated several theoretical models of criminal behaviour in economic literature. According to Becker's theory the decision to commit crime is rational and is based on the expected costs and benefits of committing a crime. The expected cost in this case is the charge/sentencing given to criminals that are apprehended, while the expected benefit is the difference between the wealth from the crime and the opportunity cost of the crime. To date Becker's theory has had numerous extensions, for example Ehrlich (1973); Heineke (1978); Witte (1980); Fajnzylber et al. (2002); Becker et al. (2006); Mehlkop and Graeff (2010); Miyoshi (2011); and Haddad and Moghadam (2011). The extensions that are of relevance to the theme of this paper are the research works of Fajnzylber et al. (2002) and Haddad and Moghadam (2011). Fajnzylber *et al.*, (2002) extended *Becker's theory* by making four main extensions in the core model. Firstly, the core model excludes deterrence factors because of limited cross-sectional data. Secondly, the model includes the effects of illegal drugs. Thirdly, demographic factors which include urbanisation and age composition are incorporated in the model. The fourth and final extension is to include two cultural factors which are geographic region and religion dummies. With these extensions, Fajnzylber *et al.*, (2002) use a panel data of 45 Western industrialised countries, Latin America and the Caribbean, Eastern Europe and Central Asia, the Middle East and North Africa and Asia (excluding sub-Saharan Africa) to test their hypothesis. Fajnzylber *et al.*, (2002) re-estimate a second model in which they include deterrence factors. Their argument for this approach is based on the fact that significance of the income variable may be a result of unobservables such as unequal distribution of crime prevention efforts, and that this effect is likely to be eliminated by using deterrence proxies. The deterrence proxies that they use include the

[97] Silber and Geffen (2009) state that crime cannot be eliminated without providing basic services such as sanitation and water. Hough (2009) notes that violence is likely to occur in deprived groups as a result of these groups comparing themselves with groups that are better off. Lau *et al.* (2010) states that protests occur in wards that have higher unemployment and lack access to services, while Vromans *et al.* (2011) asserts that violence is a result of post-apartheid expectations of the disadvantaged not being met.

number of police personnel per 100,000 population and the second proxy is the existence of the death penalty. Similarly, Haddad and Moghadam (2011) also estimate an extension of *Becker's theory* in Iran using panel data on provinces. Specifically, their model includes the following: deterrence (the probability of arrest which is defined as the number of arrested criminals over number of reported crime), economic factors (unemployment and income inequality), social factors (family instability and literacy rate) and demographic factors (population density). The extensions that are of relevance to the theme of this paper are the works of Fajnzylber *et al.* (2002) and Haddad and Moghadam (2011). Hence staying in the Beckerian setting and following Fajnzylber *et al.* (2002) and Haddad and Moghadam (2011) our theoretical framework is based on two core assumptions: firstly individuals act rationally in their decision to commit crime using cost benefit analysis and secondly individuals are assumed to be risk neutral. The expected *net benefit* (nb) of committing crime is equal to the probability of not being *apprehended* $(1-\mathrm{pr})$ multiplied by the *loot* (l), less the total *cost of executing the crime* (c), less forgone wages from *legitimate activities* (w). From this we then subtract the expected punishment from committing crime $(\mathrm{pr} * \mathrm{pu})$:

$$\mathrm{nb} = [(1-\mathrm{pr}) * l - \mathrm{c} - \mathrm{w}] - [\mathrm{pr} * \mathrm{pu}] \tag{4.1}$$

In order for an individual to commit crime $(\mathrm{d}=1)$ or not commit crime $(\mathrm{d}=0)$ the expected *net benefit* needs to exceed the *moral stance* (m). The *moral stance* incorporates moral values, and possesses the threshold of moral value beyond which a crime will be committed.

$$\mathrm{d} = 1 \text{ when } \mathrm{nb} \geq \mathrm{m}$$

$$\mathrm{d} = 0 \text{ when } \mathrm{nb} < m \tag{4.2}$$

The presence of *family disruption* (fdi) implies that the formal and informal associational ties which assist in collective supervision, social support and social control at household and community level are broken. When this occurs there is likely to be a lower *moral stance* among individuals which in turn increases the likelihood of them committing a crime. Furthermore *family disruption* is also likely to decrease the *cost of executing the crime*, since young individuals who are not supervised are more likely to become involved in street gangs and hence have a higher chance of engaging in risky behaviours. Closely aligned to *family disruption* is *residential instability* (res). Interestingly, *residential instability* disturbs the social networks of friendship and kinship in a community and consequently reduces the positive social relations among community members. This suggests that *residential instability* is likely to create a conducive environment for criminal behaviour as it reduces the *cost of executing crime* and *moral stance*.

By the same token, an increase in the level of *community disadvantage* (dis) has the effect of lowering the community's *moral stance* and decreasing *legitimate activities* since it captures and takes advantage of the characteristics of a community with social decay. This notion is based on the premise that communities with low socioeconomic status lack adequate resources to participate in both formal and informal organisational activities that are designed to consolidate and reinforce community structures. High *unemployment levels* (uem) may be associated with lower levels of *legitimate activities* and thereby is likely to increase the probability of criminal behaviour. The *lack of basic services* (abs) implies the absence of resources required for basic survival as such this is likely to decrease *moral stance* and limit *legitimate activities*.

Therefore the effects of the aforementioned social factors[98] on the *cost of executing crime* can be formally written as $\frac{\Delta c}{\Delta fdi} < 0$, and $\frac{\Delta c}{\Delta res} < 0$, while the effects on the *moral stance* is $\frac{\Delta m}{\Delta fdi} < 0$ and $\frac{\Delta m}{\Delta res} < 0$. On the other hand the effects on the *loot* (*l*) can therefore be formally represented by $\frac{\Delta l}{\Delta uem} > 0$, $\frac{\Delta l}{\Delta dis} > 0$, and $\frac{\Delta l}{\Delta abs} > 0$, while the effects on *legitimate activities* (w) is $\frac{\Delta w}{\Delta uem} < 0$ and $\frac{\Delta w}{\Delta dis} < 0$ and finally the effects on the *moral stance* (m) is $\frac{\Delta m}{\Delta uem} < 0$, $\frac{\Delta m}{\Delta dis} < 0$ and $\frac{\Delta m}{\Delta abs} < 0$. Recall that, fdi is *family disruption*, res is *residential instability*, uem is the *unemployment level*, dis is the *community disadvantage,* while abs is the *lack of basic services*. Substituting these factors into equation (4.1) and (4.2) yields equation (4.3) which depicts the condition under which an individual is likely to commit a crime:

$$d = 1 \text{ if } l(uem, dis, abs) - c(fdi, res) - w(uem, dis) - m(uem, dis, abs, fdi, res) \quad (4.3)$$

Equation (4.3) can be rewritten as a function (f) of the social factors in a reduced form as depicted in equation (4.4).

$$d = 1 \text{ if } f(uem, dis, abs, fdi, res) = f(\psi) \geq 0 \quad (4.4)$$

Where ψ represents the factors that determine crime and $\frac{\Delta f}{\Delta uem} > 0$, $\frac{\Delta f}{\Delta dis} > 0$, $\frac{\Delta f}{\Delta abs} > 0$, $\frac{\Delta f}{\Delta fdi} > 0$ and $\frac{\Delta f}{\Delta res} > 0$. We further assume that the decision to commit crime is based on a linear probability model and that f has a linear functional form. The individual equation then takes the form as portrayed in (4.5).

$$d = \beta'\psi + \mu \quad (4.5)$$

Equation (4.5) is an individual regression equation, however as previously mentioned, the current data at our disposal is aggregated crime data at community level and not individual level. Miyoshi (2011) lists a number of criminology studies that have experienced this obstacle and have instead relied on aggregated units of analysis such as country or province as opposed to analysis at individual level. The study therefore adopts a similar approach and follow Fajnzylber *et al.* (2002), and Haddad and

[98] The social factors that we include are embedded in the international evidence of significant interaction between socially disorganised communities and community crimes (Sampson and Groves 1989; Kawachi *et al.* 1999).

Moghadam (2011) and convert equation (4.5) to an aggregate community level regression by obtaining the average of all individuals in community (i) in a given period (t), giving rise to equation (4.6):

$$D_t = \beta'\psi_t + \mu_t \quad (4.6)$$

Substituting ψ in equation (4.6), namely factors that determine crime, gives as the following equation:

$$D_{it} = \alpha + \beta_1 . dis_{it} + \beta_2 . abs_{it} + \beta_3 . fdi_{it} + \beta_4 . res_{it} + \beta_5 . uem_{it} + \mu_{it} \quad (4.7)$$

4.5 Empirical Framework

The main purpose of this section is to describe the data and our estimation methods. Specifically, section 4.5.1 introduces the data used to test our model in equation (4.7). Section 4.5.2 defines the outcome variable and the regressors used in our estimation model. Thereafter we proceed to section 4.5.3 which describes the likely estimation challenges in our model. We then end by outlining the strategy to overcome the estimation challenges.

4.5.1 Area of Study and Sources of Data

The study employs data from the Cape Area Panel Study (CAPS)[99] and Cape Town crime statistics from the South African Police Service (SAPS). CAPS is a longitudinal study of young adults in the Metropolitan Cape Town area of South Africa, the first wave was conducted in 2002 with the most recent wave, wave 4, carried out in 2006. The survey focuses on health, education, employment, family formation and intergenerational support systems of the young adults. The data which is the primary source of our regressors is the young adult data and household data.

The crime statistics from the South African Police Service (SAPS) include statistics on 33 types of crime that occurred in the Metropolitan Cape Town area between 2001/2002 and 2005/2006. Table 4.A1 in Appendix 4.A shows the distribution of the categories across the five years. We match the CAPS young adult dataset and household dataset of waves 1 (2002), 3 (2005) and 4 (2006), to the 2001/2002, 2004/2005 and 2005/2006 crime data. As the crime statistics are grouped at the community (sub-place) level we also aggregate all the CAPS data at community level. Hence the unit of analysis in this study is the community (defined by sub-place).

4.5.2 Variable Specification

The section defines each of the variables used in our model which attempts to explain the impact of social disorganisation on contact and contact related crime. The model has two outcome variables:

[99] http://www.caps.uct.ac.za or http://www.datafirst.uct.ac.za

contact crime and *contact related crime*, while the regressors include *family disruption*, *community disadvantage*, *residential instability*, *lack of basic services* and *unemployment*.

4.5.2.1 Contact and Contact Related Crime

As indicated earlier this study uses two types of outcome variable, namely *contact crime* and *contact related crime*. Our motivation for using *contact crime* and *contact related crime* is mainly because these crimes are a result of the underlying social structures, while other crimes (*property related crimes*, *crimes detected by the police* and *other serious crimes*) are mainly committed for economic reasons (SAPS 2010). Furthermore, because the determinants of crime are likely to vary by the type of crime (Breetzke 2010; Miyoshi 2011), we test the two types of crimes separately. Below we describe each of the five main types of crime as defined by the South African police service (SAPS)[100].

Contact crime involves physical contact between the perpetrators and the victims of a crime. This category of crime is also known as *social contact crime* because it often occurs between individuals who are familiar with each other. Thus the types of crime that fall into this category mainly involve crimes of social or group behaviour. The category includes the following types of crime: murder, attempted murder, rape, all forms of assault, and all forms of robbery.

Contact related crime usually occurs as a result of individual or collective behaviours, for example crimes that occur as a result of frustration. The types of crime that falls in this category include: malicious damage to property, arson, public violence[101] and *crimen injuria*[102]. *Property related crimes* are usually classified as non-violent crimes as they occur in the absence of the victims. Burglary at residence and businesses, theft of motor vehicle and motorcycle, theft out of motor vehicle and stock theft are the types of crimes that fall into this category. *Crimes detected by police* are the fourth type of the broad category of crime. The crimes in this category are a result of police proactive action such as road blocks and random searches. The crimes in this category include illegal possession of firearms, drug related crimes and driving under the influence. The fifth and last category is the *other serious crimes* category which includes commercial crime, shoplifting and all theft not mentioned elsewhere.

4.5.2.2 Family Disruption

There are various proxies that have been used for *family disruption*. For example Sampson (1989) used an index consisting of the percentage of divorced/separated adults and percentage of single parents as

[100] According to the SAPS 2009/2010 report, the various crimes in South Africa are grouped into five categories which include: contact crime, contact related crime, property related crime, crimes detected by police and other serious crimes.

[101] This includes all forms of contact crime that occur in public places.

[102] This includes actions that impair the dignity or privacy of a person, for example offensive language or psychological abuse.

a proxy for *family disruption*. On the other hand, Breetzke (2010) used divorce/separation, father deceased/estranged and female headed household in the analysis. Lastly Haddad and Moghadam, (2010) included the number of divorces in the province as one of their regressors. In our view the above mentioned indicators are unlikely to be good proxies for *family disruption* in the case of South Africa. This is because and as previously mentioned absentee fathers and female headed households have become the 'norm' in South Africa SAIRR 2011; Schatz *et al.* 2011; Mathews *et al.* 2011), where the family unit differs from the conventional international 'nuclear family'. Breetzke (2010) for instance found female headed household to be insignificant in predicting crime in Tshwane, South Africa.

This study therefore uses the frequency of interaction between the youth and both parents as a proxy for family disruption, where the lower the frequency of interaction with both parents the higher the family disruption. Instead of capturing whether the household has a single female parent, which we have highlighted to be the 'norm' in South Africa, we rather focus on the amount of time parents spend with the youth. This proxy is obtained from the following two questions in the CAPS survey: 'how often has mother spend the night in the past 12 months' and 'how often has father spent the night in the past 12 months'. The response choices to these questions included: 'never', 'rarely', 'sometimes' and 'often'. Family disruption is represented by the percentage of youth who never and rarely see both parents.

4.5.2.3 Community Disadvantage

There are various factors that have been used as proxies for community disadvantage in empirical research, as recorded in the literature. For example Hannon and Defronzo (1998) include the percentage of female headed households, percentage of black population, and family poverty rate. Similarly Gorman *et al.* (2001) include the unemployment rate, the welfare rate and household income as measures of neighbourhood poverty. On the other hand Sampson *et al.* (1989) include education level, which was measured as the percentage of community members with college education, occupation measured as the percentage of respondents in professional and managerial positions and income represented by the percentage of those with high income. On the other hand Fajnzylber *et al.* (2002) includes GNP per capita and income inequality in their model, and control for reverse causality[103] using lagged variables. Breetzke (2010) in turn uses socioeconomic deprivation captured by share of income. Reverse causality may arise in models like these given that at community level, communities with low income levels are likely to be infested with crime and other socioeconomic problems. High crime levels in itself, are likely to repel high income earners from residing in such communities, while low income earners mostly do not have any alternative but to stay in such areas.

In it is worth noting that in recognising the reverse causality between income and crime we used a proxy which is *mean years of education of a community*[104], where the lower the mean the more the disadvantaged the community, and the average *ratio of number of rooms to number of household members*[105] where lower ratios represent more disadvantaged households, and therefore a more disadvantaged community. However, education is also likely to be endogenous, see Machin et al (2004); Locher (2004); Buonanno and Leonida (2006) and Groot and Brink (2010) for this discussion. A possible channel of this endogeneity is reverse causality, for instance, low education is likely to increase criminal behaviour and by the same token high criminal behaviour is likely to influence investment in education. Further to this, communities with high crime rates are more likely to be less attractive to more educated individuals since they are likely to have more opportunities and thus have a wider choice of place of residence. We further acknowledge that the *ratio of number of rooms to number of household members* is also likely to be endogenous. One likely route of this endogeneity is a scenario where abundant houses in crime infested communities are more likely to be occupied by squatters, thus increasing the *ratio of number of rooms to number of household members*.

With these caveats in mind we consider an alternative proxy measure which is the *ratio of number of rooms to number of household members* where we only include households who either own or rent the house, so as to purge out the squatter effects. Upon exploring our data we observe that the majority of the households either own or rent their households. Specifically, the data in the first wave show that 71.8% of the households own the house, while 24.4% rent the house and the remaining 3.7% mentioned 'other'. This 'other' category mainly consists of individuals who stay for free and as such this category is likely to contain the squatters described earlier on. Along similar lines, in the third wave, 74.1% of the households own the house, while 24.4% rent the house and those who stated 'other' comprised of only 1.2% of the households. Finally, in the fourth wave and as in the previous waves the majority of the households own the houses (75.7%), while the minority rent the houses (23.4%), with the remaining 0.7% mentioning 'other'. Overall, we found our results marginally different when using the two income proxies, namely the *ratio of number of rooms to number of household members* and the *ratio of number of rooms to number of household members* after including only households who either own or rent the house. Finally, it is worth mentioning that the *ratio of number of rooms to number of household members* is likely to be correlated with various other housing related factors such as the demand for housing, apart from income, which may in turn drive our results. Hence although the results are suggestive they are however not conclusive and this must be taken into consideration when interpreting the results.

104 This is derived from the variable: 'number of years of education' in the CAPS data, from this variable we calculate the average number of years of education of community members.

105 This indicator is derived from two variables, one being 'total number of household members that reside in the house' and 'total number of rooms contained in the house'. From this information we divide the number of rooms by household members to derive our ratio.

4.5.2.4 Residential Instability

Sampson and Groves (1989) uses the percentage of the population that grew up within 15 minutes of their current home (birth place) as a proxy for residential stability, while Gorman *et al.* (2001) uses the length of residential tenure and the proportion of the population that has moved in a five year interval and one year interval as a proxy for residential stability. Breetzke (2010) includes the percentage of Black immigrants, that is, African citizens who come from African countries outside of South Africa. Haddad and Moghadam (2010) use the number of immigrants as one of their regressors. Our proxy measure for residential instability is the proportion of individuals who have *recently moved*. Our motivation for using *recently moved* is the lack of availability of other variables that can be used as proxies for residential instability.

4.5.2.5 Lack of Basic Services

Breetzke (2010) includes a deprivation index which consists of five variables: the percentage of the population in informal settlements, the percentage of households that lack flush toilets, the percentage of households without water, the percentage of households without electricity and lastly the percentage of households without refuse removal. Following Breetzke (2010) our proxy measure includes three variables to capture *lack of basic services* including the percentage of households that do not have access to water[106], the percentage of household with no sanitation[107] and the percentage of households without access to electricity.

4.5.2.6 Unemployment

Previous empirical studies, for instance that of Breetzke (2010) include the proportion of the unemployed population who are 15 years of age and older. While Miyoshi (2011) and Haddad and Moghadam (2011) use the unemployment rate. Our proxy measure for unemployment is the percentage of households where the head of the household is unemployed.

4.5.3 Empirical Model Specification

Equation (4.7) is our econometric model that measures the relationship between community social disorganisation and contact and contact related crimes. The likely challenges facing our econometric model include: excessive collinearity among the regressors (Hannon and Defronzo 1998), measurement error which is highly prevalent in police crime data as a result of reporting bias (Fajnzylber *et al.* 2002; Demombynes and Ozler 2005; Louw 2007), omitted variable bias, and the

[106] This proxy is created by calculating the percentage of households who do not have water in their houses or in their yard. These households mainly depend on carrier tanks or public tap/kiosk or boreholes as their source of water.

[107] We obtain this proxy by aggregating across number of household who use pit latrines, bucket toilets or any other means that does not involve the use of a toilet for their sanitation needs, we then calculate the percentage of these households at the community level.

influence of community unobservables (Fajnzylber *et al.* 2002; Blackmore 2003). Another estimation challenge worth mentioning is reverse causality. Recall that section 4.5.2 controlled for the reverse causality between community disadvantage and *contact crime* and *contact related crime* by using a *community disadvantage* proxy that does not contain income but rather captures income effects.

These challenges have been identified as some of the causes of the variation in results for this type of research, especially in cross-sectional analysis (Hannon and Defronzo 1998; Fajnzylber *et al.* 2002; Blackmore, 2003). Table 4.D1 in Appendix 4.D shows the estimation challenges that are likely to occur in the model. High multicollinearity is a result of theoretically related regressors in models of social determinants of crime, such that the regressors approach a near perfect linear relationship with each other. The presence of high multicollinearity among the regressors implies that the coefficients are likely to have large standard errors and hence low levels of significance, even in cases where the coefficients are significant. High collinearity may also affect the sign (wrong or unexpected sign) and size of the coefficient (Greene 2002; Wooldridge 2002). Table 4.C1 and 4.C2 in Appendix 4.C show the high collinearity between the variables used to construct a *community disadvantage* index and those that were employed in the *lack of basic services* index respectively[108]. To cushion against high collinearity we construct indices of the related variables, as proposed by among others Greene (2002)[109] and following the work of Hannon and Defronzo (1998) and Breetzke (2010). Hence the *community disadvantaged* index consists of *mean years of education of a community*, *percentage of the population that is non-white*, and the *average ratio of number of rooms to number of household members*. The *lack of basic services* index is composed of the *percentage of households that do not have access to water*, *percentage of household with no sanitation* and *percentage who do not have access to electricity*. To determine whether we have indeed controlled for high collinearity we perform a simple test that detects the presence of multicollinearity: by computing a correlation matrix (Cameron and Trivedi 2005). Table 4.C3 in Appendix 4.C shows the correlation matrix of our regressors. It is important to note that some of our regressors are related, for example people who have flush toilets are likely to have running water. This implies that these regressors are highly correlated, with the correlation coefficients ranging from 0.6 to 0.8. In an effort to curtail this we developed indices (see Greene, 2002; Cameron and Trivedi, 2005) of related variables (for example flush toilet and running water). This had the effect of reducing the correlation coefficients to less than 0.3 which is lower in comparison to the previous correlations among our regressors prior to constructing the indices. We do acknowledge the fact that using indices as regressors introduces a lack of clarity into what is being estimated by these regressors (see Greene, 2002), as such this caveat should be taken into consideration when interpreting the results.

[108] For example table 4.C1 in appendix 4.C shows that the correlation between ratio of rooms to household members and education is 0.5544. Furthermore table 4.C1 also shows the pairwise correlation between the community disadvantage index and mean average income is positive and very high at 0.8302, therefore confirming our earlier statement: the community disadvantage proxy is likely to also capture the income effects.

[109] Hannon and Defronzo (1998), Fajnzylber *et al.* (2002) and Breetzke (2010) also construct indices to eliminate the high collinearity in their respective studies.

The second likely estimation challenge is measurement error in the outcome variable. This type of error is not necessarily harmful to the estimators depending on whether the error is correlated or uncorrelated with the regressors. If the measurement error is uncorrelated with the regressors, this does not cause any bias in the coefficient. The only likely effect is to inflate the standard errors. However if the measurement error is correlated with the regressors then the coefficients become biased and inconsistent (Greene 2002; Bricker and Engelhardt 2008; Benoit *et al.* 2009).

We follow Wooldridge (2002) and Bricker and Engelhardt (2008) to illustrate this. Recall from equation (4.7) that D_{it} is the reported/police crime statistics. Let $D^*{}_{it}$ be the true crime statistics, such that $D_{it} \neq D^*_{it}$. The measurement error is therefore the difference between the true crime statistics and the reported/police statistics: $\varepsilon_{it} = D_{it} - D^*_{it}$. Where ε_{it} is the measurement error, this equation can be re-written as $D_{it} = D^*_{it} + \varepsilon_{it}$, substituting this into equation (4.7) we obtain: $D_{it} = \alpha + \mathbf{X}\beta + \mu_{it} + \varepsilon_{it}$ where $\mathbf{X}$ is a vector of regressors. In the case where $\mathrm{cov}(D^*_{it}, \varepsilon_{it}) = 0$, the measurement error is uncorrelated with the true crime statistic. The coefficients are consistent although inefficient as long as measurement error is not correlated with the regressors ($\mathrm{cov}(\mathbf{X}, \varepsilon_{it}) = 0$). Hence in this scenario one would proceed with the estimating equation (4.7), and ignore the fact that our dependent variable is imperfect (Wooldridge 2002; Bricker and Engelhardt 2008). However, when $\mathrm{cov}(D^*_{it}, \varepsilon_{it}) \neq 0$ that is, the measurement error is correlated with the true crime statistics and that relationship is of the form: $\varepsilon_{it} = \delta D^*_{it} + \gamma_{it}$ then equation (4.7) becomes: $D_{it} = \alpha + \mathbf{X}(1 + \delta)\beta + \mu_{it} + \varepsilon_{it}$. Hence in this second scenario the coefficients will be biased (proportional bias of δ) and inconsistent. Using these two types of scenarios outlined above we determine the types of crimes that we can ignore in the presence of measurement error and the ones we cannot ignore. Specifically we are looking for those crime statistics where the measurement error is unsystematically related to the explanatory variables (this measurement error can be ignored), and the measurement error that is systematically related to the explanatory variables (cannot be ignored).

The main channel of measurement error in crime literature is *reporting bias* which occurs as a result of underreporting of crime occurrence, and has been identified as a pervasive problem of police crime statistics (Fajnzylber *et al.* 2002; Blackmore 2003; Demombynes and Ozler 2005). The types of crimes that are more likely to be afflicted with *reporting bias* are robbery and rape[110] (Fajnzylber *et al.* 2002; Demombynes and Ozler 2005). For example in South Africa members from economically advantaged communities are more likely to report cases of robberies than those from disadvantaged communities, mainly because the latter communities mistrust the police force. In addition the disadvantaged communities are more likely to choose vigilante justice than to report any crime that has been committed (Blackmore, 2003).

[110] It is worth noting that the crime that has the least *reporting bias* is murder, and this is because it is impossible to ignore a deceased individual (Fajnzylber *et al.* 2002; Demombynes *et al.* 2005; Louw 2007; SAPS, 2010).

In the case of rape there is more likely to be under reporting given the sensitive nature of sexual crimes and the stigmatisation that goes along with this type of crime (Fajnzylber *et al.* 2002; Nleya and Thompson 2009; Clay-Warner and McMahon-Howard 2009; Wolitzky-tailor *et al.* 2011). A similar pattern is observed in South Africa where most rape cases go unreported (Nleya and Thompson 2009; Wolitzky-tailor *et al.* 2011)[111]. In addition it is stated that rape reporting differs by the demographic and socioeconomic characteristics of victims (Clay-Warner and McMahon-Howard 2009). It is clear from this that these crimes exhibit a systematic relationship between the existing measurement error and the regressors. This is mainly because under reporting is likely to be higher among disadvantaged communities, as such we cannot ignore the measurement error in these crimes.

Two approaches are used to address measurement error[112]. It is worth noting that this error only affects the *contact crime* outcome variable and not the *contact related crime*, and this is because *contact crime* includes rape and robbery. The first is the 'naive approach' where we constrain the analysis to include only those crime statistics that are not prone to measurement error. In this study this involves excluding rape and robbery from *contact crime*. The second approach follows the procedure of Demombynes and Ozler, (2005) who controlled for measurement error by adjusting the crime statistics with the probability of reporting crime which was derived from data from the 1998 Victims of Crime Survey (VCS)[113]. Using the VCS we regress the probability of reporting crime on a set of household characteristics common to VCS and CAPS, that is, $\ln C_h^{VCS} = \beta^{VCS} X^{VCS} + e_h$, where X is household characteristics common in both VCS and CAPS. Using the estimated coefficients we then impute the probability of reporting crime, $\widehat{\ln} C_h^{CAPS} = \hat{\beta}^{VCS} X^{CAPS}$, thereafter we predict the probabilities[114].

The third likely estimation challenge that is often ignored in crime literature relates to the *mobile* nature of crimes (Breetzke 2010). This mobility is often found in crimes that are committed for economic purposes as opposed to social crimes. Such economic crimes include house robbery, bank robbery, and cash in transit heists (shaded grey in Table 4.D1 in Appendix 4.D). The *mobile* nature of these

[111] Existing studies have shown that some of the barriers to reporting rape include demographic factors, socioeconomic conditions and stigma (Fajnzylber *et al.* 2002; Clay-Warner and McMahon-Howard 2009; Wolitzky-tailor *et al.* 2011).

[112] All the previous studies that are closely related to the current study acknowledge the presence of measurement error; however different approaches were used to overcome this econometric challenge. Sampson *et al.* (1989), Hannon and Defronzo (1998) and Breetzke (2010) do not accommodate measurement error. Fajnzylber *et al.* (2002) only uses homicides and robbery, stating that these two crimes are least likely to have measurement error, although with regard to robbery this is questionable as we outlined in section 4.4.

[113] The 1998 VCS is the first nationwide survey in South Africa to collect information on crime incidence, crime reporting, the public's perception of police, and the existing support structures for victims of crime. The data was collected from 3899 households and 12167 individuals throughout the nine province of South Africa (Stats SA, 2001).

[114] Table 5.A in Appendix 5 shows the Probit regressions that were used to impute and predict the probability of reporting crime. Using the merged VCS and CAPS household survey data we regress *household reporting of crime* on household characteristics common in both datasets and include: level of schooling, gender, age, household income and province. The regression output shows that reporting of crime differs by type of crime and in general by gender of victim where it is likely to be higher among men than among women.

crimes makes it difficult to establish their relationship to community social or economic characteristics as these crimes are not area specific. For example, crimes may be committed in area *i* by members from area *j*. Hence the characteristics of area *i* are unlikely to pick up the effects of the crime committed in area *i*, as the characteristics only include those of area *i* when they should also include those of area *j*[115].

We illustrate the likely result of omitting the variables following Cameron and Trivedi (2005). Recall our model: $D_{it} = \alpha + \mathbf{X_i}\beta + \mu_{it}$. Where D_{it} is the crime statistic that is prevalent in community *i* in time t and $\mathbf{X_i}$ is a vector of regressors representing the structural characteristics in community *i*. Suppose that the crime that has occurred in community *i* is economically driven, representing for example house robberies. Then as argued above, the equation would be of the form $D_{it} = \alpha + \mathbf{X_i}\beta + \mathbf{X_j}\varphi + \mu_{it}$ where $\mathbf{X_j}$ represents characteristics from community *j*. However, if we do not include the characteristics from community j it implies that we are instead estimating the following equation $D_{it} = \alpha + \mathbf{X_i}\beta + (\mathbf{X_j}\varphi + \mu_{it})$ where $\mathbf{X_j}\varphi + \mu_{it}$ is the new error term. The regressors $\mathbf{X_i}$ will be correlated with the error term ($\mathbf{X_j}\varphi + \mu_{it}$) if $\mathbf{X_i}$ is correlated with $\mathbf{X_j}$. Hence our estimated coefficients will be inconsistent due to the omitted variable bias.

The omitted variable bias causes endogeneity in our model since the regressors ($\mathbf{X_i}$) are correlated with the error term ($\mathbf{X_j}\varphi + \mu_{it}$), hence requiring the use of instruments to counteract the effects of omitted variable bias[116] (Wooldridge 2002; Cameron and Trivedi 2005). To militate against this bias we follow the 'naive approach' and do not include the crime statistics that are prone to omitted variable bias[117]. In our case this is robbery under *contact crime*. This error does not affect *contact related crimes*, as robbery only forms part of *contact crimes*.

[115] To drive this point home we illustrate with an example, the usual model in most crime literature, which is: $\text{Economic crime in area } i = \text{f (characteristics of area } i)$ where area represents country, state, province or community, depending on the unit of analysis. However, our example illustrates that the model is more likely to have the following form: $\text{economic crime in area } i = \text{f (characteristics of area } i, \text{ characteristics of area } j)$ when *mobile* crimes are involved. A practical example would be the case of individuals from disadvantaged communities who are more likely to cross over to more advantaged communities to commit economic crimes such as house robbery. Omitting the characteristics of community *j* introduces omitted variable bias and will result in the coefficients being biased. However, the difficulty in estimating the model is further compounded by the fact that the model is likely to include more than one area, that is the individual who commits a crime in area *i* may be from area *j*, *z* or *q* and so on, and therefore establishing the community's boundary becomes a challenge.

[116] If the omitted variables were time invariant then fixed effect model will suffice to cater for the omitted variable, however because $\mathbf{X_j}$ contains time varying omitted variables then IVs are also required (Wooldridge 2002). An alternative approach would be to use proxy variables for $\mathbf{X_j}$ (Wooldridge 2002). However because of the *open-endness* of $\mathbf{X_j}$, that is $\mathbf{X_j}$ represents characteristics of adjacent communities (community *j*) from the community where crimes are committed (community *i*), determining the boundary of these adjacent communities becomes a challenge.

[117] From the list of studies that are closely related to this study, Gorman *et al.* (2001) and Breetzke (2010) acknowledge the likely presence of mobility in some of the crime data. However, these studies do not take any measures to counteract the effects. Fajnzylber *et al.* (2002) use instrumental variables (IVs) in their model, stating that it is likely that all the variables are endogenous. However they do not specify the likely cause of the endogeneity.

The fourth and final likely estimation challenge is unobserved community heterogeneity which refers to the intercommunity differences that cannot be measured, for example visible policing or police structures. The study applies the fixed effects model to counteract these effects[118]. Lastly, we include the effects of community racial composition in our analysis. This is done by interacting the *majority population group*[119] dummy with our regressors. The motivation for this is to capture the effects of the differences in community based on the type of racial composition. This is because the apartheid system has led to differences in community resources along racial lines (Emmett 2003; Demombynes and Ozler 2005; Hipp 2007; Breetzke 2008; Breetzke 2010). Currently communities are still identified by their majority racial composition[120]. Equation (4.8) is our new equation with the interaction terms, where $\mathrm{p_i}$ represents the majority population group of community *i*, where ∂_t is the year dummy which controls for variations across time periods.

$$D_{it} = \alpha + \beta_1 . dis_{it}p_i + \beta_2 . abs_{it}p_i + \beta_3 . fdi_{it}p_i + \beta_4 . res_{it}p_i + \beta_5 . uem_{it}p_i + \partial_t + \mu_{it} \quad (4.8)$$

4.6 Estimation Results

4.6.1 Descriptive Statistics

Table 4.1 shows the descriptive statistics of our variables. The variables have been grouped by the *majority population group* in the community. As earlier indicated our data is a panel of three years. The sample consists of 77 communities which implies a total of 231 observations. Also as previously mentioned all variables are aggregated at the community level, they are then normalised by dividing them with the largest value in the sample.

There are two key observations that the descriptive statistics in Table 4.1 show. Firstly, contact and contact related crime appear to be higher among communities where the Black population group are in the majority, lower among the majority Coloured communities and least among the majority White communities. The second key observation is that property crime appears to be the highest among majority White communities, hence supporting our earlier argument of the mobile nature of crime: individuals from disadvantaged communities are likely to cross over to advantaged communities to commit these economic crimes as opposed to social crimes (contact and contact related) which mainly occur as a result of social behaviour.

118 It is worth noting that a fixed effects model controls for unobservables that are time invariant. Hence any unobservables that are time variant are likely to be absorbed by the error term, and if these time variant unobersvables are correlated with the regressors it will result in omitted variable bias and cause endogeneity (Cameron and Trivedi 2005). The time variant unobserved community heterogeneity that is likely to affect our model is that which will result from the *mobile* crimes such as robbery, and our strategy is to exclude these crimes.

119 Majority racial population group categories define communities by population group that is most common in the area, this can be either be Black population group, a Coloured population group or a White population group.

120 An alternative approach is to divide the sample by population and determine the effects of the individual variables by the various population groups. However, given the small sample size of our data, this approach is not ideal.

The descriptive statistics also show that community disadvantage is highest amongst the Black communities (0.801), followed by the Coloured communities (0.792) while the mean community disadvantage for the White communities is very low (0.267). The other remaining regressors also appear to be highest among the Black communities illustrating the weak social structures in these communities which is consistent with findings in the existing literature.

Table 4.1: Descriptive Statistics of our Sample

Variable	Obs	Mean	Std. Dev.	Min	Max	Source
Black Population Group						
Dependent Variable						
Contact crime	57	0.453	0.297	0.014	0.918	Crime data
Contact related crime	57	0.499	0.315	0.025	1.000	Crime data
Property crime	57	0.162	0.094	0.042	0.316	Crime data
Police detection crime	57	0.156	0.093	0.020	0.304	Crime data
Other crime	57	0.240	0.111	0.020	0.398	Crime data
Independent Variables						
Community disadvantage index	57	0.801	0.117	0.323	0.912	CAPS household survey
Residential instability	57	0.050	0.087	0.000	0.385	CAPS young adult survey
Family disruption	57	0.591	0.128	0.393	0.846	CAPS young adult survey
Service deliver index	57	0.466	0.220	0.022	0.967	CAPS household survey
Unemployment (head of household)	57	0.862	0.120	0.553	0.993	CAPS household survey
Coloured Population Group						
Dependent Variable						
Contact crime	129	0.104	0.104	0.031	0.516	Crime data
Contact related crime	129	0.225	0.145	0.082	0.755	Crime data
Property crime	129	0.189	0.117	0.050	0.545	Crime data
Police detection crime	129	0.131	0.097	0.038	0.465	Crime data
Other crime	129	0.173	0.158	0.020	0.916	Crime data
Independent Variables						
Community disadvantage index	129	0.792	0.090	0.552	1.000	CAPS household survey
Residential instability	129	0.080	0.137	0.000	0.671	CAPS young adult survey
Family disruption	129	0.503	0.137	0.352	1.000	CAPS young adult survey
Service deliver index	129	0.348	0.163	0.031	1.000	CAPS household survey
Unemployment (head of household)	129	0.839	0.133	0.355	0.999	CAPS household survey
White Population Group						
Dependent Variable						
Contact crime	45	0.074	0.052	0.010	0.224	Crime data
Contact related crime	45	0.251	0.165	0.082	0.890	Crime data
Property crime	45	0.328	0.198	0.078	1.000	Crime data
Police detection crime	45	0.255	0.220	0.046	1.000	Crime data
Other crime	45	0.155	0.202	0.017	1.000	Crime data
Independent Variables						
Community disadvantage index	45	0.267	0.118	0.050	0.512	CAPS household survey
Residential instability	45	0.207	0.261	0.000	1.000	CAPS young adult survey
Family disruption	45	0.440	0.116	0.352	0.909	CAPS young adult survey
Service deliver index	45	0.453	0.152	0.208	0.759	CAPS household survey
Unemployment (head of household)	45	0.780	0.131	0.526	1.000	CAPS household survey

We examine the amount of variation across communities and across time, see Cameron and Trivedi (2005). Several features are worth mentioning in this regard. Firstly, we observe that the within (across time) variations is less than half of the between (across communities) variation for contact crime, whereas in the contact related crime the within variation is half as large as the between variation. Turning to the regressors, we observe a similar pattern, in that the within variation is somewhat less than the between variation. As expected all the variability in the wave dummy is within variation. In general the between variation ranges between 0.11 and 1.64, while the within variation ranges from 0.06 to 0.86. Thus the variation is larger between communities than it is within communities. By incorporating fixed effects a large portion of total variation is therefore absorbed. Since the Hausman test, which discriminates between fixed effects and random effects model, rejected the random effects specification, the models presented in this paper are based on fixed effects models. Also worth reporting is the fact that an alternative analysis using the random effects models, although not reported in the paper, yielded reasonably similar results to our fixed effects models. Importantly, given that our panel consist of 77 communities covering a period of three years we do take note that the fixed effects estimates are likely to be imprecise, since most of the variation is across communities rather than across time. Notwithstanding, our results are reasonably consistent with the current literature on crime.

4.6.2 Fixed Effects Model Results

Our results are based on two types of regression models: the first model only controls for unobserved community heterogeneity by applying the fixed effects model. The results of these regressions are depicted in Table 4.2. The second model controls for unobserved community heterogeneity, measurement error and omitted variable bias. The unobserved heterogeneity is controlled by fixed effects and the measurement error bias as well as omitted variable bias is controlled for by using the 'naive approach and secondly by adjusting for the error approach, as depicted in Table 4.3.

Table 4.2 shows both the main and interaction effects of community social disorganisation and racial composition on contact and contact related crime. Specifically, the main results are depicted in panel *A*, where column 1 of panel *A* depicts the results based on *contact crime* and column 2 shows the results when using *contact related crime* as the outcome variable. Panel *B* (column 3 and 4) shows the interaction effects where we interact our regressors with the majority population group dummy. Worth noting is the fact that the models presented in this paper are based on robust standard errors with clustering at the community level. Also worth noting is the fact that we found the results to be similar when we used robust standard errors. See Cameron and Trivedi (2005) and Peterson (2009) for discussion on standard errors in panel data.

We first observe that the interaction models in panel *B* are more significant than the models without interaction in panel *A*, showing that the racial composition of the community is significant. We also notice that in general when the community has either a Coloured or White racial composition the effect of any of the regressors on crime is less (as shown by the interaction effects). For example the effect of community disadvantage in column 3 for the Black population group is 3.663, while for the Coloured population group this is 3.663 + (-3.397), and the White population group is 3.663 + (-3.552). *Community disadvantage* is statistically significant in most of the models but more so for contact crime (column 3). The same pattern is observed for *residential instability* in both the main effect model shown in panel A and the interaction model in panel B, where *residential instability* appears to be more significant among the Black communities than in the communities defined by a dominance of other population groups.

Family disruption which captures the frequency of interaction between the youth and both parents is significant in the main effects (panel *A*) and even more significant when we introduce interactions, as shown in panel B. We also observe the effects to be more significant in the Black communities, and surprisingly in White communities as well although the effects are not significant in the Coloured communities. The results are consistent with findings in the current literature that show malicious damage to property (this crime falls under *contact related crime*) such as graffiti and broken windows to be predicted by evidence of unsupervised youth (Sampson and Groves 1989; Kawachi *et al*. 1999). It is worth noting that our *family disruption* proxy, frequency of interaction between the youth and both parents is not available in year 2006 of the CAPS data. To derive the variable we impute using the frequency of interaction between the youth and both parents of 2002 and 2005 and individual characteristics (age, gender, population group and education) common in all the three years. Table 4.F1 of Appendix 4.F replicates the results of Table 4.2 with the only difference being that we exclude the year 2006. Hence the results consist of a two year panel of 77 communities, making the sample size 154. The results between the two tables are similar indicating that our imputation has not biased the results.

In addition the main effects show that *lack of basic services* is significant in determining both contact and contact related crime. The interaction effects show that the results are significant in the Black and Coloured communities, while in the White communities these effects are insignificant. In Appendix 4.G we substitute the *lack of basic services* variable with the access to water, electricity and sanitation measure. The results are shown in Tables 4.G1, 4.G2 and 4.G3 respectively.

Table 4.2: Base Results From Fixed Effects Model

	Panel A[1]		Panel B[2]	
Dependent Variable	(1) Contact Crime[3]	(2) Contact related crime[4]	(3) Contact Crime[3]	(4) Contact related crime
Community disadvantage	0.653***	0.316*	3.663***	2.526***
	(0.246)	(0.188)	(0.663)	(0.506)
Community disadvantage*Majority Coloured			-3.397***	-2.394***
			(0.775)	(0.592)
Community disadvantage*Majority White			-3.552***	-2.703***
			(0.737)	(0.563)
Residential instability	0.323*	0.280*	0.998**	0.802**
	(0.191)	(0.146)	(0.487)	(0.372)
Residential instability*Majority Coloured			-0.816	-0.818*
			(0.545)	(0.416)
Residential instability*Majority White			-0.903	-0.389
			(0.577)	(0.441)
Family disruption	0.0497	0.0494**	0.155**	0.120**
	(0.0306)	(0.0234)	(0.0628)	(0.0480)
Family disruption*Majority Coloured			-0.103	-0.0661
			(0.0756)	(0.0578)
Family disruption*Majority White			-0.172*	-0.129*
			(0.0944)	(0.0721)
Lack of basic services	0.0307***	0.0273***	0.0697***	0.0539***
	(0.0114)	(0.00873)	(0.0211)	(0.0161)
Lack of basic services *Majority Coloured			-0.0559**	-0.0371*
			(0.0279)	(0.0213)
Lack of basic services *Majority White			-0.0364	-0.0306
			(0.0294)	(0.0225)
Head unemployed	0.0461	-0.0253	0.375	0.169
	(0.169)	(0.129)	(0.278)	(0.212)
Head unemployed*Majority Coloured			-0.587	-0.383
			(0.362)	(0.277)
Head unemployed*Majority White			-0.338	-0.266
			(0.405)	(0.309)
Year 2005	-0.107***	-0.0781***	-0.0798***	-0.0614***
	(0.0247)	(0.0189)	(0.0251)	(0.0192)
Year 2006	-0.184***	-0.137***	-0.154***	-0.115***
	(0.0239)	(0.0183)	(0.0240)	(0.0183)
Constant	-32.31*	-27.77*	-36.96*	-26.97*
	(19.12)	(14.61)	(19.64)	(15.00)
Observations	231	231	231	231
R-squared	0.315	0.299	0.451	0.439
Number of communities	77	77	77	77

- Robust Standard errors in parentheses
- *** p<0.01, ** p<0.05, * p<0.1 indicates significance at 1%, 5% and 10%

[1] Panel A is based on main effects only

[2] Panel B is based on main effects and interaction effects of all the regressors. As evident by R square, the interaction models in panel B offer a better fir than those in panel A.

[3] As previously described in section 4.5.2, contact crimes include murder, rape, assault, kidnapping, abduction and robbery

[4] Also as defined in section 4.5.2 contact related crimes consist of arson, malicious damage to property, public violence and *crimen injuria.*

Table 4.3 presents the results after controlling for measurement error bias and omitted variable bias in our model. Since measurement error only affects *contact crime* and not *contact related crime*, the results are based on *contact crime* alone. Panel *D* is the *contact crime* base results from Table 4.2 in panel A. In Panel *E* we attempt to eliminate the bias using the 'naive approach' by excluding the crimes that are prone to measurement error. That is, we remove rape and robbery from *contact crime,* for the reason that the measurement error is likely to be systematically correlated with the regressors. This can occur if members from economically advantaged communities report these crimes while those from disadvantaged communities do not, as a result of mistrusting the police and therefore having a preference for vigilante justice. Also recall that we control for omitted variable bias using the 'naive approach' as the crime (robbery) prone to this bias is also eliminated.

In Panel *F* we adjust for measurement error bias by calculating the probability of reporting crime from the Victims of Crime Survey data using these probabilities to correct the crime statistics. The results, shown in panel *D*, *E* and *F* are similar with the only difference being the size of the coefficients. This finding is similar to those of the study by Demombynes and Ozler (2005) who, using the Victims of Crime Survey data, after controlling for measurement error found the results to be similar to the results obtained when the error was not controlled.

Table 4.3: Results Based on Fixed Effects Model Controlling for Measurement Error and Omitted Variable Bias

	Panel D[1]		Panel E[2]		Panel F[3]	
	Base results: column (1) and (3) in Table 4.2		Measurement error soln: 'Naïve approach'		Measurement error soln: Adjusting the crime data	
Dependent Variable	(1) Contact Crime	(2) Contact Crime	(3) Contact Crime	(4) Contact Crime	(5) Contact Crime	(6) Contact Crime
Community disadvantage	0.653***	3.663***	0.451***	2.485***	0.419***	2.341***
	(0.246)	(0.663)	(0.165)	(0.445)	(0.160)	(0.433)
Community disadvantage*Majority Coloured		-3.397***		-2.290***		-2.181***
		(0.775)		(0.520)		(0.507)
Community disadvantage*Majority White		-3.552***		-2.403***		-2.262***
		(0.737)		(0.495)		(0.482)
Residential instability	0.323*	0.998**	0.225*	0.717**	0.206*	0.631**
	(0.191)	(0.487)	(0.129)	(0.327)	(0.124)	(0.318)
Residential instability*Majority Coloured		-0.816		-0.608*		-0.510
		(0.545)		(0.365)		(0.356)
Residential instability*Majority White		-0.903		-0.624		-0.578
		(0.577)		(0.387)		(0.377)
Family disruption	0.0497	0.155**	0.0367*	0.108**	0.0308	0.0980**
	(0.0306)	(0.0628)	(0.0206)	(0.0421)	(0.0199)	(0.0410)
Family disruption*Majority Coloured		-0.103		-0.0698		-0.0645
		(0.0756)		(0.0507)		(0.0494)
Family disruption*Majority White		-0.172*		-0.113*		-0.109*
		(0.0944)		(0.0633)		(0.0617)
Lack of basic services	0.0307***	0.0697***	0.0220***	0.0480***	0.0193**	0.0433***
	(0.0114)	(0.0211)	(0.00769)	(0.0141)	(0.00743)	(0.0138)
Lack of basic services *Majority Coloured		-0.0559**		-0.0371**		-0.0345*
		(0.0279)		(0.0187)		(0.0182)
Lack of basic services *Majority White		-0.0364		-0.0255		-0.0222
		(0.0294)		(0.0197)		(0.0192)
Head unemployed	0.0461	0.375	0.0178	0.249	0.0433	0.247
	(0.169)	(0.278)	(0.114)	(0.186)	(0.110)	(0.181)
Head unemployed*Majority Coloured		-0.587		-0.408*		-0.374
		(0.362)		(0.243)		(0.237)
Head unemployed*Majority White		-0.338		-0.240		-0.196
		(0.405)		(0.272)		(0.265)
Year 2005	-0.107***	-0.0798***	-0.0660***	-0.0477***	-0.0693***	-0.0522***
	(0.0247)	(0.0251)	(0.0166)	(0.0169)	(0.0161)	(0.0164)
Year 2006	-0.184***	-0.154***	-0.119***	-0.0984***	-0.121***	-0.102***
	(0.0239)	(0.0240)	(0.0161)	(0.0161)	(0.0155)	(0.0157)
Constant	-32.31*	-36.96*	-22.49*	-25.80*	-20.61*	-23.54*
	(19.12)	(19.64)	(12.86)	(13.18)	(12.43)	(12.84)
Observations	231	231	231	231	231	231
R-squared	0.315	0.451	0.302	0.444	0.317	0.447
Number of communities	77	77	77	77	77	77

- Robust Standard errors in parentheses
- *** p<0.01, ** p<0.05, * p<0.1 indicates significance at 1%, 5% and 10%

[1] Panel D depicts our base results in table 4.2. The purpose for including these results in table 4.3 is to aid a comparison of results between our base results and the results after controlling for measurement error.

[2] In Panel E we control for measurement error using the 'naïve approach' in which we omit the crimes that are prone to measurement error. Hence we have omitted rape and robbery. The contact crime dependant variable therefore consists of murder, assault, kidnapping, and abduction. See section 4.6.3.

It is also worth noting that the naïve approach also controls for omitted variable bias, as the crimes (robbery) prone to omitted variables are no longer included.

In Panel F we control for measurement error by multiplying the crime statistics with 1/p, where p is the probability of reporting crime, following Demombynes *et al.* (2005). See section 4.6.3

4.7 Conclusion

This study examined how community social disorganisation influence the perpetration of *contact crime* and *contact related crime*. Community social disorganisation was captured using *family disruption*, *residential mobility*, *community disadvantage*, *lack of basic services* and *unemployment* following conventional research detailed in the literature. The motivation behind this study is the high frequency of contact crime in South Africa (Emmett 2003; Blackmore 2003; Demombynes and Ozler 2005; Breetzke 2010; Whyte 2010), and the research on current literature identifying South African communities as being socially disorganised (Emmett 2003; Demombynes and Ozler 2005; Breetzke 2010; Budlender and Lund 2011; Schatz *et al.* 2011).

To determine the impact of these social factors on crime we follow Fajnzylber *et al.* (2002) and Haddad and Moghadam (2011) who extend Becker's theory by including social and demographic structures in the theoretical framework. To arrive at the results we first need to control for excessive collinearity (Hannon and Defronzo 1998), measurement error as a result of reporting bias of police crime data (Fajnzylber *et al.* 2002; Demombynes and Ozler 2005; Louw 2007), omitted variable bias and the influence of community unobservables (Fajnzylber *et al.* 2002; Blackmore 2003).

Firstly, the study adopts a fixed effects model that controls for unobservable community heterogeneity using panel data covering the Cape Town Metropolitan region of South Africa. To eliminate the measurement error bias of police crime statistics, mainly as a result of misreporting, we use two approaches. The first is the 'naive approach' where we eliminate those crimes identified as being prone to reporting bias, rape and robbery, from our *contact crime group.* In the second approach we use the 1998 Victims of Crime Survey data to obtain the probability of reporting crime and we use these probabilities to adjust our crime statistics for measurement error. In addition the 'naive approach' has the effect of curbing the omitted variable bias as we have eliminated robbery (mobile crime that is likely to have this bias) from our *contact crime* outcome.

Additionally, all regressors are interacted with *majority racial population group* dummies to capture the effects of the differences in community based on the type of racial composition. This is an influence of the apartheid era on community structures, where individuals were housed by race with the Black and Coloured population groups having the least resources allocated to their communities. In general, our results show that the Black communities are systematically worse off than Coloured and White communities.

Our findings reveal that *lack of basic services* breeds crime, as we find *lack of basic services* to be positive and significant in determining *contact crime* and *contact related crime* especially among the Black communities. Additionally, *residential instability* is also positive and significant in determining both crime types. Other social structures, such as *community disadvantage* and *family disruption* also appear to be strong determinants of crime in general, but once again more so among the Black communities.

As previously mentioned the current literature states that there is a paucity of empirical literature to substantiate the effects of socially disorganised communities on crime in developing countries in general, and in sub-Saharan Africa in particular (Ward 2007; Breetzke 2010; SAIRR 2010; Daniels and Adams 2010). This study has demonstrated that community social disorganisation in the Cape Town Metropolitan area of South Africa creates a setting conducive to criminal behaviour and this research thereby supports the findings in the current literature that suggest that the aforementioned factors may have a role to play in the current high crime rates in South Africa (Emmett 2003; Louw 2007; Landau and Misago 2009; Whyte 2010; Lau *et al.* 2010; SAIRR 2010).

In addition our study can be seen as complementary to the cross sectional study by Breetzke (2010) which was conducted in the Tswane area of South Africa. By using panel data, alternative proxies and the majority racial group interaction effects we are able to achieve positive and significant coefficients, which point to social disorganisation leasing to the perpetration of both *contact crimes* and *contact related crimes*. This is found across communities represented by all racial groups but especially among the majority Black communities.

From a policy perspective, our results indicate that an investment in social capital is essential at the community level, while improving family structures is of outmost importance in curbing crimes of social behaviour.

4.8 Reference

Agnew, R. (1985). A revised strain theory of delinquency, *Social Forum*, 64: 672-682.

Agnew, R. (1999). A general strain theory of community differences in crime rates, *Journal of Research in Crime and Delinquency*, 36 (2): 123-155.

Akers, A., Muhammad, M. and Corbie-Smith, G. (2011). "When you got nothing to do, you do somebody": A community's perceptions of neighbourhood effects on adolescent sexual behaviours, *Social Science and Medicine*, 72: 91-99.

Alexander, P. (2010). Rebellion of the poor: South Africa's service delivery protests - A preliminary analysis, *Review of African Political Economy*, 37 (123): 25-40.

Anderson, K. and Beutel, A., (2007). HIV/AIDS prevention knowledge among youth in Cape Town, South Africa, *Journal of Social Sciences*, 3 (3): 143-151.

Atkinson, M., Greenstein, T., and Lang, M. (2005). For women, breadwinning can be dangerous: Gendered resource theory and wife abuse, *Journal of Marriage and Family*, 67: 1137-1148.

Baier, C. and Wrigt, B. (2001). "If you love me, keep my commandments": A meta-analysis of the effect of religion on crime, *Journal of Research in Crime and Delinquency*, 38 (3): 3-21.

Baron, S. (2003). Self-control, social consequences and criminal behaviour: street youth and the general theory of crime, *Journal of Research in Crime and Delinquency*, 40 (4): 403-425.

Becker, G. (1968). Crime and punishment: an economic approach, *The Journal of Political Economy*, 76 (2): 169-217.

Becker, R. and Mehlkop G. (2006). Social class and delinquency: An empirical utilisation of rational choice theory with cross-sectional data of the 1990 and 2000 German general population surveys (ALLBUS), *Rationality and Society*, 18 (2): 193-235.

Becker, B., Gaunt, C., Eberhard, A. and Marquard, A. (2008). Uncertainties with South Africa's goal of universal access to electricity by 2012, *Journal of Energy in Southern Africa*, 19 (2): 4-13.

Benoit, K., Laver, M. and Mikhaylov, S. (2009). Testing words as data with error: Uncertainty in text statements of policy positions, *American Journal of Political Science*, 53 (2), 495-513.

Berry, J. and Levitt, S. (1999). Crime, urban flight and the consequences for cities, *The Review of Economics and Statistics*, 81 (2): 159-169.

Blackstock, C. and Trocme, N. (2005). Community-based child welfare for aboriginal children: Supporting resilience through structural change, *Social Policy Journal of New Zealand*, 24: 12-24.

Blackmore, F. (2003). A panel data analysis of crime in South Africa, *Sajems*, 6 (3): 439-458.

Breetzke, G. and Horn, A. (2008). Key requirements in the development of spatial-ecological theory of crime in South Africa, *Acta Criminologica*, 21 (1): 123-143.

Breetzke, G. (2010). A Socio-structural analysis in the city of Tshwane, South Africa, *South African Journal of Science*, 106 (11/12): 1-7.

Bricker, J. and Engelhardt, G. (2008). Measurement error in earnings data in the health and retirement study, *Journal of Economics and Social Measurement*, 33: 39-61.

Broidy, L. and Agnew, R. (1997). Gender and crime: A general strain theory perspective, *Journal of Research in Crime and Delinquency*, 34 (4): 275-306.

Brown, K. (2001). The determinants of crime in South Africa, *South African Journal of Economics*, 69 (2): 269-299.

Budlender, D. and Lund, F. (2011). South Africa: A legacy of family disruption, *Development and Change*, 42 (4): 925-946.

Buonanno P. and Leonida L. (2006). Education and crime: evidence from Italian regions, Applied Economics Letter, 13 (11): 709-713.

Cameroon, C. and Trivedi, P. (2005). *Microeconomics Methods and Applications*, Cambridge, UK: Cambridge University Press.

Caughy, M. and O'Campo, P. (2006). Neighbourhood poverty, social capital, and the cognitive development of African American preschoolers, *American Journal of Community Psychology,* 37 (1): 141-154.

Clay-Warner, J. and McMahon-Howard, J. (2009). Rape reporting: 'Classic rape' and the behaviour of law, *Violence and Victims*, 24 (6): 723-743.

Cohen, D., Spear, S., Scribner, R., Kissinger, P., Mason, K. and Wildgen, J. (2000). "Broken windows" and the risk of gonorrhoea, *American Journal of Public Health*, 90: 230-236.

Cornwell, C. and Trumbull, W. (1994). Estimating the economic model of crime with panel data, *The Review of Economics and Statistics*, 76 (2): 360-366.

Daniels, D. and Adams, Q. (2010). Breaking with township gangsterism: The struggle for place and voice, *African Studies Quarterly*, 11 (4): 45-57.

Davies, R. and Thurlow, J. (2010). Formal-informal economy linkages and unemployment in South Africa, *South African Journal of Economics*, 78 (4): 437-459.

Dembo, R., Belenko S., Childs K., Wareham J. and Schmeidler J. (2009). Individual and community risk factors and sexually transmitted diseases among arrested youths: A two level analysis, *Journal of Behavioural Medicine*, 32: 303-316.

Demombynes, G. and Ozler, B. (2005). Crime and local inequality in South Africa, *Journal of Development Economics*, 76: 265-292.

DiIulio, J. (1996). Help wanted: Economists, crime and public policy, Journal of Economic Perspectives, 10: 3-24.

Ehrlich, I. (1973). Participation in illegitimate activities: A theoretical and empirical investigation. *The Journal of Political Economy*, 81, 521-565.

Ehrlich, I. (1996). Crime, punishment, and the market for offenses. *The Journal of Economic perspectives*, 10: 43-67.

Emmett, T. (2003). Social disorganisation, social capital and violence prevention in South Africa, *African Safety Promotion*, 1: 4-18.

Fajnzylber, P., Lederman, D. and Loayza, N. (2002). What causes violent crime? *European Economic Review*, 46: 1323-1357.

Fougere, D., Pouget, J. and Kramarz, F. (2009). Youth unemployment and crime in France, *Journal of the European Economic Association*, 7 (5): 909-938.

Gabrysch, S., Edwards, T. and Glynn, J. (2008). The role of context: Neighbourhood characteristics strongly influence HIV risk in young women in Ndola, Zambia, *Tropical Medicine and International Health*, 13 (2): 162-170.

Gaviria, A. (2000). Increasing returns and the evolution of violent crime: The case of Colombia, *Journal of Development Economics*, 61: 1-25.

Gillespie, G., Greener, R., Whitworth, J. and Whiteside, A. (2007). Poverty, HIV and AIDS: Vulnerability and impact in Southern Africa, *AIDS*, 21 (7): S1-S104.

Glaser, L., Sacerdote, B. and Scheinkman, A. (1996). Crime and social interactions, *The Quarterly Journal of Economics*, 111 (2): 507-548.

Goldin, J. (2010). Water policy in South Africa: Trust and knowledge as obstacles to reform, *Review of Radical Political Economics*, 42 (2): 195-212.

Gould, E., Weinberg, B. and Mustard, D. (2002). Crime rates and local labor market opportunities in the United States: 1979-1997, *The Review of Economics and Statistics*, 84(1): 45-61.

Gorman, D., Speer, P., Groenewol,d P. and Labourne, F. (2001). Spatial dynamics of alcohol availability, neighbourhood structure and violent crime, *Journal of Studies on Alcohol*, 62: 628-636.

Greene, W., (2002). *Econometric Analysis*, 5th Edition. Upper Saddle River, New Jersey: Prentice-Hall.

Groot W. and Brink H. (2010). The effects of education on crime, *Applied Economics*, 42 (3): 279-289.

Haddad, G, and Moghadam, H. (2011). The socioeconomic and demographic determinants of crime in Iran: A Regional Panel Study, *European Journal of Law Economics*, 32: 99-114.

Hannon, L. and Defronzo, J. (1998). The truly disadvantaged, public assistance, and crime, *Social Problems*, 45 (3): 383-392.

Heineke, J. (1978). Substitution among crimes and the question of deterrence: An indirect utility function approach to the supply of legal and illegal activity, Economic models of criminal behaviour, North-Holland Publishing Company.

Hipp, J., (2007). Income inequality, race and place: Does the distribution of race and class within neighbourhood affect crime rates? *Crime*, 45: 665-697.

Hobfoll, S. (2001). The influence of culture, community, and the nested-self in the stress process: Advancing conservation of resources theory, *Applied Psychology: An International Review*, 50 (3): 337-421.

Horvath, R. and Kolomaznikova, E. (2002). Individual decision making to commit a crime: Early models, *Law and Economics*, No. 0210001: 1-17.

Hough, M. (2009). Violent protest at local government level in South Africa: Revolutionary potential, *South African Journal of Military Studies*, 36 (1): 1-14.

Jarrette, R. (1997). African America family and parenting strategies in impoverished neighbourhood, *Qualitative Sociology*, 20 (2): 275-288.

John, V. (2010). Community development in a post-conflict context: Fracture and depleted social capital, *Community Development Journal*, 46 (S1): i51-i65.

Kawachi, I., Kennedy, B. and Wilkinson, R. (1999). Crime: Social disorganisation and relative deprivation, *Social Science Medicine*, 48: 719-731.

Kelly, M., (2000). Inequality and crime, *Review of Economics and Statistics* 82: 530-9.

Kenyon C., and Zondo S., (2011). Why do some south African ethnic groups have very high HIV rates and others not?, *African Journal of AIDS Research*, 10 (1): 51-62.

Klasen, S. and Woolard, I. (2008). Surviving unemployment without state support: Unemployment and household formation in South Arica, *Journal of African Economies*, 18 (1): 1-51.

Kling, J., Ludwig, J. and Katz, L. (2005). Neighbourhood effects on crime for female and male youth: Evidence from a randomised housing voucher experiment, *Quarterly Journal of Economics*, 120 (1): 87-130.

Krug, E., Mercy, J., Dahlberg, L. and Zwi, A. (2002). The world report on violence and health, *Lancet*, 360: 1083-1088.

Landau, L. and Misago, J. (2009). Who to blame and what's to gain? Reflections on space, state and violence in Kenya and South Africa, *Africa Spectrum*, 44 (1): 99-110.

Lau, U., Seedat, M. and Shahnaaz, S. (2010). Research and theory, collective violence in South Africa: Explaining the explanations, *African Safety Promotion Journal*, 8 (2): 1-19.

Lauritsen, J. and Heimer, K. (2010). Violent victimisation among males and economic conditions: The vulnerability of race and ethnic minorities, *Criminology and Public Policy*, 9 (4): 665-692.

Lin, M. (2008). Does unemployment increase crime? Evidence from US data 1974-2000, *Journal of Human Resources*, XLII (2): 413- 454.

Lochner L. (2004). Education, work and crime: a human capital approach, International Economic Review, 45 (3): 811-843.

Louw, A. (2007). Crime and perceptions after a decade of democracy, *Social Indicators Research*, 81: 235-255.

Ludwig, J., Duncan, G. and Hirschfield, P. (2001). Urban poverty and juvenile crime: Evidence from a randomised housing-mobility experiment, *The Quarterly Journal of Economics*, 116 (2): 655-679.

Machin S., Marie O. and Vujic S. (2011). The crime reducing effects of education, The Economic Journal, 121: 463-484.

Mah, T. (2010). Prevalence and correlates of concurrent sexual partnerships among young people in South Africa, *Sexually Transmitted Diseases,* 37 (2): 105-108.

Mah, T. and Halperin, D. (2010b). Concurrent sexual partnerships and the HIV epidemics in Africa: Evidence to move forward, *AIDS Behaviour*, 14: 11-16.

Mah, T. and Halperin, D. (2010c). The evidence for the role of concurrent partnerships in Africa's HIV epidemics: A response to Lurie and Rosenthal, *AIDS Behaviour*, 14: 25-28.

Manski, C. (1993). Identification of endogenous social effects: The reflection problem, *The Review of Economic Studies*, 60 (3): 531-542.

Manski, C. (2000). Economic analysis of social interactions, *The Journal of Economic Perspectives*, 14 (3): 115-136.

Manski, C., (2003). Identification problems in the social sciences and everyday life, *Southern Economic Journal*, 70 (1): 11-21.

Mare, G., (2001). Race counts in contemporary South Africa: 'An illusion of ordinaries', *Transformation*, 47: 75-93.

Mathews, S., Jewkes, R.. and Abrahams, N. (2011). Exploring childhood adversity in the shaping of masculinities among men who killed an intimate partner in South Africa, *British Journal of Criminology*, 51: 960-977.

Mehlkop, G. and Graeff, P. (2010). Modelling a rational choice theory of criminal action: Subjective expected utilities, norms and interactions, *Rationality and Society*, 22 (2): 189-222.

Miyoshi, K. (2011). Crime and local labor market opportunities for low skilled workers: Evidence using Japanese prefectural panel data, *Pacific Economic Review*, 16 (5): 565-576.

Myers, S. (1983). Estimating the economic model of crime: Employment versus punishment effects, *The Quarterly Journal of Economics*, 98 (1): 157-166.

Naude, W. (2010). Does sexual strategies theory help explain high violent crime in South Africa? *Applied Economics Letters*, 17: 571-574.

Nisim, D., Halperin, E., Hobfoll, S. and Johnson, R.. (2006). Xenophobia towards Palestine citizens of Israel among Russian immigrants in Israel: Heightened by failure to make gains in a new democratic society, Working Paper 327.

Nleya, N. and Thompson, L. (2009). Survey methodology in violence-prone Khayelitsha, Cape Town, South Africa, *IDS Bulletin*, 40 (3): 50-57.

Osgood, D. (2000). Poisson-based regression analysis of aggregate crime rates, *Journal of Quantitative Criminology*, 16(1): 21-43.

Oster, A. and Agell, J. (2007). Crime and unemployment in turbulent times, *Journal of the European Association*, 5 (4): 752-775.

Patalinghug, E. (2009). Identifying the Link between Unemployment and Crime in the Philippines, *International Journal of Comparative and Applied Criminal Justice*, 33 (1): 84-95.

Petersen M. (2009). Estimating standard errors in finance panel data sets: Comparing approaches, *The Review of Financial Studies*, 22 (1): 435-480.

Ragnarsson, A., Townsend, L., Ekström, A., Chopra, M. and Thorson, A. (2010). The construction of an idealised urban masculinity among men with concurrent sexual partners in a South African township, *Global Health Action*, 3 (5092): 1-7.

Rapheal, S. and Winter-Ebnmer, R. (2001). Identifying the effects of unemployment on crime, *Journal of Law and Economics*, 44(1): 259-283.

Roux, N. and Nyamukachi, P. (2005). A reform model for the improvement of municipal service delivery in South Africa, *Journal of Public Administration*, 40 (4.1): 687-705.

Ruiters, G., (2009). Free basic electricity in South Africa: A strategy for helping or containing the poor in human sciences research council, Electricity capitalism-recolonising Africa on the power grid, Edited by McDonald D. HSRC Press. South Africa.

Sampson, R. and Groves, W. (1989). Community structure and crime: Testing social disorganisation theory, *The American Journal of Sociology*, 94 (4): 774-802.

Sampson, R., Morenoff, J. and Earls, F. (1997). Neighbourhoods and violent crime: A multilevel study of collective efficacy, *Science,* 277: 918-924.

Sampson, R., Morenoff, J. and Earls, F. (1999). Beyond social capital: Spatial dynamics of collective efficacy for children, *American Sociological Review*, 64 (5): 633-660.

Schatz, E., Madhavan, S. and Williams, J. (2011). Female-headed households contending with AIDS-relate hardship in rural South Africa, *Health and Place*, 17: 598-605.

Seekings, J., Jooste, T., Langer, M. and Maughan-Brown, B. (2005). Inequality and diversity in Cape Town: An introduction and user's guide to the 2005 Cape Area Study**,** CSSR Working Paper No. 124.

Silber, G. and Geffen, N. (2009). Race, class and violent crime in South Africa: Dispelling the 'Huntley thesis', *South Africa Crime Quarterly*, 30: 35-43.

Statistics South Africa, (2001). Victims of Crime Survey 1998, Pretoria, South African Data Archive [distributor], South Africa.

South African Institute of Race Relation (SAIRR), (2011). First steps to healing the South African Family, SAIRR and Donaldson Trust.

South Africa Police Service (SAPS), (2010). The crime situation in South Africa, South Africa Police Service 2009/2010.

Super, G. (2010). The Spectacles of crime in the new South Africa: A historical perspective (1976-2004), *British Journal of Criminology*, 50: 165-184.

Tregenna, F. (2011). Earnings inequality and unemployment in South Africa, *International Review of Applied Economics*, 25 (5): 585-598.

UNAIDS, (2009). AIDS epidemic update, Joint UNAIDS and WHO, Geneva, Switzerland.

UNAIDS (2010). Joint action for results: UNAIDS outcome framework, 2009-2011, Joint United Nations Programme on HIV/AIDS Geneva, Switzerland.

Usher, D. (1997). Education as a deterrent to crime, *The Canadian Journal of Economics*, 30: 367-384.

Veysey, B. and Messer, S. (1999). Further testing of social disorganisation theory: An elaboration of Sampson and Grove's "Community structure and crime", *Journal of Research in Crime and Delinquency*, 36 (156): 156-174.

Vincent, L. (2008). The limitations of 'inter-racial contact': Stories from young South Africa, *Ethnic and Racial Studies*, 31 (8): 1426-1451.

Vromans, L., Schweitzer, R., Knoetze, K. and Kagee, A. (2011). The experience of xenophobia in South Africa, *American Journal of Orthopsychiatry*, 8 1(1): 90-93.

Ward, C. (2007). Neighbourhood indicators: Monitoring child rights and well-being at small area level in monitoring child wellbeing concepts and context. Assessed 17 January 2011. http://www.hsrcpress.ac.za.

Ward, C. and Bakhuis, K. (2010). Intervening in children's involvement in gangs: Views of Cape Town's young people, *Children and Society*, 24: 50-62.

Webber, C. (2007). Revaluating relative deprivation theory, *Theories of Crime*, 11: 97-120.

Whyte, E. (2010). *Aluta Continua*: The struggle continues in South Africa - Against violent crime, *Dialogue*, 7 (1): 1-30.

WHO, (2002). World report on violence and health: Summary, World Health Organisation (WHO), Geneva, Switzerland.

WHO, (2008). Manual for estimating the economic costs of injuries due to interpersonal and self directed violence, WHO and Department of Health and Human Services Centres for Disease Control and Prevention, Geneva, Switzerland.

Witte, A. (1980). Estimating the economic model of crime with individual data, *The Quarterly Journal of Economics*, 94(1): 57-84.

Wolitzky-Tailor, K., Resnick, H., McCauley, J., Amstadter, A., Kilpatrick, D. and Ruggiero, K. (2011). Is reporting of rape on the rise? A comparison of women with reported versus unreported rape experiences in the national women's study, *Journal of Interpersonal Violence*, 26 (4): 807-832.

Wong, Y. (1995). An economic analysis of the crime rate in England and Wales, 1857-92, *Economica*, 62 (246): 235-246.

Wooldridge, J. (2002). Econometric analysis of cross section and panel data, Cambridge, the MIT Press.

Wooldridge, J. (2003). *Introductory Econometrics: A Modern Approach*, 2nd Edition, Thomson South-Western.

Zamani, G., Gorgievski-Duijvesteijn, M. and Zarafshani, K. (2006). Coping with drought: Towards a multilevel understanding based on conservation of resources theory, *Human Ecology*, 34 (5): 677-692.

Appendix 4.A: Distribution of Crime Rates

Table 4.A1: Metropolitan Cape Town Crime Indicators

Crime Statistics-Mean	2001/ 2002	2002/ 2003	2003/ 2004	2004/ 2005	2005/ 2006
Contact Crimes					
· Murder	69.8	88	61.4	46.2	47.8
· Rape	107	99.2	94.9	82.2	75.8
· Attempted murder	67.5	95.8	77.6	45	31.4
· Assault with the intent to inflict grievous bodily harm	478.8	479.9	518.3	359.4	278.4
· Common assault	428	502.1	561.8	504.1	419.3
· Indecent assault	19.9	21.5	24.8	28.5	28.3
· Neglect and ill-treatment of children	7.5	16.5	22.2	14.5	9.4
· Culpable homicide	16.7	20.6	16.7	15.3	13.6
· Robbery with aggravating circumstances	328.7	388.8	352.3	306.1	302.5
· Common robbery	292.1	345.2	287.6	248.7	163.7
· Bank robbery (subcategory of aggravated robbery)	1	0	0	0.1	0.1
· Robbery of cash in transit (subcategory of aggravated robbery)	0.2	0.6	0.3	0.2	0.9
· Robbery at residential premises (subcategory of aggravated robbery)	0	9.1	6.4	8.7	12.5
· Robbery at business premises (subcategory of aggravated robbery)	0	3.5	1.8	1.8	2.8
Contact Related Crimes					
· Arson	11.9	13.9	17.1	10	7.1
· Malicious damage to property	341.3	408.8	441.2	380.1	365.5
· Crimen injuria	101.4	129	135	116.5	77.1
· Public violence	2.1	5.6	4	2.7	4.1
Property Related Crime					
· Burglary at residential premises	685.2	731.5	676.8	554.7	492.9
· Burglary at business premises	179.8	153.6	138.4	99.6	99.3
· Theft of motor vehicle and motorcycle	283.4	328.7	317	268.2	267.4
· Theft out of or from motor vehicle	905.8	869	726.7	637.8	654.4
· Stock-theft	1.5	0.9	1.3	0.6	0.4
Crime Detected as a Result of Police Action					
· Illegal possession of firearms and ammunition	44.6	44.3	42.2	36.9	29.9
· Drug-related crime	152.6	140.1	185.6	239.2	275.3
· Driving under the influence of alcohol or drugs	48.5	49.2	59.7	70	81
Other Serious Crime					
· All theft not mentioned elsewhere	1459	1732	1897	1715	1415
· Commercial crime	168.3	154.1	160.4	149	158
· Shoplifting	172.3	185.3	201.3	186.1	188.3
· Carjacking (subcategory of aggravated robbery)	22.7	29.7	32.6	22.5	23.6
· Truck hijacking (subcategory of aggravated robbery)	5.3	3.1	0.9	0.3	0.2
· Kidnapping	14.4	11	10.3	5	4.3
· Abduction	8.9	12.1	8.9	9.5	7.2

Appendix 4.B: Family Disruption Statistics in South Africa

Table 4.B1: Evidence of Family Disruption in South Africa

Proportion of children with absent fathers:	
• Black	Up from 46% (1996) to 52% (2009)
• Coloured	Up from 34% (1996) to 41% (2009)
• Indian	Down 17% (1996%) to 12% (2009)
• White	Up from 13% (1996%) to 15% (2009)
Proportion of children with absent, living fathers	Up from 42% (1996) to 48% (2009)
Proportion of female urban single parents by race group:	
• Black	79%
• Coloured	84%
• Indian	64%
• White	69%
Proportion of children (0-17) in HH with employed adult	34%
Urban single parents by age:	
• 16-24 years	13%
• 25-34 years	33%
• 35-44 years	24%
• 45-64 years	23%
Number of registered civil marriages	Down176 521 (2004) to 171 989 (2009)
Number of registered customary marriages	Down 20 301 (2004) to 13 506 (2009)

Source: South Africa Institute of Race Relations (SAIRR), 2011

Appendix 4.C: Correlation Matrix of the Regressors

Table 4.C1: Correlation Matrix of the Community Disadvantaged Proxies Vs. Average Income

	Ratio of no. of rooms to household members	Mean years of education	Community disadvantage index	Average community income
Ratio of no. of rooms to household members	1			
Mean years of education	0.5544*	1		
Community disadvantage index	0.9368*	0.7303*	1	
Average community income	0.8440*	0.4653*	**0.8302***	1

*** p<0.01, ** p<0.05, * p<0.1

Table 4.C2: Correlation Matrix of the Absolute Deprivation Proxies

	percentage of household with no access to sanitation	percentage of household with no access to electricity	percentage of household with no access to water
percentage of household with no access to sanitation	1		
percentage of household with no access to electricity	0.6067*	1	
percentage of household with no access to water	0.9697*	0.6397*	1

*** p<0.01, ** p<0.05, * p<0.1

Table 4.C3: Correlation Matrix of the Independent Variables

	Community disadvantage	Residential instability	Family disruption	Lack of basic services	Head unemployed
Community disadvantage	1				
Residential instability	0.3170*	1			
Family disruption	0.1016	-0.0139	1		
Lack of basic services	0.2917*	0.0925	-0.0715	1	
Head unemployed	0.2123*	0.0777	0.0290	0.1458*	1

*** p<0.01, ** p<0.05, * p<0.1

It is worth noting that the correlation between *community disadvantage* and *lack of basic services* is low and negative when *type of toilet* variable is included in the *lack of basic services* index. Once the *type of toilet* variable is removed from the *lack of basic service* index (leaving behind electricity and water) the correlation between this index and *community disadvantage* becomes positive and significant. A likely explanation for this is that the *type of toilet* variable does not capture the income effects across communities as this variable does not capture the complete and detailed variation in the *type of toilet* facilities present in communities. That is, in the CAPS dataset the following options exists in the variable: 'flush toilet', 'improved pit latrine with ventilation', 'other pit latrine', 'bucket toilet', 'chemical toilet' and 'none'. While other datasets such as the National Income and Dynamic Survey (NIDS), for example, offers a wider range of *type of toilet*, specifically, 'flush toilet with onsite disposal', 'flush toilet with offsite disposal', 'chemical toilet', 'pit latrine with ventilation pipe', 'pit latrine without ventilation pipe', 'bucket toilet', 'none' and 'other'. In our CAPS dataset all households with flush toilet, for instance, are therefore clustered as one category without distinguishing the type of 'flush toilet' in detail – whether it is 'flush toilet' with onsite or offsite disposal. The lack of this detailed explanation in the *type of toilet* is likely to make this variable bias in relation to income effects. Furthermore, a simple correlation between the *household income* and *type of toilet* variables reveals a negative correlation.
In addition and most importantly the CAPS dataset does not capture information on whether the household shares the toilet facilities or not, (this information is captured in NIDS for instance) which is also likely to influence the toilet condition across communities. For example the 'pit latrine with ventilation pipes' toilet facilities in one community will not be distinguished from another community who may have the same (or similar – 'pit latrine without ventilation pipe') toilet conditions but contains households who share these facilities.

Appendix 4.D: Types of Crime and Likely Estimation Challenge

Table 4.D1: Types of Crimes and Likely Estimation Challenge

Type of Crime Forming the Dependent Variable	Measurement error[1]	Omitted variable bias[2]	Excessive Collinearity	Unobservables
	Errors emanating from type of outcome variables		Errors emanating from regressors/model[3]	
Contact Crime				
Rape	√		√	√
Murder or Culpable homicide or Attempted murder			√	√
Assault			√	√
Kidnapping or Abduction			√	√
Neglect and ill-treatment of children			√	√
Robbery	√	√	√	√
Contact Related Crime				
Arson			√	√
Malicious damage to property			√	√
Crimen injuria			√	√
Public violence			√	√
Property Crime				
Burglary	√	√	√	√
Theft of motor vehicle and motorcycle		√	√	√
Theft out of or from motor vehicle	√	√	√	√
Stock-theft		√	√	√
Policing Crime				
Illegal possession of firearms and ammunition		√	√	√
Drug-related crime		√	√	√
Driving under the influence of alcohol or drugs		√	√	√
Other Crimes				
All theft not mentioned elsewhere	√	√	√	√
Commercial crime		√	√	√
Shoplifting		√	√	√

[1] These models show a systematic relationship between the measurement error (under reporting of crime) and the explanatory variable community disadvantage. This is because under reporting is likely to be higher among disadvantaged communities. Because of this systematic relationship we cannot ignore the measurement error in the models containing these crimes as the dependent variables. These crimes are usually executed for economic gain, (refer to section 4.5.3 for more detail).

[2] Omitted variables are as a result of the *mobile* nature of crimes that are committed for economic reasons. This mobility makes it difficult to establish a relationship between the crimes and social, economic or structural characteristics of communities as these crimes are not area specific. This is further detailed in section 4.5.3.

[3] The errors in these columns are influenced by model specification and not specifically by the type of outcome variable.

The grey area of the table shows the crimes that are not included in our current study. We are interested in contact and contact related crimes only which are crimes that occur as a result of social behaviour and not economic behaviour.

Appendix 4.E: Probit Results of Household Crime Reporting

Table 4.E1: Household Reporting of Crime

Dependent Variable	(1) Burglary	(2) Property	(3) Hijacking	(4) Vehicle
Female	-0.0372	-0.140**	-0.0390	-0.213**
	(0.0367)	(0.0581)	(0.130)	(0.100)
Coloured	-0.0418	0.441***	-0.919***	-0.103
	(0.0881)	(0.117)	(0.0370)	(0.181)
White	-0.145*	0.280**	-0.262	0.317***
	(0.0762)	(0.112)	(0.280)	(0.120)
Income bracket 3000-5999	-0.110*	-0.0653	-0.995***	0.327**
	(0.0637)	(0.0860)	(0.00365)	(0.144)
Income bracket 6000-11999	-0.0355	0.144		0.150
	(0.0635)	(0.102)		(0.195)
Income bracket 12000-23999	-0.128	0.00865	-0.991***	-0.00175
	(0.0791)	(0.103)	(0.00705)	(0.220)
Income bracket 24000-47999	-0.0640	-0.136	-0.999***	0.247
	(0.0799)	(0.0873)	(0.00106)	(0.174)
Income bracket 48000-95999	-0.0996	0.0962	-0.998***	0.281
	(0.0935)	(0.133)	(0.00348)	(0.174)
Income bracket 96000 and above	-0.125	-0.0109	-0.926***	0.424***
	(0.118)	(0.141)	(0.0372)	(0.125)
Education, grade1-4	-0.179*	-0.197**		-0.470**
	(0.0979)	(0.0880)		(0.202)
Education, grade5-7	-0.0688	-0.233***	0.337***	-0.169
	(0.0819)	(0.0834)	(0.0904)	(0.295)
Education, grade8-11	-0.0685	-0.258**	0.968***	-0.185
	(0.0713)	(0.107)	(0.0396)	(0.267)
Education, diploma/certificate without matric	0.164***	-0.0803		-0.169
	(0.0551)	(0.145)		(0.347)
Education, matric (grade 12)	-0.0228	-0.291***	0.325***	-0.158
	(0.0800)	(0.0635)	(0.117)	(0.287)
Education, higher than matric	-0.0755	-0.222**	0.421***	-0.100
	(0.102)	(0.0872)	(0.146)	(0.297)
Free State	-0.0718	0.0697		-0.246
	(0.0936)	(0.120)		(0.189)
Gauteng	0.0211	-0.0638		-0.0502
	(0.118)	(0.127)		(0.249)
KwaZulu Natal	-0.0280	0.155	-0.953***	0.0987
	(0.114)	(0.152)	(0.0274)	(0.215)
Limpopo	-0.0244	0.124	-0.994***	0.120
	(0.0907)	(0.145)	(0.00594)	(0.228)
Mpumalanga	-0.0556	0.0682	-0.953***	-0.576***
	(0.114)	(0.186)	(0.0267)	(0.0833)
North West	0.0672	0.0977	-0.971***	-0.372**
	(0.0735)	(0.133)	(0.0426)	(0.172)
Northern Cape	-0.0594	0.180	-0.934***	0.197
	(0.106)	(0.166)	(0.0512)	(0.265)
Western Caoe	-0.126	0.366**		-0.200
	(0.115)	(0.167)		(0.219)
Observations	548	308	48	147

- Standard errors in parentheses
- *** p<0.01, ** p<0.05, * p<0.1 indicates significance at 1%, 5% and 10%
- Reference for gender is male; reference for population group is Black; reference dummy for income bracket is 1-2999; reference dummy for education is none; reference for province is Eastern Cape
- The empty coefficients in the hijacking model are as a result of the small sample size. As a result all those in income bracket6000-11999 for example stated that they had reported crime, as a result Stata drops this coefficient. This is the same for the education and province coefficients.

Appendix 4.F: Results Based on 2 Waves: 2002 and 2005

Table 4.F1: Results Based on 2002 and 2005 Only

Dependent variable	(1) Contact crime	(2) Contact related crime	(3) Contact crime	(4) Contact related crime
Community disadvantage	0.902**	0.604**	3.057***	2.183***
	(0.374)	(0.301)	(0.697)	(0.591)
Community disadvantage*Majority Coloured			-2.873***	-2.344***
			(0.958)	(0.812)
Community disadvantage*Majority White			-3.010***	-2.075***
			(0.843)	(0.714)
Residential instability	0.117	0.209	1.060	0.984*
	(0.247)	(0.200)	(0.659)	(0.558)
Residential instability*Majority Coloured			-1.196	-1.160*
			(0.727)	(0.616)
Residential instability*Majority White			-1.152	-0.766
			(0.735)	(0.622)
Family disruption	0.0624	0.0420	0.252***	0.184**
	(0.0432)	(0.0348)	(0.0910)	(0.0771)
Family disruption*Majority Coloured			-0.235**	-0.151*
			(0.105)	(0.0891)
Family disruption*Majority White			-0.266**	-0.221**
			(0.115)	(0.0977)
Lack of basic services	0.0353**	0.0319***	0.0667***	0.0587***
	(0.0144)	(0.0116)	(0.0225)	(0.0190)
Lack of basic services *Majority Coloured			-0.0581*	-0.0491*
			(0.0294)	(0.0249)
Lack of basic services *Majority White			-0.0359	-0.0235
			(0.0315)	(0.0266)
Head unemployed	0.0214	-0.108	0.843	0.222
	(0.443)	(0.357)	(0.526)	(0.446)
Head unemployed*Majority Coloured			-1.556**	-0.956*
			(0.625)	(0.530)
Head unemployed*Majority White			-1.346	-0.421
			(0.832)	(0.705)
Year 2005	-0.113***	-0.0829***	0.0867***	0.0705***
	(0.0306)	(0.0247)	(0.0285)	(0.0241)
Constant	-11.81	-20.78	-16.85	-18.46
	(24.77)	(19.98)	(24.38)	(20.65)
Observations	154	154	154	154
R-squared	0.336	0.296	0.618	0.553
Number of communities	77	77	77	77

- Standard errors in parentheses
- *** p<0.01, ** p<0.05, * p<0.1 indicates significance at 1%, 5% and 10%
- This table is of similar to table 4.2 in section 4.6.2. The only difference being that we use two years, 2002 and 2005 and omit 2006. This is to check whether the imputation of the missing family disruption variable in 2006 is valid. The results of the two tables are similar implying that our imputation is indeed valid.

Appendix 4.G: Results Based on Individual Basic Services

Table 4.G1: Results Based on Water Services

Dependent variable	(1) Contact crime	(2) Contact related crime	(3) Contact crime	(4) Contact related crime
Community disadvantage	0.642***	0.307	2.329***	1.670***
	(0.245)	(0.187)	(0.697)	(0.539)
Community disadvantage*Majority Coloured			-1.944**	-1.467**
			(0.796)	(0.615)
Community disadvantage*Majority White			-2.245***	-1.868***
			(0.765)	(0.591)
Residential instability	0.350*	0.303**	0.933*	0.777**
	(0.190)	(0.145)	(0.479)	(0.370)
Residential instability*Majority Coloured			-0.619	-0.700*
			(0.537)	(0.414)
Residential instability*Majority White			-0.930	-0.438
			(0.565)	(0.437)
Family disruption	0.0506	0.0502**	0.103	0.0871*
	(0.0306)	(0.0234)	(0.0628)	(0.0485)
Family disruption*Majority Coloured			-0.0538	-0.0355
			(0.0746)	(0.0576)
Family disruption*Majority White			-0.123	-0.0998
			(0.0919)	(0.0710)
Water	0.234***	0.209***	0.00728	0.0626
	(0.0857)	(0.0654)	(0.129)	(0.0998)
Water *Majority Coloured			0.433***	0.268**
			(0.141)	(0.109)
Water *Majority White			0.345**	0.201*
			(0.143)	(0.110)
Head unemployed	0.0557	-0.0175	0.260	0.102
	(0.168)	(0.128)	(0.276)	(0.213)
Head unemployed*Majority Coloured			-0.407	-0.269
			(0.361)	(0.279)
Head unemployed*Majority White			-0.258	-0.235
			(0.382)	(0.295)
Year 2005	-0.170***	-0.135***	0.184***	0.142***
	(0.0384)	(0.0293)	(0.0374)	(0.0289)
Year 2006	-0.244***	-0.191***	-0.0691***	-0.0495***
	(0.0360)	(0.0275)	(0.0236)	(0.0182)
Constant	-35.02*	-30.14**	-40.97**	-30.14**
	(18.98)	(14.49)	(19.26)	(14.87)
Observations	231	231	231	231
R-squared	0.316	0.301	0.475	0.451
Number of communities	77	77	77	77

- Standard errors in parentheses
- *** p<0.01, ** p<0.05, * p<0.1 indicates significance at 1%, 5% and 10%

Table 4.G2: Results Based on Electricity Services

Dependent variable	(1) Contact crime	(2) Contact related crime	(3) Contact crime	(4) Contact related crime
Community disadvantage	0.546**	0.219	2.514***	1.612***
	(0.247)	(0.190)	(0.686)	(0.526)
Community disadvantage*Majority Coloured			-2.242***	-1.497**
			(0.802)	(0.614)
Community disadvantage*Majority White			-2.491***	-1.836***
			(0.764)	(0.585)
Residential instability	0.358*	0.322**	1.214**	0.984**
	(0.196)	(0.151)	(0.512)	(0.392)
Residential instability*Majority Coloured			-1.004*	-0.987**
			(0.566)	(0.433)
Residential instability*Majority White			-0.962	-0.450
			(0.594)	(0.455)
Family disruption	0.0553*	0.0530**	0.0938	0.0692
	(0.0315)	(0.0243)	(0.0709)	(0.0543)
Family disruption*Majority Coloured			-0.0419	-0.0172
			(0.0831)	(0.0637)
Family disruption*Majority White			-0.0984	-0.0601
			(0.102)	(0.0781)
Electricity	0.227	0.138	-0.616	-0.536
	(0.190)	(0.147)	(0.445)	(0.341)
Electricity *Majority Coloured			0.954*	0.802**
			(0.497)	(0.381)
Electricity *Majority White			1.093*	0.775*
			(0.561)	(0.430)
Head unemployed	0.0940	0.0410	0.580**	0.337
	(0.179)	(0.138)	(0.287)	(0.220)
Head unemployed*Majority Coloured			-0.742**	-0.503*
			(0.372)	(0.285)
Head unemployed*Majority White			-0.656	-0.439
			(0.466)	(0.357)
Year 2005	-0.107***	-0.0728***	0.0876***	0.0660***
	(0.0294)	(0.0227)	(0.0294)	(0.0225)
Year 2006	-0.179***	-0.130***	-0.0786***	-0.0565***
	(0.0254)	(0.0196)	(0.0265)	(0.0203)
Constant	-35.89*	-32.00**	-46.86**	-34.55**
	(19.55)	(15.08)	(20.39)	(15.62)
Observations	231	231	231	231
R-squared	0.288	0.257	0.417	0.400
Number of communities	77	77	77	77

• Standard errors in parentheses
• *** p<0.01, ** p<0.05, * p<0.1 indicates significance at 1%, 5% and 10%

Table 4.G3: Results Based on Sanitation Services

Dependent variable	(1) Contact crime	(2) Contact related crime	(3) Contact crime	(4) Contact related crime
Community disadvantage	0.568**	0.248	3.081***	2.002***
	(0.247)	(0.188)	(0.635)	(0.489)
Community disadvantage*Majority Coloured			-2.841***	-1.919***
			(0.749)	(0.576)
Community disadvantage*Majority White			-2.954***	-2.145***
			(0.717)	(0.551)
Residential instability	0.359*	0.302**	1.145**	0.888**
	(0.194)	(0.148)	(0.495)	(0.380)
Residential instability*Majority Coloured			-1.111**	-0.994**
			(0.559)	(0.430)
Residential instability*Majority White			-1.170*	-0.558
			(0.609)	(0.468)
Family disruption	0.0495	0.0491**	0.136**	0.108**
	(0.0311)	(0.0238)	(0.0635)	(0.0488)
Family disruption*Majority Coloured			-0.0824	-0.0508
			(0.0763)	(0.0587)
Family disruption*Majority White			-0.127	-0.103
			(0.0918)	(0.0706)
Sanitation	0.0972	0.112**	-0.0105	0.0569
	(0.0651)	(0.0496)	(0.0770)	(0.0593)
Sanitation *Majority Coloured			0.184**	0.0876
			(0.0781)	(0.0601)
Sanitation *Majority White			0.260*	0.141
			(0.133)	(0.102)
Head unemployed	0.184	0.0987	0.619**	0.385*
	(0.164)	(0.125)	(0.279)	(0.214)
Head unemployed*Majority Coloured			-0.932**	-0.668**
			(0.373)	(0.287)
Head unemployed*Majority White			-0.475	-0.389
			(0.405)	(0.311)
Year 2005	-0.0705***	-0.0417**	0.0437	0.0307
	(0.0264)	(0.0202)	(0.0272)	(0.0209)
Year 2006	-0.204***	-0.164***	-0.143***	-0.120***
	(0.0335)	(0.0256)	(0.0422)	(0.0325)
Constant	-36.05*	-30.12**	-29.99	-22.48
	(19.39)	(14.79)	(20.12)	(15.48)
Observations	231	231	231	231
R-squared	0.292	0.277	0.435	0.414
Number of communities	77	77	77	77

- Standard errors in parentheses
- *** p<0.01, ** p<0.05, * p<0.1 indicates significance at 1%, 5% and 10%

Chapter 5:
Conclusion: Synopsis of Findings, Policy Implications, and Areas for Future Study

To reiterate, this thesis presented an empirical investigation of the role of social interaction in risk taking, including. This is against the backdrop of an exponential growth of interest in including social interaction in the economic modelling of human behaviour and premised on the fact that the presence of the identification problem (the effects of unobservables, selection bias and simultaneity and the related *reflection problem*) has made the current literature skewed in favour of theory rather than empirical investigation.

The thesis investigated three different research areas under the overall theme of social interaction while taking into consideration background information regarding identification into the economic modelling of our research problems involving risk taking behaviours. These three research areas include the role of *social norms* and *social pressure* in influencing multiple partnerships, the effects of *expected income*, *expected health* and contextual effects of *health uncertainty* on *risky sex* and how community social disorganisation namely *family disruption*, *residential instability*, *community disadvantage*, *lack of basic services* and *unemployment* can induce individuals' to resort to *crimes of social behaviour,* specifically *contact crime* and *contact related crime*.

Each of these three research areas provided a stand-alone paper that constitutes a chapter in the thesis. To enable the investigation three different data sources were utilised. This included data from the Cape Area Panel Study (CAPS) of young adults in Metropolitan Cape Town, South Africa. The second data source is the cross sectional Demographic and Health Surveys (DHS) of six sub-Saharan African (SSA) countries, namely Tanzania, Kenya and Uganda in East Africa; and Swaziland, Lesotho, and Zimbabwe from Southern Africa. The final data source consists of crime statistics for Cape Town from the South Africa police service (SAPS). The current concluding chapter is structured as follows. Section 5.1 provides a synthesis of our findings from the three stand-alone papers covered in chapter two, three and four. Thereafter section 5.2 gives the policy implication of our research findings and the final section 5.3 recommends areas for future research.

5.1 Synopsis of the Findings

As previously stated the aim of this thesis was to offer an economic perspective on the influence of social interaction on risk taking. In general, the thesis found social interaction to be significant in influencing risky behaviours and this appears to be the case for both men and women and in both the East and Southern African regions. We provide three key findings.

First this study moved outside the arena of current conventional individualistic empirical research that has explained multiple partnerships by individual level factors and has presented the influence of *social norms* and the related *social pressure* as equally important. We do indeed find that *social norms* and *social pressure* caused by social group interaction to be positive and significant in influencing risky sexual behaviour, in the choice of number of sex partners. The results also show that for both men and women choosing the number of sex partners to engage with is affected by the prevailing *social norms* and *social pressure*.

Secondly, we discovered the pivotal role played by *expected health* and the underlying contextual *health uncertainty* in predicting the sexual risk associated with HIV infection. Surprisingly, this pattern is observed for both the East Africa and Southern Africa regions. This is despite of the fact that the two regions vary by the level of risk related to sexual behaviour where the Southern Africa region is deemed to be more risky than the Eastern Africa region in existing literature. The implication of this finding is that *expected health* and the associated contextual *health uncertainty* are instrumental in influencing sexual behaviour regardless of the magnitude of sexual risk taking.

Thirdly, we find that community social disorganisation, specifically *family disruption*, *residential instability*, *community disadvantage*, *lack of basic services* and *unemployment*, creates conditions for *contact crime* and *contact related crime* to flourish. This is especially observed among the majority Black population group communities in comparison to the majority Coloured or majority White population group communities.

5.2 Policy Implication

Overall the thesis has established that social interaction matters in risk taking behaviours and thus should be incorporated into policies directed at changing human behaviour.

For the reason that we find social interaction especially social norms have great implication for the '*zero new HIV infection*' vision promoted by UNAIDS, advocating for among others sexual behaviour change. This is because the prevailing *social norms* determine the number of sex partners an individual is likely to have. As such it becomes apparent that in order to change behaviour, especially multiple sexual partnerships and achieve a zero new infection rate we need to establish interventions that target the prevailing *social norms*.

In addition our findings do offer support for the validity of the UNAIDS '*know your local epidemic*' which advocates for knowing the contextual factors, as health and health uncertainty appear to be the background factors hindering change in sexual behaviours. Consequently, this study points to confrontation of these factors being pivotal in the fight against HIV/AIDS.

Lastly, our results suggest that an investment in social capital is essential at the community level, while improving family structures is of outmost important at the household level in curbing crimes of social behaviours, specifically *contact crime* and *contact related crime*.

5.3 Areas for Future Research

It is often argued that social norms are essential in understanding choices and variation in behaviour. This thesis has provided empirical evidence of how underlying social norms control the choice of multiple partnerships. It would be of interest to determine social norms that address other sexual risk taking behaviours such as social norms related to *condom use* or social norms related to *sexual debut*.

Another contextual factor that is dominant in the sub-Saharan Africa (SSA) region is *food insecurity*, especially given that most countries in this region are the most vulnerable to climate change. It is stated that climate change will lead to a decrease in food availability. It would be of interest to determine how this contextual factor would affect sexual risk taking. Particularly, given that research has highlighted that one of the reasons for transactional sex is dietary needs. The question then becomes how an increase in *food insecurity* as a contextual effect will influence sexual behaviours associated with HIV risk.

The crimes of social behaviour that were included in this thesis are the *contact crimes* and *contact related crimes*. It would be of interest to establish how other crimes such as *property crimes* and *policing crimes* are influenced by prevailing social structures. However the difficulty arises as a result of the *mobile* nature of these crimes which makes it difficult to match the criminals' place of residence to places where these crimes occur. Our current data sources do not permit such analysis, and these crimes where therefore excluded in the current study.