Sexual Risk Taking among Young Adults in Cape Town

- Effects of Expected Health and Income

Mintewab Bezabih*

Andrea Mannberg†

Martine Visser‡

Abstract

This paper empirically assesses links between expectations of future health and income on sexual risk taking on a sample of young adults in Cape Town, South Africa. An important contribution of the paper lies in combining a wide range of variables measuring risky sexual behavior such that the maximum information possible is extracted from, and adequate weights are attached to each measure, as opposed to previous studies that are based on individual measures or arbitrary aggregations. The findings indicate that expected income and health and future uncertainty are significant determinants of current patterns of sexual risk taking. From a policy perspective, the results suggest that reducing poverty and improving social insurance as well as reducing the taboo related to talking about HIV may constitute important issues to be addressed.

Keywords: HIV/AIDS, Health risk, Risk aversion

JEL-Classification: D81, D84, D91, I10

* University of Portsmouth

† Umeå University and the Doctoral Programme in Population Dynamics and Public Policy

‡ University of Cape Town

Financial support from SIDA/SAREC, the Graduate School in Public Policy and Population Dynamics and the EfD Initiative is gratefully acknowledged. The authors are indebted to K.G. Jöreskog for helpful instructions related to the LISREL 8.80 software, Kurt Brännäs, Jesper Stage, Jörgen Hellström, and William Greene for guidance on econometric issues, participants at the SALDRU seminar at Univeristy of Capetown, and staff at the Department of Economics at Umeå University for helpful comments and suggestions. The usual disclaimer applies.

1. Introduction

In 2008, 67 percent of all HIV infections in the world could be found in sub-Saharan Africa and HIV/AIDS is today one of the main causes of premature death in the region (UNAIDS, 2009; Rao *et al.,* 2006). In spite of rising knowledge levels and a nearly global awareness of HIV, sexual risk taking still prevail in many HIV infested countries. Indeed, although HIV prevalence among commercial sex workers (CSW's) can be as high as 70 percent in some areas, demand for unprotected commercial sex has been found sufficiently high to produce a price of 4 times that of commercial sex with a condom (Abdool *et al.*, 1995; Swart-Kruger and Richter, 1997; Robinson & Yeh, 2008).

The purpose of this paper is to address the issue of seemingly irrational sexual risk taking in the presence of HIV/AIDS. More specifically, we test the hypothesis that future prospects of health and income are interlinked with present sexual risk taking on a dataset consisting of young adults in Cape Town, South Africa.

The analysis of how contextual uncertainty and expectations of future health and income affect incentives to engage in safe sex practices is of interest for many reasons. First and foremost, interventions aimed to reduce the spread of HIV have to date been relatively ineffective. Although educational campaigns have contributed to a nearly universal awareness of HIV and to some reductions in sexual risk taking in terms of an increase in condom use and a reduction in the number of sex partners, condom use remains at best inconsistent and the lack of behavioral change in high risk groups effectively means that there are almost no signs of significant reductions in HIV rates[1] (e.g. Caldwell, 1999; Caldwell, 2000; Bloom *et al.*, 2000; Buvé *et al.*, 2001; Williams *et al.*, 2003; Mwaluko *et al.,* 2003; Hearst & Chen, 2004; UNAIDS, 2009).

In addition, although the development of antiretroviral (ARV) treatment has improved health and extended the life length of HIV positive individuals, HIV remains incurable and AIDS is still associated with premature death. Similarly, although access to ARV treatment has improved substantially, many individuals remain both undiagnosed and untreated.[2]

[1] Lagarde *et al.* (2001) estimate condom coverage at 27-31 percent for men and 11-17 percent for women in a number of highly affected cities in sub-Saharan Africa. In Zambia the share of unprotected sexual acts with non-cohabiting partners actually increased during the late 1990's.

[2] The percentage of pregnant women receiving antiretroviral treatment has improved substantially over that past couple of years. In many countries with reports of coverage rates, numbers are as high as 80 percent. However, although the trend in antiretroviral treatment distribution in poor countries has been positive, the number of new HIV infections each year outnumbers the increase in individuals on ARVs, by 2.5 to 1; See e.g. UNAIDS (2008); Rao *et al.* (2006).

Our findings indicate that AIDS policy, in order to be effective, may need to take a more holistic approach than only focusing on information and condom availability. If uncertainty about future prospects affects incentives to invest in health in the present, public policy on information campaigns and condom distribution may need to be complemented by additional investments in social insurance and poverty reduction.

Health related uncertainty and low expectations of the future are naturally not the only explanations to excessive risk taking. Myths and misinformation about HIV/AIDS and condoms still exist, and most likely constitute a part of the explanation. However, the lack of HIV knowledge cannot fully explain persisting unsafe sex practices; a number of surveys and studies on sexual behavior show that sexual risk taking persists in spite of relatively adequate HIV knowledge and access to condoms (e.g. Campbell, 1997; Varga, 1997; Pettifor *et al.* 2000; MacPhail and Campbell, 2001; Pettifor *et al,* 2004). HIV susceptible populations, in both rich and poor countries, are mainly comprised of individuals living under a greater degree of uncertainty than the average population. In sub-Saharan Africa, the main populations at risk are constituted by refugees, commercial sex workers (CSW's), miners, military personnel and transport workers[3] (UNAIDS & IOM, 2003). For these groups, HIV is likely to represent a significant threat but it may not be dominant in daily life.[4]

The empirical analysis presented below is based on a longitudinal dataset on young adults in Cape Town, South Africa (The Cape Area Panel Study (CAPS))[5]. In South Africa, the spread of HIV/AIDS has been explosive: HIV prevalence has increased from less than 1 percent in 1990, to over 20 percent only 10 years later[6] (Pettifor *et al.*, 2004; Simbayi *et al.*, 2004; Connolly *et al.*,

---

[3] For example, miners in sub-Saharan Africa face daily threats of mutilation or death in work related accidents (2.9 % probability of being killed in a work related accident and 42 % probability of suffering a reportable injury during a 20-year working life, according to the South African Chamber of Mines). Likewise, military personnel live under great risk of having their future cut short. In Thailand soldiers have been found to visit prostitutes to a higher extent, and to use condoms to a lesser extent, than other Thai men (VanLandingham *et al.* 1993). The relatively higher level of consumption of commercial sex in these groups is perhaps partly explained by the fact that, for example, miners and migrant workers live away from their families. However, this fact does not explain the low degree of condom use.

[4]In Sierra Leone, estimated HIV prevalence was estimated lie between 60 and 70 percent among soldiers in 2002 and HIV frequencies among prostitutes increased from 26.7 percent to 70.6 during the civil war (UNAIDS/WHO (2002).

[5] The Cape Area Panel Study Waves 1-2-3 were collected between 2002 and 2005 by the University of Cape Town and the University of Michigan, with funding provided by the US National Institute for Child Health and Human Development and the Andrew W. Mellon Foundation. Wave 4 was collected in 2006 by the University of Cape Town, University of Michigan and Princeton University. Major funding for Wave 4 was provided by the National Institute on Aging through a grant to Princeton University, in addition to funding provided by NICHD through the University of Michigan.

[6] Estimates are based on pregnant women visiting public clinics in 1990 and 2000. HIV prevalence estimates based on antenatal clinic visitors are problematic. Pregnant women are bound to be both sexually active and to have engaged in unprotected sex. Hence, the estimates presented above are biased upward. However, national figures on HIV are no less alarming than previous estimates from antenatal clinics. Nationally representative surveys on HIV

2004). Age group analysis suggests that HIV incidence is substantially overrepresented among youth in their late teens and that HIV prevalence varies both between population groups and between regions (e.g. MacPhail & Campbell, 2001; Lam *et al.,* 2008).

2. Previous economic research

Sexual behavior may not be a conventional subject within the economic field of research. However, research in economics has provided interesting insights concerning the incentives to engage in sexual risk taking in the presence of HIV. Important contributions include Philipson & Posner (1993; 1995), Schroeder and Rojas (2002), Kremer (1996), Gertler *et al.,* (2005), Robinson & Yeh (2008) and Oster (2007a). Philipson and Posner (1993) show that the inclusion of marginal costs and benefits associated with unprotected casual sex improves prediction accuracy of epidemiological models. Early epidemiological models predicted that HIV prevalence would explode as the pool of potential virus hosts increased. However, since an increase in HIV prevalence increases the marginal cost of unsafe sex for HIV negative individuals, it also increases incentives to abstain from risky sex. The predictions of Philipson and Posner (1993) thus suggest that the rate at which HIV spread will not necessarily increase as HIV prevalence increases (see also Schroeder and Rojas, 2003). Philipson and Posner (1995) further suggest that the prevailing differences in HIV prevalence rates between rich and poor countries can be explained in terms of an inelastic supply of sexual services among CSW's in poor countries. This hypothesis is supported by a number of empirical studies in poor countries suggesting that CSW's accept unprotected sex due to the substantial income loss associated with condom use. The driving mechanism behind the price differential for condom-free sex is partly explained by strong preferences for unprotected sex among clients (e.g., Rao *et al.,* 2003; Gertler *et al.,* 2005; Robinson & Yeh, 2008).

Kremer (1996) provides an interesting alternative explanation to observed differences in epidemiological development. Kremer note that the marginal cost associated with HIV risk does not only depend on the per-coital risk[7] of an HIV infection, it also depends on the probability of being HIV positive since, disregarding altruistic objectives, an HIV positive individual has little to

prevalence have only recently been available. In 2003, HIV prevalence amounted to 19.2 percent in a representative sample of the adult population, and the HIV frequency among young adults was estimated at around 10 percent (Pettifor *et al.*, 2004; Simbayi *et al.*, 2004; Connolly *et al.*, 2004)

[7] i.e., the probability of acquiring HIV through 1 unprotected sexual act

lose from having unprotected sex.[8] This implies that, at sufficiently high perceived risk of HIV, individuals may deem the likelihood of already being infected with the virus to be sufficiently high to make abstaining from risky sex relatively meaningless. Hence, within countries or groups of individuals where sexual activity is initially high, an increase in HIV prevalence may create reckless sexual behavior.

The empirical analysis presented in this paper relates most closely to a number of economic papers that incorporates the effect of physical and social context on health related behavior.[9] The empirical analysis of Dow *et al.* (1999) suggests that health policy in poor countries may have positive spill-over effects on private health seeking behavior. Interventions, such as immunization programs, reduce pressing mortality risks and thereby increase the marginal benefit of private health investments. Benz (2005) follows the same line of argument and suggests that the correlation between high HIV prevalence and experiences of civil conflict can at least partly be explained by increased in uncertainty associated with crumbling civic institutions. Dinkelman *et al.* (2007) do not explore the direct link between economic shocks and sexual risk taking in South Africa. However, the empirical analysis suggests that the sexual debut of youth from poor households occurs at younger age than young adults from richer households. In addition, poor young adults use condoms less frequently than richer young adults. In the paper most closely related to our analysis, Oster (2007a) suggests that a reduction in expected life length or future earnings reduces the expected cost of risky sexual behavior. Using cross-sectional data from nine African countries[10], her findings indicate that, while HIV knowledge and HIV frequency does not have a significant effect on the number of sex partners, income levels and expected life length do affect sexual risk taking.[11]

Although the above mentioned research has contributed with important insights on the mechanisms behind sexual risk taking, several questions remain unanswered. Gertler *et al.* (2005) provide an answer to why commercial sex workers demand a higher price for supplying unprotected sexual services. However, their analysis does not address the fact that the buyer of the sexual service is willing to pay a higher price for the risky alternative. Given the high risk of

---

[8] Admittedly, unprotected sex for an HIV positive individual is still risky since the individual may acquire another STI which in turns reduces the immune system's ability to keep maintain the viral load at low levels. See, e.g., Piasani, 2008

[9] See e.g. Dow *et al.* (1999); Benz (2005); Dinkelman *et al.* (2007); Oster (2005; 2007a).

[10]Demographic and Health data are available for Benin, Burkina Faso, Ethiopia, Ghana, Kenya, Malawi, Mali, Namibia and Zimbabwe.

[11]In order to deal with the obvious endogeneity between life expectancy and HIV prevalence, malaria frequency and maternal mortality is used as a proxy for life expectancy. Oster further acknowledges the link between sexual promiscuity and HIV prevalence, and therefore uses the distance to the (believed) origin of the HIV virus (the Democratic Republic of Congo) as an instrument for HIV frequency.

infection, a rational buyer should intuitively be willing to pay more for *safe* sex rather than for *unsafe* sex. Similarly, although the consideration of the net effect of an increase in HIV prevalence on the marginal cost of unprotected sex may explain reckless behavior, it does not fully explain differences in reactions towards changes in HIV prevalence and HIV knowledge *within* high- or low activity groups.[12] Oster (2005, 2007a, 2007b), Dow (1999), and Benz (2005) all analyze the link between health behavior and future prospects, but none explicitly study how sexual behavior and uncertainty of future prospects are interlinked. The empirical analysis presented below builds on the theoretical work by Mannberg (2010) and can be seen as a complement to the work of Oster (2007b). Mannberg's theoretical analysis suggests that health related uncertainty contributes to a lower expected benefit of abstaining from pleasurable but risky sex in the present. This implies that, if the social and physical context is more uncertain in sub-Saharan Africa than in other parts of the world, then investments in health should be more risky there than elsewhere. Hence, although individuals in poor countries appear to act myopically, they may be acting completely rational within the context of uncertainty. In order to give a full understanding of the model, the theoretical framework is presented in more detail below.

3. Theoretical Framework

The theoretical framework used for the empirical analysis below is a simplified version of the theoretical work of Mannberg (2010), representing a 2-period model where an individual derives utility from consumption, health and sexual activity.

The preferences in period t are described by the utility function, $U(\boldsymbol{c}, \boldsymbol{h}, \boldsymbol{x_{us}}, \boldsymbol{x_s})$ where $\boldsymbol{c} = (c_1, c_2)$ is consumption of other goods, $\boldsymbol{h} = (h_1, h_2)$ is health, $\boldsymbol{x_s} = (x_{s,1}, x_{s,2})$ is consumption of safe sex, and $\boldsymbol{x_{us}} = (x_{us,1}, x_{us,2})$ is consumption of unsafe sex. The instantaneous utility function is assumed to be twice continuously differentiable, increasing in each argument, and strictly concave. Safe and unsafe sex are assumed to be additively separable. Safe sex can be represented by either having only one sex partner[13] or using a condom. In addition, the utility of safe sex is scaled by a factor $\alpha; 0 < \alpha < 1$, implying that unsafe sex confers a higher utility than safe sex. Since the individual is assumed to die in the end of time period 2,

[12] Within the high risk groups during the early phases of the epidemic, such as the gay community, sexual risk taking fell as a response to HIV knowledge and increasing HIV prevalence. However, within HIV susceptible groups such as intravenous drug users, migratory workers, and refugees there are few signs of changes in sexual risk taking (UNAIDS & IOM, 2003).

[13] Naturally, being monogamous is not a sufficient condition for sex to be safe. The individual also need to make sure that his sex partner is monogamous and HIV negative to begin with.

unsafe sex does not have any health related cost in this period. Let us therefore disregard sexual consumption in time period 2.

The probability of contracting HIV via unprotected sexual intercourse is given by $\Pr(HIV)$, with the corresponding probability of staying HIV negative $1 - \Pr(HIV) = \varphi$. If the individual acquires HIV, she/he is assumed to die before reaching time period 2. In addition to the risk of acquiring HIV, the individual also faces an exogenous risk of experiencing changes in health via stochastic health shocks leading to high or low health outcomes, represented by probabilities ρ and $(1-\rho)$, respectively. The spread between high and low health is given by $\gamma = (h_{high} - hlow$. Thus, if $h2$ is the expected value of future health conditional on the individual being alive in period two, then $h_{high} = \tilde{h}_2 + (1-\rho)\cdot\gamma$, and $h_{low} = \tilde{h}_2 - \rho\cdot\gamma$. We define the expected utility function as:

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

where $\beta; 0 < \beta < 1$, is an exogenous discount parameter, and where

$$E[u(c_2, h_2)] = \varphi^{x_{us}} \cdot [\rho \cdot u(c_2, h_{high}) + (1-\rho) \cdot u(c_2, h_{low})] \tag{2}$$

is the expected utility in period two.[14] The intertemporal budget constraint is given by:

$$m_1 = c_1 + p \cdot x_s + q \cdot x_{us} + S \tag{3}$$

$$m_2 + (1+r) \cdot S = c_2 \tag{4}$$

where m_1 and m_2 is income in each time period, assumed to be exogenously given. p is the price of unprotected sex and q the price of unprotected sex with exogenous income in each time period and non-negative savings. Eliminating the intertemporal budget constraint, the first order conditions for safe sex, unsafe sex and saving are given by:

$$\frac{\partial E[U_t]}{\partial x_s} = U_{x_s} = -p \cdot u_{c_1} + \alpha \cdot g_{x_s} = 0 \tag{5}$$

[14] Since an HIV positive individual is assumed to die before entering time period 2, the expected utility of an HIV positive individual in this time period is zero.

$$\frac{\partial E[U_t\,]}{\partial x_{us}} = U_{x_{us}} = -q \cdot u_{c_1} + g_{x_{us}}$$
$$+ \beta \ln(\varphi)\, \varphi^{x_{us}}\left[\rho u_{c_2}(c_2, h_{high}) + (1-\rho) u_{c_2}(c_2, h_{low})\right] = 0 \quad (6)$$

$$\frac{\partial E[U_t\,]}{\partial S} = U_S$$
$$= -u_{c_1} + \beta \cdot \varphi^{x_{us}}\left[\rho \cdot u_{c_2}(c_2, h_{high})\right. \quad (7)$$
$$\left. + (1-\rho) \cdot u_{c_2}(c_2, h_{low})\right] \leq 0$$

The first order conditions together implicitly define the optimal amount of unprotected sex as $x_{us}^* = x_{us}^*(p, q, \rho, \gamma, \varphi, m_1, m_2, \tilde{h}_2)$. In order to derive a hypothesis concerning the effect of health related uncertainty on sexual incentives, we follow Mannberg's (2010) analysis. Accordingly, we analyze the effect of an increased variability of future health which corresponds to an increase in the spread (γ). The effect of improvements in expected health and income is analyzed by performing comparative statics for an increase in $\tilde{h}_2$ and m_2, respectively. The results of the comparative statics are summarized in *proposition 1* below:

Proposition 1

a. $\partial x_{us}/\partial \gamma > 0$ *: An increase in the uncertainty of future health reduces incentives to abstain from unsafe sex*

b. $\partial x_{us}/\partial \tilde{h}_2 < 0$*: And increase in the expected level of future health increases incentives to abstain from unsafe sex.*

c. $\partial x_{us}/\partial m_2 \lessgtr 0$*: An increase in future income has ambiguous effects on incentives to engage in unsafe sexual practices.*

Proof: See the appendix/or for details on the derivations see; Mannberg (2010).

The results show that health related uncertainty reduces the expected value of the future, not only by lowering the expected utility of future health but also in terms of reducing the expected value of future consumption. Hence, incentives to postpone consumption for the future, in terms of abstaining from risky sex, diminish. The negative relationship between risky sexual behavior and an individual's higher expected health value is a reflection of that an individual that expects to be healthy at an old age consumes less unsafe sex than an individual expecting to have poor health later in life at a given level of exogenous uncertainty. An improvement in expected health makes the cost of unsafe sex more salient in terms of an increase in the potential loss induced by unsafe sex consumption. As health and consumption are assumed to be complements, the improvement in expected health not only increases the value of the future directly, but also raises the marginal value of future consumption. Finally, as in the case of improvements in health, an increase in future income increases the expected benefit of surviving to old age, thus reducing incentives to consume unsafe sex. However, the increase in future income also reduces the need to save for old age. A reduction in savings implies the budget space for current consumption, including unsafe sex, increases. Hence, an expected increase in future income presents the individual with conflicting incentives.

4. Empirical approach

The theoretical model, presented above, suggests an empirical model of the form:

$$y_i = \alpha + \beta_1 \cdot \tilde{h}_2 + \beta_2 \cdot \gamma_{health} + \beta_3 \cdot E[m_2] + \boldsymbol{X_i}'\boldsymbol{\beta} + u_i \qquad (8)$$

where y is a measure of sexual risk taking, $\tilde{h}_2$ as in the above is expected future health, γ_{health} is a measure of uncertainty of future health, $E[m_2]$ is expected future income, $\boldsymbol{X}$ is a vector of control variables, and u is the error term. In this section we set up the econometric framework for analyzing the links between risky sexual behavior, economic and health conditions of the individual, and uncertainties associated with them. We start by giving a short description of the data used for the empirical study, and then proceed with a discussion of the operationalization of each set of variables. In *section 4.3* we describe our econometric considerations.

4.1 Description of the data

The data used for the econometric analysis below stems from the Cape Area Panel Study (CAPS), a longitudinal dataset consisting of a representative sample of Cape Town youth.[15] The first wave of CAPS was implemented in 2002 covering 5250 households and 4750 adolescents between the ages 14 and 22. Three additional survey waves were conducted in the years 2004, 2006 and 2009. The CAPS data is extremely rich and has been used in a number of studies including HIV related issues (e.g. Dinkelman *et al.,* 2007; Anderson & Beutel, 2007; Anderson *et al.,* 2007). The dataset includes detailed information on current socioeconomic characteristics, income and health of young adults, their households (parents) and in the community. In addition,

[15] The Cape Area Panel Study Waves 1-2-3 were collected between 2002 and 2005 by the University of Cape Town and the University of Michigan, with funding provided by the US National Institute for Child Health and Human Development and the Andrew W. Mellon Foundation. Wave 4 was collected in 2006 by the University of Cape Town, University of Michigan and Princeton University. Major funding for Wave 4 was provided by the National Institute on Aging through a grant to Princeton University, in addition to funding provided by NICHD through the University of Michigan.

CAPS contains information on the young adults' sexual behavior, their HIV related knowledge and on how often the young adult discusses HIV issues with parents or step-parents.

The empirical analysis is based on the third survey wave, collected in 2004, since it consists of a wider range of information regarding risky sexual behavior of young adults and regarding expectations of future health[16]. While the previous waves also include such information, certain variables are missing, which made use of the full panel difficult for our purpose. In addition, in order to use the information provided by parents we had to restrict the sample further to young adults whose parents, step-parents or guardians provided information in the parent questionnaire. Taken together this implies that of the nearly 5000 young adults taking part in the CAPS study, only a sample of 2618 individuals were included in the study. The descriptive statistics of the sample is presented in *Table 1*, below.

As can be seen in *Table 1*, most of the respondents in the sample consider using a condom as an effective method to protect against HIV (83 percent), and roughly half of the sample considers abstaining from sex altogether to protect against HIV (47 percent). However, only 35 percent of the sample used condoms consistently[17] with their last sex partner, and 49 percent of the young adults that used condoms inconsistently did this with a non-spousal partner. The descriptive statistics reveal large differences between population groups regarding household per capita income. A *t-test* of group means shows that black young adults have significantly lower mean per capita income than colored and white young adults (p<0.0001), and white young adults are significantly richer than their colored and black counterparts (p<0.0001). However, the difference in subjective life expectancy[18] between black and white young adults is not significant at the 10 percent level, and black respondents have significantly higher life expectancy than colored young adults (p<0.0001). Colored respondents also live in areas with significantly higher variances in life expectancy and use condoms less frequently than white and black young adults (p<0.0001).

---

[16] For example, wave 3 contains information about whether or not the young adult had simultaneous sex partners during his or her last sexual relationship, a well known risk factor for HIV (c.f. Piasani, year). This wave also includes questions concerning the respondent's relationships to his or her last 10 sexual partners, the young adult's life expectancy, and the behavior and expectations of the young adult's parent or guardian.

[17] Where consistent condom use is defined as having "always" used condoms with most recent sexual partner, and inconsistent condom use is defined as having used condoms "usually", "sometimes", or "rarely"

[18] Respondents were asked to estimate their expected life length in terms of answering yes or no to the question: "Do you expect to be alive at the age of 30? …40? …50? …60? …70? …80?"

Table 1: Descriptive statistics

	All	Females	Males	Black	Coloured	White
Number of observations	2618	1420	1198	1011	1346	261
Percent	100	54	46	39	51	10
Life expectancy	63.72	62.91	64.62	68.44	58.88	70.19
Parent life expectancy	68.46	68.74	68.14	71.54	65.21	73.03
Houshold per capita income	1399.40	1342.78	1462.66	517.06	1221.47	5589.64
Variance in life expectancy	311.86	311.38	312.41	266.75	363.94	220.9
Education[a]	0.86	0.89	0.82	0.80	0.87	0.98
Age in 2002	17.38	17.39	17.37	17.57	17.35	16.62
HIV discuss[b]	2.32	2.39	2.24	2.45	2.27	2.07
Relation to HIV/AIDS infected	27	32	22	50	14	9
HIVprotct_abstain (%)	47	49	44	40	47	73
HIVprotct_1sexp (%)	11	22	20	10	26	39
HIVprotct_limit (%)	21	10	11	4	13	25
HIVprotct_condom (%)	83	81	85	88	80	80
Age at first sex	16.46	16.87	16.01	15.98	16.86	17.40
Condom use (ordinal)[c]	1.46	1.58	1.33	1.74	1.37	0.79
Inconsistent condom use[d]	35	40	28	39	35	17
Inconsistent condom use with non-spouse[e]	49	58	39	42	59	39
Sex partners past 12 months[d]	1.16	0.66	0.88	1.02	0.60	0.52
Sex partners in life[d]	1.48	1.24	1.77	2.07	1.13	0.99
Type of relationship to most recent sex partner[d]						
Spouse	4.19	5.28	2.99	6.40	2.12	0.98
Someone loved but not married to	76.94	83.10	70.13	81.77	72.42	69.61
Someone known but not loved	10.36	7.16	13.90	4.93	15.13	20.59
Casual	8.51	4.46	12.99	6.90	10.33	8.82
Time before having sex with most recent sex partner[d]						
More than 1 year	39.52	43.83	34.73	29.40	49.36	50.49
less than 1 year but more than 1 month	36.67	39.01	34.07	38.59	35.08	33.01
less than a month but more than 2 days	18.49	14.69	22.72	25.06	11.74	13.59
2 days or less	5.32	2.47	8.49	6.95	3.82	2.91
Other sex partners in most recent sexual relationship[d]						
Definetely no	79.61	87.70	70.61	69.58	88.68	96.12
Unsure	4.56	3.16	6.11	7.39	1.84	0.97
Definetely yes	15.83	9.13	23.28	23.03	9.48	2.91

a: Dummy for secondary education or higher, b: Ordinal variable for how often YA discussed HIV with parent/step parent during the past 12 months preceeding the interview (1=never. 4=Often), c: Ordinal variable for condom use (1=Always. 4=Rarely), d: % of sexually active, e: % of inconsistent condom users

Given the prevalence of HIV in South Africa in general and in Cape Town in particular, the focus of risky sexual behavior is likely to be highly correlated with HIV incidence, making causal inference difficult. Hence, one likely empirical problem could be disentangling the effect of the influence of HIV prevalence on risky sexual behavior from that of expected income and health (our main variables of interest). The focus of our analysis being on *individuals* enables us to control for the behavioral effect of HIV prevalence, as, for an individual agent, HIV frequency is likely to be exogenous.

4.2 Variable operationalization

4.2.1 Sexual risk taking

Theoretically, sexual risk taking related to HIV may be defined as the quantity of unprotected sexual acts with sexual partners whose HIV status is unknown. However, empirically, it is difficult to separate sexual *activity* from sexual *risk taking*. In general, economic research on sexual risk taking related to HIV or other STIs utilize measures such as condom use, age at sexual debut, and the number of sexual partners in a specified time interval (e.g., Dinkelman *et al.*, 2007). Although it may be argued that inconsistent condom use *or* having numerous sex partners is always to some degree risky, especially in countries where HIV prevalence is high and simultaneous sex partners common, these variables in isolation tell us relatively little about sexual risk taking. For example, inconsistent condom use may be relatively safe if the young adult only has sex with his or her spouse and knows that this partner is HIV negative and monogamous. Likewise, having many sexual partners is risky only if the young adult uses condoms inconsistently. Hence, using *only* condom use, *or* number of sexual partners tells us very little about sexual risk taking.

Despite this, previous research in both economics and psychology have usually resorted to using single variables as indicators of sexual risk taking, such as age of first sex or condom use at last sex. Given no solid theoretical foundation to direct our choice of variables to include in the combined measure or how to combine these measures, we have two challenges at hand: what variables to include and how to sensibly aggregate the alternative risky sex measurement variables. A few papers on sexual risk taking conducted by psychologists have used more adequate measures of sexual risk taking by combining condom use with number of sex partners (i.e. by explicitly asking about condom use with each sexual partner during a specified time interval, see van der Pligt *et al.*, 1996). However, even in these instances, simple aggregations are used where the number of sexual partners over a specified time interval is multiplied with information about condom use.

In light of this, one of the major aims of the paper is to generate a measure of risky sexual behavior by combining a wide range of variables such that the maximum information possible is extracted from, and adequate weights are attached to each measure, as opposed to previous studies that are based on individual measures or arbitrary aggregations. In our case, since we have information on the young adult's sexual behavior in terms of condom use, type of relationship with most recent sexual partner, the time the young adult knew the partner before having sex, whether or not the young adult or his/her partner had additional sex partners during the relationship, and the young adult's number of sexual partners in both life and in the most recent 12 months, we are able to construct such a measure of sexual risk taking.

In order to do this, we draw on a series of psychological studies by van der Velde & van der Pligt (1991), van der Velde *et al.,* (1994) and van der Velde *et al.,* (2002). In these studies, information concerning condom use, number of sexual partners and type of partner is combined in order to measure sexual risk taking among high risk individuals in Amsterdam.[19] We augment their measure by adding more detailed information about the type of relationship, the time the young adult knew the sex partner before having sex, and whether or not the young adult had additional sex partners during the relationship. This gives us a measure of the degree of riskiness associated with the young adult's last sexual partner. In addition, in order to include information related to the frequency of sexual encounters, we also include a measure of sexual activity in terms of the number of sex partners relative to the average in the young adult's age group.[20] However, simply adding and multiplying different indicators of sexual risk taking is problematic for several reasons. First, the weights of the individual variables are ad hoc. Second, due to the ordinal nature of the indicators available in CAPS, it is difficult to interpret the risk level associated with the different values of the constructed variable. In order to derive adequate weights of each indicator, we analyze the data with explanatory factor analysis (EFA) in SPSS and thereafter derive latent factors scores from a confirmatory factor analysis (CFA) for ordinal variables in the LISREL 8.80 software (Jöreskog, 2005; Jöreskog & Moustaki, 2001; Jöreskog *et al.,* 2006).

As mentioned above, it is important to not only analyze the degree of risk associated with the young adult's last sex partner, but also to include information about condom use. However, for rational individuals with a higher number of casual sexual relationships, condom use is likely to be positively correlated with the number of sex partners and with the degree of risk posed by each sex partner. This correlation causes CFA to produce a positive relationship between the degree of condom use and sexual risk taking. Although a high frequency of casual sexual relationships is always to some extent risky, an individual with a consistent use of condoms

[19] In these studies, the type (i.e. private or prostitution partner) and number of sexual partners was multiplied with the degree of condom use (past 4 months preceding the interview). In addition, information related to the frequency of vaginal intercourse with each partner was included in the measure. The measure of condom use and frequency of vaginal intercourse was measured on ordinal scale (from never to always). The measure was constructed by transforming the Likert scales into multipliers. For example: A participant who had 5 private partners with whom the participant often had vaginal intercourse and with whom condoms were never used, and 5 prostitution partners with whom the participant always had vaginal intercourse while used condoms half of the time received the risk score: (5*0.75*1)+(5*1*0.5).

[20] This variable was created by dividing the young adult's lifetime sex partners by the average number of sex partners in his/her age group. In order to facilitate the analysis in LISREL (discussed below) we transformed the variable into an ordinal variable where each group represents quartiles of the distribution.

should intuitively be considered as less exposed to risk than an individual with inconsistent condom use, given that the two individuals have the same number and type of sex partners. Hence, in order to avoid the positive association between condom use and sexual risk, the variable related to condom use is not included in the combined measure constructed by CFA. Information about condom use is instead included in the estimation procedure by multiplying the combined measure by 1 if condoms were used inconsistently and 0 if condoms were used consistently during the young adult's last sexual relationship (see section 4.3, below).

The basic idea behind factor analysis is that the correlation between a set of observed variables may be explained in terms of one or more latent factors theoretically hypothesized to underlie the observed variables. EFA enables us to investigate the number of factors underlying the set of observed variables in terms of an analysis of the variance explained in the indicator variables by each latent construct. EFA also enables us to evaluate the relative strength in the relationship between each indicator variable and the latent construct. Using CFA we may then test whether our hypothesized relationship between the set of indicator variables and the latent construct is significantly better than an unrestricted model assuming no relationships between the observed variables and the latent factor. The basic model that we want to estimate is:

$$\boldsymbol{y} = \boldsymbol{\tau}_y + \boldsymbol{\Gamma}_\mathbf{y} \cdot \boldsymbol{\eta} + \boldsymbol{\delta} \qquad (9)$$

where $\boldsymbol{y}$ is a q-vector of observed variables, and $\boldsymbol{\eta}$ is a vector of latent factors assumed to underlie the observed variables. The $\boldsymbol{\Gamma}_\mathbf{y}$ is a matrix of factor loadings describing the relationship between the observed variables and the latent factors, $\boldsymbol{\tau}_y$ is the mean vector of $\boldsymbol{y}$, and $\boldsymbol{\delta}$ represents a vector of errors. If the latent factors are assumed to be uncorrelated and of unit variances, the individual factor loadings can be estimated by an analysis of the correlation and covariance matrix of $\boldsymbol{y}$ (e.g., Jöreskog & Sörbom, 1979; Kaplan, 2000). However, as can be seen in *Table 1,* the relevant observed variables related to sexual risk taking in CAPS are of ordinal scale. With ordinal variables, conventional correlation and covariance matrices cannot be used to estimate the model parameters. Hence, for EFA and CFA of ordinal variables the *polychoric* correlation is used to estimate the system in equation (9). The polychoric correlation may be estimated by several procedures. The method used in LISREL is based on the "underlying response variable approach". This approach assumes that each observed ordinal variable has an underlying, normally distributed, continuous variable. The procedure estimates the probability of giving a response in a specific category and then estimates the correlation between the unobserved continuous response variables (e.g., Jöreskog & Moustaki 1994; a more detailed description of the estimation procedure of the polychoric correlations is given in appendix A3).

Our hypothesized model, described in equation (10), contains the observed variables; 1) relative number of sex partners to age group (*sexp*), 2) time before having sex (*timesex*), 3) relationship type (*rel type)*, and 4) presence of multiple sex partners (*multiple sexp*). With four observed variables the degrees of freedom are insufficient to test a model of more than one latent factor.[21] Hence, a CFA of four observed variables implies a structure of the following form (A full description of the variables can be found in the appendix):[22]

$$\begin{pmatrix} sexp \\ Timesex \\ rel\ type \\ multiple\ sexp \end{pmatrix} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{pmatrix} \cdot \xi + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \end{pmatrix} \qquad (10)$$

The latent factor scores are estimated in accordance with the approach outlined by Anderson & Rubin (1956) (see also, Jöreskog, 2000).

4.2.2 Expected health, expected income and uncertainty of future health

The main variables of interest for testing the health risk model relate to expected health, expected income and uncertainty of future health. The CAPS dataset contains several measures of the young adults' perception of current and expectation of future health. The variable perhaps most closely related to expected health is the young adult's life length expectancy. However, this variable is potentially endogenous to the young adult's present level and expectation of future sexual risk taking behavior. Hence, we would ideally like to find an instrument that affects life expectancy (or expected health) but is unrelated to sexual risk taking. In previous studies, malaria frequency has been used (Oster, 2007a). However, in the CAPS dataset, and for the Cape Town region, no such variables exist. In order to deal with the endogeneity of the life expectancy variable, we instead instrument this variable with the mean life expectancy in the community in

[21] A naive analysis of all variables related to sexual risk taking was also conducted. Variables associated with relatively high factor loadings (above 0.5) include: age at first sex, number of sex partners relative to the average in the young adult's age group, number of sex partners during the past 12 months, and whether or not the young adult had multiple sex partners at the same time. However, in combination these variables do not produce a statistically viable model. Indeed, age of first sex in general reduce model fit regardless of the other variables included in the analysis. Results are available from the authors upon request.

[22] The variables "multiple sex partners" take the value 0 if the young adult (or his or her partner) had no other sexual partners during the relationship and 1 if the young adult (or his or her partner) did have other partners. Finally, the "relationship type" variable takes the value 1 if the sexual partner was the young adult's spouse, 2 if the sexual partner was a loved girl or boy friend, 3 if the sexual partner was someone known but not loved and 4 if the sexual partner was someone that the young adult did not know well. Since it may be the case that a spouse became infected with HIV before marriage, even marital sex without condoms are measured as mildly risky. For individuals not yet sexually active, the "riskpartner" measure is always zero.

which the young adult resides and with the life expectancy of the young adult's parent (Examples of constructing instruments using aggregate village level variables include Quigley & Raphael, 2005; D'Hombres *et al.*, 2010). It may be argued that the mean of life expectancy suffers from neighborhood effects (i.e. risky individuals may decide to reside in areas where other risky individuals live). However, since the sample used in the econometric analysis consists of young adults, dwelling location is likely to be determined not by the respondents themselves, but rather by the head of the household. Likewise, although there may be some links between the young adult's risk taking behavior and the parent's life expectancy (for example if the parent engages in risk and this affects the risk taking behavior of the young adult), this link is likely to be relatively weak.

The theoretical model described above predicts that *uncertainty* of future health reduces incentives to abstain from unsafe sex in the present. More specifically, *proposition 1a* in *section 3* above, suggests that the *spread* between potential future health outcomes matters for sexual behavior. We therefore estimated the variance in subjective life expectancy in each enumeration area[23] and use this as a proxy for uncertainty of future health.

Finally, *proposition 1c* suggests a link between expectations of future income and current sexual risk taking behavior. Since risky sexual behavior may affect expectations of future income (via the effect of health on earnings), we use mean household per capita income as a direct proxy for measuring expected income. Since it is more likely that older adults and parents rather than the young adults themselves are breadwinners in the household, we treat this variable as exogenous in the regressions.

In order to control for HIV knowledge, we use information on how often the young adult talks about HIV with parents or step parents, a dummy for whether the young adult knows someone with HIV or someone that has died of AIDS, and a set of dummies for different methods stated by the young adult as effective methods to prevent an HIV infection.[24] In addition we include a dummy for whether the young adult had secondary education or higher in 2003.[25] The different measures of HIV knowledge is included in order to capture both formal HIV knowledge and more deeply rooted relational knowledge of the consequences of an HIV infection.

---

[23] Primary sampling units corresponded to enumeration areas within the 1996 census and are defined in terms of subplace and placename. In calculating the means and variance of subjective life expectancy at a neighborhood level, we used all observations within a subplace (excluding the individual's own data) if the total number of observations were greater or equal to 20. Alternatively we estimated the means and variances at a placename level, but in such cases restricted the sample by population group.

[24] Abstaining from sex, only having one sex partner, limiting the number of sex partners and using condoms

[25] A dichotomous variable measuring whether the young adult only has primary education or lower, or at least secondary education.

4.3 Econometric specification

To study whether a significant impact of the economic and health conditions on the decision to engage in risky sexual behavior exists, we start by presenting the econometric relationships between our measure of risky sexual behavior and its determinants. We then discuss the econometric issues involved in the measurement of the dependent and independent variables and how we address them.

The main hypothesis is that the degree of riskiness of an individual's sexual behavior depends on the level and uncertainty of future prospects in terms of health and income. In estimating this relationship a number of important factors need to be considered. First, the sample consists of a relatively large share of individuals not yet sexually active. Merely estimating the relationship for the sexually active sample, or simply treating these individuals as "zero-risk" takers in the regression may lead to selection bias. First, since abstaining from sex may constitute a strategy to protect against HIV the non-sexually active sample may contain important information. However, estimating a Tobit where not sexually active young adults are treated as zeroes is problematic. First, the Tobit approach implies that we assume that the decision to not have sex is qualitatively the same as the decision related to the *degree* of risk taken. However, there may be important differences related to these decisions. Consider for example the effect of age: the older the young adult is, the more likely it is that he or she has had sex, simply due to biological reasons. However, concerning the *degree* of risk taken, age may have other implications such as experience or the desire to engage in more stable relationships. Second, sexually active individuals may differ from non-sexually active individuals with regards to unobserved characteristics such as risk aversion. In other words, treating the decision to abstain from sex as a corner solution may provide biased results due to a correlation of the error terms. Hence, studying the riskiness of sexual behavior needs to account for the presence of individuals who have no sexual experience at all and to correct for possible sample selection. In addition, individuals who are sexually active may choose to abstain from *risky* sexual behavior by choosing to use condoms consistently. As in the case of the choice to become sexually active, this decision *may* be distinct from the decision related to the degree of risk taken. This would for example be true for someone who does not use condoms, but engage in sex with only one partner consistently, while expecting their partner to be mutually monogamous. One way of estimating the model is to consider a double selection framework. This approach suggests the following system of equations:

Selection model 1: The decision to become sexually active or not

$$y_{s,i} \quad = \begin{cases} 1 & if \quad y_{s,i}^{*} = \boldsymbol{\alpha}'\boldsymbol{w_i} + \varepsilon_i > 0 \\ 0 & otherwise \end{cases} \qquad (11)$$

where for an individual i , y_s is a dichotomous variable taking the value 1 if the young adult is sexually active and 0 otherwise, and $\boldsymbol{w}$ is a vector of variables that determine the individual's decision to engage in sex or not; $\boldsymbol{\alpha}$ represents the vector of coefficients to be estimated, and ε represents the error term.

Selection model 2: The decision to use condoms consistently or not, given that the young adult is sexually active

$$y_{c,i}|y_{s,i} = 1 \quad = \begin{cases} 1 & if \quad y_{c,i}^{*} = \boldsymbol{\vartheta}'\boldsymbol{z_i} + \varepsilon_i > 0 \\ 0 & otherwise \end{cases} \qquad (12)$$

where for individual *i,* y_c is a dichotomous variable taking the value 1 if condoms were used consistently with the individual's last sex partner and 0 if condoms were used inconsistently and $\boldsymbol{z}$ is a vector of exogenous variables affecting the decision to use condoms consistently. Note that, since our main interest is to investigate sexual risk taking, and in order to avoid endogeneity in the selection equation, we exclude variables related to health and income in these equations.[26] The selectivity corrected outcome model for sexual behavior given inconsistent condom use is given by:

$$y_{sr,i}|\, y_{s,i}, y_{c,i} = 1 \quad = \boldsymbol{X_i}'\boldsymbol{\beta} + \sigma_1 \cdot \lambda_{1,i} + \sigma_2 \cdot \lambda_{2,i} + v_i \qquad (13)$$

where for an individual i y_{sr} is the sexual risk index, and λ are the estimated inverse Mill's ratios emanating from selection equation (11) and (12) respectively. The $\boldsymbol{\beta}$, σ_1 and σ_2 are the corresponding coefficients to be estimated. The model can be estimated even if the variables included in the different regressions are identical. However, if this is the case, the identification of the model relies on the nonlinearity of the inverse Mills ratios. Since this nonlinearity is often weak, the lack of exclusion restrictions is likely to cause a high correlation between the estimated inverse Mills ratio and the regressors in the outcome model (e.g. Wooldridge, 2002). This multicollinearity makes the estimated coefficients unreliable and may thus lead to erroneous conclusions about the significance levels of the estimated parameters.

In order to reduce multicollinearity we thus need to find a set of valid exclusion restrictions for each selection equation. For the first selection equation (the decision to become sexually active) two exclusion restrictions are available in the dataset: a dummy variable for whether or not the

[26] A preliminary analysis shows that the coefficients on these variables are insignificant. Results are available from the authors upon request.

young adult expected to have a child in three years time in 2002, and whether or not the young adult stated abstinence as an effective measure to avoid an HIV infection. However, finding valid exclusion restrictions that separates the decision related to condom use from the decision related to the sexual risk index is difficult from both theoretical and data availability perspectives. Research in psychology and sociology suggests that perceived ability to prevent an HIV infection, social support for condom use and discussing HIV prevention with parents constitute important determinants for condom use (e.g., Adih & Alexander, 1999; Diclimente *et al.*, 1992; Miller *et al.*, 1998; Meekers and Klein, 2002). However, these factors may also affect the degree of sexual decision for individuals that use condoms inconsistently. In addition to making estimation difficult, the lack of theoretically viable exclusion restrictions also questions the adequacy of modeling the condom decision as a selection issue. A preliminary analysis of the double selection model show that, due to the strong correlation between the error terms of the condom equation and the sex risk equation in a finite sample, the selectivity corrected standard errors are in some instances negative (see Greene, 1981 for a discussion of this problem). We take this as an additional argument for treating the condom and sex risk decision to be intrinsically the same,[27] and therefore consider a mildly different approach in terms of viewing the sexual behavior and condom decision as qualitatively the same.[28] Since it may be important to differentiate between the "zero-risk takers" that are not sexually active, from the "zero-risk takers" that are condom users, we maintain the selection model for the decision to become sexually active in the first step[29]. For the second step, we adjust the LISREL risk index for condom use in terms of multiplying the latent factor scores with 1 if condoms were used inconsistently, and 0 if condoms were used consistently. This produces a zero-inflated dependent variable. In order to adjust for the large stack of zeroes we therefore estimate the second step with a Tobit. This approach has previously been used by, for example, Czarnitzki & Kraft (2004) and Hallman *et al.* (2005) in other settings.

The selection model thus reduces to selection model 1 (equation 11), while the outcome equation now has the condom use adjusted LISREL index as the dependent variable:

$$y^*_{sr,i} \big| \, y_{s,i} = 1 \quad = \boldsymbol{X_i}'\boldsymbol{\beta} + u_i \tag{14}$$

[27] Results for the double selection model are available from the authors upon request. The results are qualitatively the same as the Tobit model with selectivity

[28] Results for the double selection model are available from the authors upon request. The results are qualitatively the same.

[29] Naturally, using condoms does not guarantee full protections against HIV since condoms may break or come off during sexual intercourse. However, since, to our knowledge, there are no estimates on the degree of risk associated with condoms we disregard from this effect here.

$$y_{sr,i} = max(0, y^*_{sr,i}) \qquad (15)$$

The log likelihood function for the Tobit model with selectivity is given by:

$$Log\ L = \sum_{y_s=0} log\Phi[-\alpha' w] + \sum_{y_s=1, y_{sc}=0} log\Phi_2[-\beta_1' \boldsymbol{x}_{i1}, \alpha' w, -\rho]$$
$$+ \sum_{y_s=1, y_{sc}>0} -1/2\,[log 2\pi + log\sigma + (u_i/\sigma)^2] \qquad (16)$$
$$+ log\Phi\left[(\alpha' w + \rho u_i/\sigma)/(1-\rho^2)^{\frac{1}{2}}\right]$$

where Φ_2 is the standard bivariate normal cumulative distribution function (CDF) and Φ is the univariate standard normal CDF. The ρ is the correlation between u and ε, and σ is the error variance (Greene, 2007, p. R17-69, E31-8).

The potential endogeneity issue is addressed using instrumental variables techniques. Due to the nonlinear characteristic of the outcome model, the instrumental variable approach involves two steps that are mildly different from the classical 2SLS procedure (e.g. Angrist and Pischke, 2009; Greene, 2007). In the first step, the endogenous variable is regressed on all exogenous variables in the outcome model in combination with a set of instruments for the selected sample:

$$\tilde{h}_{t+1,i} = \boldsymbol{X_i}'\boldsymbol{\beta_h} + \boldsymbol{V_i}'\boldsymbol{\kappa} + e_i \qquad (17)$$

where, for individual *i*, $\tilde{h}_{t+1}$ is the subjective life expectancy, $\boldsymbol{X}$ the vector of exogenous variables in equation (15), and $\boldsymbol{V}$ is a vector of instrumental variables corresponding to the endogenous variables. The $\boldsymbol{\beta_h}$ and $\boldsymbol{\kappa}$ are vectors of parameters to be estimated, and e is the error term. The predicted value from equation (17) is then used as an instrument in the outcome equation. Standard errors are corrected for the generated regressor problem (Greene, 2007, p. E30-26). The selection, outcome and instrumentation is carried out in one step with maximum likelihood in Limdep.

As mentioned above, our candidate instruments for the endogenous variable subjective life expectancy consists of the life expectancy of the young adult's parent and mean life expectancy at the sub-place level (net of the young adult's own life expectancy). Each of these instruments is a plausible instrumental variable whose validity depends on the degree to which it is correlated with subjective life expectancy only via its effect on life expectancy and not via any direct link to risky sexual behavior.

In order to test for over-identification, we perform an Amemiya-Lee-Newey test (Baum *et al.,* 2006). The test indicates that the specification shows no over-identification. As we fail to reject the null hypothesis of validity of mean of young adults' and parents' life expectancy, these instruments are used in our model specifications. In order to test exogeneity, we perform a Wald test.[30] The results indicate insufficient evidence to reject exogeneity. This result may be interpreted in three ways: As subjective life expectancy not being endogenous to sexual risk taking, as subjective life expectancy not measuring expected health, or as the explanatory variables not being adequate as instruments. From a theoretical perspective, life expectancy is without doubt endogenous to the level of sexual risk taking as long as HIV implies a premature death. Hence, some caution is needed when we interpret the results presented below.

Heteroscedasticity may constitute a problem. This is especially so when the outcome equation is a Tobit, since the presence of heteroscedasticity makes Tobit coefficients both inefficient and inconsistent. In order to correct for possible heteroscedasticity, we first estimate a heteroscedastic Tobit model without correcting for endogeneity (see *Table A4* in the appendix). However, since the estimation does not reveal any severe heteroscedasticity, the results presented in the next section are for a Tobit selectivity model with robust standard errors and instrumentation for the variable life expectancy.

Finally, due to the relatively low reliability of the sexual risk index estimated by factor analysis, we also estimate a selection model that focus on the decision related to condom use. More specifically, we estimate a model where the dependent variable is a measure of how often the young adult used condoms with his or her last sexual partner. As was argued above, this approach is only adequate if we consider with *whom* the individual had sex. In order to do this, we include information about the type of relationship the young adult had with the individual that he or she had sex with. In order to use all information available, we use the condom variable of ordinal scale available in the CAPS dataset. This implies that we need to estimate an ordered probit with selection. This model is given by:

$$\begin{aligned} y_{c,i}^{*} &= \boldsymbol{\vartheta_c}'\boldsymbol{z_i} + \zeta_i, \quad E[\zeta_i] = 0 \\ y_{c,i} &= 1 \quad if \quad y_{c,i} > \mu_0, \\ y_{c,i} &= 2 \quad if \quad y_{c,i} > \mu_1, \end{aligned} \tag{18}$$

[30] The intuition behind endogeneity is that the error terms of the structural and reduced form equations are correlated. This is because the added explanatory power of the instrument would be reflected in the reduced form equation (Wooldridge, 2002). However, it should be noted that even an invalid instrument which has its own explanatory variable could pass the test for endogeneity, implying that the test for endogeneity is not a test for the validity of the instrument per se.

$$y_{c,i} = 3 \quad if \quad y_{c,i} > \mu_2,$$
$$y_{c,i} = 4 \quad if \quad y_{c,i} > \mu_4$$

where $\boldsymbol{z_i}$ now also contains the explanatory variable relation to last sex partner (rel_type). As in the above, $y_{c,i}$ is only observed if $y_{s,i} = 1$. As in the Tobit model with selectivity, subjective life expectancy is instrumented with mean life expectancy and parent life expectancy. The log likelihood function for the model with sample selection is:

$$Log\ L = \sum_{y_s=0} log\Phi[-\alpha'\boldsymbol{z}] + \sum_{j,y_s=1} log\{\Phi_2[\mu_j - \boldsymbol{\vartheta_c}'\boldsymbol{z_i}, \alpha'w, \rho] - \Phi_2[\mu_{j-1} - \boldsymbol{\vartheta_c}'\boldsymbol{z_i}, \alpha'w, \rho]\} \quad (19)$$

where μ_j are the thresholds (Greene, 2007).

5. Empirical results

In this section we present the results of the empirical analysis. We start presenting the results of the CFA for the sexual risk index in section 5.1 and then proceed with the main results of the empirical analysis in section 5.2.

5.1 The sexual risk index

A CFA in LISREL 8.80 of the equation system in equation (10) produces a model with the relationships presented in *Figure 1* below.

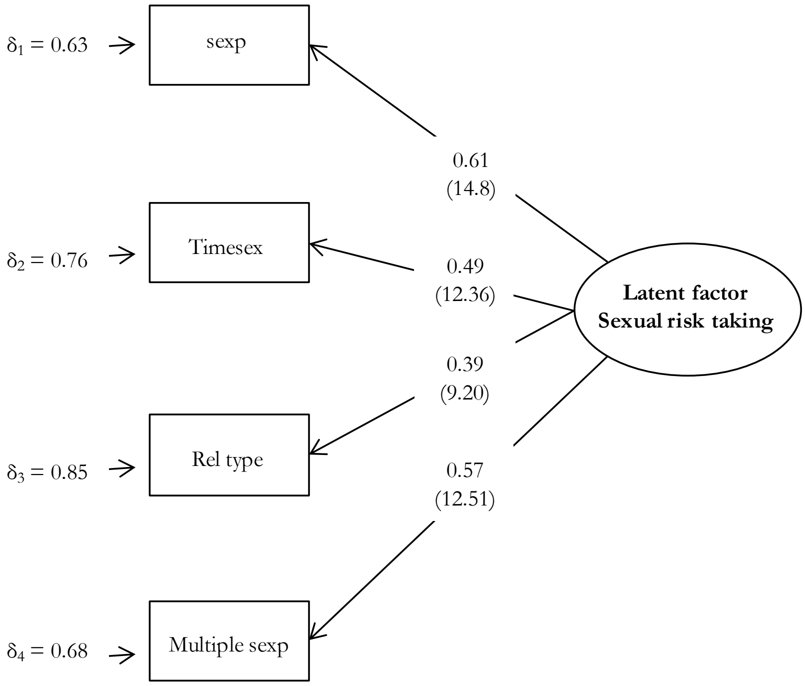

Figure 1: Path diagram for sexual risk taking

Observed variables represented by rectangular boxes and latent factors by elliptic boxes. Factor loadings are presented on each path from the latent factor to the observed variable with *t-values* in parenthesis. Fit statistics: Satorra-Bentler Scaled Chi-Square = 0.99 (P = 0.61), Root Mean Square Error of Approximation (RMSEA) = 0.0, Goodness of Fit Index (GFI) = 1.00 , Adjusted Goodness of Fit Index (AGFI) = 1.00, Parsimony Goodness of fit index (PGFI) = 0.20.

The path diagram in *Figure 1* describes the relationship between the set of observed variables and the hypothesized latent factor sexual risk taking. The *factor loading* for each variable describes how much of the variance of a variable that is explained by the latent factor. As can be seen in *Figure 1* all manifest variables are associated with positive loadings on the latent factor, thus suggesting that a higher degree of sexual risk taking is associated with more sexual partners, more casual partners, shorter time before having sex and the presence of simultaneous sex partners. However, as can also be seen in the figure, the estimated factor loadings are relatively small. This implies that the construct reliability and total extracted variance[31] are relatively low (0.59 and 0.27, respectively).[32] However, the overall model fit is satisfactory with a Santorra-Bentler scaled Chi-square equal to 0.99 (p = 0.61) and a root square mean error corresponding to 0.0. The parsimony goodness of fit index suggests that the model contains too many indicators. Since the model becomes saturated with less than four indicators, it is not possible to improve the

[31] $Construct\ reliability = \frac{\left(\sum_{i=1}^{p}\lambda_i\right)^2}{\left(\sum_{i=1}^{p}\lambda_i\right)^2+\sum_{i=1}^{p}e_i},\ extracted\ variance = \frac{\sum_{i=1}^{p}(\lambda_i)^2}{\sum_{i=1}^{p}(\lambda_i)^2+\sum_{i=1}^{p}e_i}$

[32] Ideally, construct reliability should be above 0.7 and extracted variance above 0.5

parsimony by dropping an indicator. Given the above results, it is clear that our measure of sexual risk taking is not 'perfect'. However, in comparison with alternative measures, this measure has been estimated and tested with suitable statistical methods and the individual weights of the indicator variables are in all cases highly significant. We therefore argue that the estimated sex risk index is a more adequate measure of sexual risk taking than using individual variables like, say age of first sex.[33]

5.2 Main results

Let us now turn to the empirical findings. As can be seen in *Table* 2, the results only give weak support for the propositions derived from the theoretical model in *section 3*. *Proposition 1b* suggests that expectations of good health in the future increases incentives to abstain from sexual risk taking in the present. However, as is depicted in *Table 2*, although the coefficient on life expectancy has the expected sign, it is not statistically significant. Indeed, this result persists even if we do not use instrumental variables. Hence, if subjective life expectancy is a valid measure of expected health we cannot reject the null hypothesis that expectations of the level of future health is unrelated to sexual risk taking in the present.

Proposition 1a suggests that a high degree of uncertainty of future health reduces the expected utility of safe sexual practices. The positive and significant coefficient on variance in life expectancy lends support to this hypothesis. The result suggests that individuals living in communities with a greater variability in life length scores significantly higher on the sexual risk index.

Finally, *Proposition 1c* suggests that expectations of future income may both increase and decrease incentives to engage in unsafe sex practices in the present. The empirical results suggest that a high household income is associated with a lower degree of sexual risk taking. Given the assumption that a young adult's expectations of own future income is a function of their parents' income level, this suggests that future income levels have a dampening effect on sexual risk taking.

For most of the covariates, the results are relatively unsurprising. In accordance with the UNAIDS (2009) report on the AIDS epidemic, women score significantly higher than men on the sexual risk index. Similarly, young adults with only primary education or lower, and individuals that do not discuss HIV with their parents are more exposed to risk, than young adults with secondary school education and young adults from families that talk about HIV.

[33] In a forthcoming paper, the differences between different measures of sexual risk are discussed further.

Table 2: *Tobit selectivity model: Determinants of the degree of sexual riskt aking by young adult*

	B	std.err	z	P[|Z|>z]
Direct effects in the regression				
Constant	0.761	0.684	1.111	0.266
Black	-0.377	0.379	-0.995	0.32
Colored	0.3	0.345	0.869	0.385
Female	**1.086**	0.193	5.629	0.000
Education	**-0.476**	0.225	-2.12	0.034
infHIV	**-0.199**	0.069	-2.901	0.004
HIVknow	0.046	0.177	0.259	0.795
LifeEXP	-0.007	0.046	-0.147	0.884
Variance lifeEXP	**0.026**	0.007	3.492	0.001
Income	**-0.07**	0.03	-2.356	0.019
HIVprotct_limit	**0.729**	0.263	2.771	0.006
HIVprotct_condom	-0.239	0.253	-0.944	0.345
Indirect effects in Lambda (mean are for all observations)				
Constant	**-8.084**	2.203	-3.67	0.000
Black	**0.515**	0.135	3.81	0.000
Colored	-0.023	0.063	-0.358	0.72
Female	-0.055	0.048	-1.136	0.256
infHIV	-0.024	0.026	-0.946	0.344
HIVknow	**0.159**	0.067	2.371	0.018
Age	**0.824**	0.237	3.482	0.001
Age^2	**-1.922**	0.607	-3.169	0.002
HIVprotct_abstain	**-0.413**	0.105	-3.946	0.000
ChildEXP_2002	**0.236**	0.119	1.99	0.047
Total effects of variables in both parts				
Black	0.138	0.402	0.345	0.73
Colored	0.276	0.351	0.79	0.43
Female	**1.031**	0.199	5.188	0.000
infHIV	**-0.224**	0.073	-3.048	0.002
HIVknow	0.205	0.189	1.081	0.28
SIGMA(1)	**2.373**	0.15	15.767	0.000
RHO(1,2)	**-0.473**	0.087	-5.436	0.000
N	1838			
Log L	-2619.331			
Finite sample AIC	2.877			

Knowing someone with HIV is not significantly associated with sexual risk taking behavior. However, it is interesting to note that the indirect effect is significant at the 5 percent level and *positive*. This result suggests that knowing someone with HIV/AIDS may be associated with more sexual risk taking behavior.[34] Naturally, if the young adult lives in an environment where sexual risk taking is fairly common, we may expect that he or she knows someone with HIV/AIDS and that the social norms related to sexual behavior affects the young adult's own sexual habits. However, it is surprising that young adults that consider limiting the number of sex partners (HIVprotect_limit) to be an effective method to protect against HIV/AIDS have higher scores on the sexual risk index than young adults not supporting this view. Finally, it is relatively

[34] We have naturally not included individuals who are HIV positive themselves. Unfortunately we cannot control for whether the person is the young adult's sex partner since this information is not available in the dataset.

unsurprising that older young adults are more likely to be sexually active and that this relationship is non-linear.[35]

The ordered probit for condom use alone (corrected for selection) shows similar results as the results of the Tobit selectivity model. The results differ with respect to one interesting aspect; as can be seen in *Table 3*, life expectancy is *significantly* associated with the frequency of condom use with the young adult's most recent sexual partner. The results suggest that the higher the life expectancy is, the more likely it is that condoms are used to a relatively higher extent. Variance of life expectancy is, as in the above associated with a lower degree of protective measures in terms of condom use and a higher income is significantly associated with a higher rate of condom use. Some caution is needed when we interpret these results as the variance explained in the condom use regression is very low (*MacFadden's Pseudo R-square* is 0.001).

Table 3: *Ordered Probit: Determinants of Condom use*

	B	std.err	z	P[|Z|>z]
Index function for probability				
Constant	-0.052	0.525	-0.099	0.922
Black	**-0.31**	0.174	-1.786	0.074
Colored	0.102	0.16	0.64	0.522
Female	**0.48**	0.065	7.422	0.000
Education	**-0.283**	0.092	-3.079	0.002
infHIV	**-0.094**	0.028	-3.368	0.001
HIVknow	-0.002	0.074	-0.023	0.982
LifeEXP	**-0.045**	0.018	-2.483	0.013
Variance lifeEXP	**0.01**	0.003	3.361	0.001
Age	**0.062**	0.021	2.984	0.003
Income	**-0.057**	0.013	-4.213	0.000
HIVprotct_condom	**-0.291**	0.098	-2.968	0.003
Rel_type	**-0.134**	0.048	-2.8	0.005
Threshold parameters for index				
Mu(1)	**0.184**	0.019	9.811	0,000
Mu(2)	**0.932**	0.042	22.256	0,000
Selection equation				
Constant	**-14.112**	1.997	-7.068	0.000
Black	**0.962**	0.121	7.963	0.000
Colored	0.004	0.107	0.038	0.97
Female	**-0.145**	0.067	-2.15	0.032
infHIV	-0.037	0.03	-1.252	0.211
HIVknow	**0.287**	0.084	3.436	0.001
Age	**1.452**	0.231	6.278	0.000
Age^2	**-3.397**	0.663	-5.12	0.000
HIVprotct_abstain	**-0.756**	0.068	-11.119	0.000
ChildEXP_2002	**0.422**	0.155	2.712	0.007
Cor[u(probit),e(ordered probit)]				
Rho(u,e)|	**-0.253**	0.151	-1.675	0.094
N	2080	Psuedo Rsqr(Mcfadden)		0.001
Log L	-2507.335	Chi squared (prob)		7.401 (0.007)
Finite Sample AIC	2.436			

[35] Since age is already included in one of the measures included in the sexual risk index, we do not include age as a covariate for the sexual risk index.

5. Conclusions and Discussion

The main focus of this paper is to empirically test the link between expectations of future health and income and sexual risk taking. The main findings indicate that uncertainty of future health is a significant determinant of current patterns of sexual risk taking. Similarly, risky sexual behavior tends to be associated with low current income. The association between household income and sexual risk taking suggests that an expectation of a prosperous future may well reduce the incentives to engage in sexual risk taking in the present. The relatively weak impact of future income could be explained by its multidirectional impact. A higher future income may increase the expected utility of surviving to old age and may also reduce the need for saving. Since the CAPS dataset does not contain information about savings or monetary costs related to sexual behavior, we cannot disentangle the two effects from each other.

The results for expected health are mixed. Although the empirical analysis lends support for a relationship between low life length expectations and a high degree of sexual risk taking in terms of condom use, the low degree of variance explained in the regression indicates that the evidence for a link between subjective life expectancy and condom use is relatively weak.

The weak relationship between life expectancy and sexual risk taking raises interesting questions about human rationality and the validity of life expectancy as a measure of expected health. If subjective life expectancy is indeed an adequate measure of expected health, the empirical analysis only lends limited support for the hypothesis derived from the theoretical model in *section 3*. However, as described in the discussion of the instrumental variable tests, the exogeneity of life expectancy cannot be rejected. Hence, since future health is likely to be a function of sexual risk taking, the empirical results implies that life expectancy may not correctly measure subjective expected health. In addition, it should be noted that the young adults' subjective expectations related to health may be biased. Indeed, research in psychology suggests that individuals tend to overestimate the per-coital risk of an HIV infection, but to underestimate their *personal* risk of acquiring the virus. This phenomenon has been labeled *optimistic bias* (e.g., van der Pligt *et al.*, 1993). Hence, if individuals fail to fully acknowledge their personal risk, subjective life expectancy may not be endogenous to the degree of sexual risk taking. However, if it is the case that individuals rationalize their behavior, in terms of (potentially falsely) convincing themselves that their protective measures are adequate, then expected health should not be expected to affect incentives related to sexual risk taking. If this is the case, traditional economic models of expected utility needs revision in order to catch the complex nature of risk perception and risk taking behavior (the mechanisms and consequences of optimistic bias and rationalizing behavior are discussed in Mannberg, 2010).

The results of the empirical analysis also provide interesting suggestions regarding the other determinants of sexual risk taking. The results in general suggest that females are exposed to HIV risk to a higher degree than men. Even though women are less likely to be sexually active, they are also less likely to use condoms frequently. This result may suggest that women dislike condoms more than men or that they are more prone to risk taking in general. A perhaps more likely explanation relates to social taboos concerning women's use of condoms and a general lack of sexual bargaining power for women. It is also interesting to note that information and knowledge of HIV in terms of discussing HIV with relatives and being educated is associated with a lower degree of sexual risk taking. Although previous research (Macintyre *et al.* 2001) has found that knowing someone with HIV reduces the degree of sexual risk taking, our analysis suggests that a personal experience of HIV actually *increases* the degree of sexual risk taking.

The major finding of the study is that poor economic circumstances and, more importantly, lack of bright economic and health prospects, explain the anomalies regarding high levels of sexual risk taking in a risky sexual environment. These findings may also hold important information from policy perspective. Since, as per the finding of this study, sexual behavior is closely tied to economic circumstances, the effectiveness of preventive mechanisms in the HIV epidemic requires addressing poverty and social insurance issues. In addition, the importance of HIV information and inter-gender imbalances in bargaining power call for more research on these issues.

While the study makes an important methodological step in combining sexual risk measures into an index, future research that analyzes the relative information contents of alternative sexual risk measures would form the basis for studies on sexual behavior. Further, studies about the impacts of reducing the taboo related to talking about HIV, and the relatively low degree of condom use of women would also illuminate our understanding of preventive measures.

Appendix A1

In this section, the proof of proposition 1 is given. First, let us totally differentiate the first order conditions (5) and (7) with respect to the parameters, $\gamma, \tilde{h}_2, Y_2$ and φ. This gives us the system:

$$\begin{pmatrix} U_{x_{us}x_{us}} & U_{x_{us}x_s} & U_{x_{us}S} \\ U_{x_s x_{us}} & U_{x_s x_s} & U_{x_s S} \\ U_{S x_{us}} & U_{S x_s} & U_{SS} \end{pmatrix} \cdot \begin{pmatrix} dx_{us} \\ dx_s \\ dS \end{pmatrix}$$
$$= \begin{pmatrix} -U_{x_{us}\gamma} & -U_{x_{us}\tilde{h}_2} & -U_{x_{us}\delta} & -U_{x_{us}\theta} & -U_{x_{us}\varphi} \\ -U_{x_s\gamma} & -U_{x_s\tilde{h}_2} & -U_{x_s\delta} & -U_{x_s\theta} & -U_{x_s\varphi} \\ -U_{x_s\gamma} & -U_{x_s\tilde{h}_2} & -U_{x_s\delta} & -U_{x_s\theta} & -U_{x_s\varphi} \end{pmatrix} \cdot \begin{pmatrix} d\gamma \\ d\tilde{h}_2 \\ d\varphi \end{pmatrix} \quad \text{(A1.1)}$$

where the individual derivatives are given in equations A1.2-A1.16, below.

$$U_{x_{us}x_s} = pq \cdot u_{c_1c_1} < 0 \quad \text{(A1.2)}$$

$$U_{x_{us}S} = q \cdot u_{c_1c_1} + \beta \cdot \ln(\varphi) \cdot \varphi^{x_{us}}\left[\rho \cdot u_{c_2}(c_2, h_{high}) + (1-\rho) \cdot u_{c_2}(c_2, h_{low})\right] < 0 \quad \text{(A1.3)}$$

$$U_{x_s S} = p \cdot u_{c_1c_1} < 0 \quad \text{(A1.4)}$$

$$U_{x_{us}\gamma} = \rho \cdot (1-\rho) \cdot \beta \cdot \ln(\varphi) \cdot \varphi^{x_{us}} \cdot \left[u_h(c_2, h_{high}) - u_h(c_2, h_{low})\right] > 0 \quad \text{(A1.5)}$$

Note that, by the second order condition for a maximum, $U_{x_{us}x_{us}}, U_{x_s x_s}, U_{SS} < 0$.

$$U_{x_s\gamma} = 0 \quad \text{(A1.6)}$$

$$U_{S\gamma} = \rho \cdot (1-\rho) \cdot \beta \cdot \varphi^{x_{us}} \cdot \left[u_{c_2h}(c_2, h_{high}) - u_{c_2h}(c_2, h_{low})\right] > 0 \quad \text{(A1.7)}$$

$$U_{x_{us}\tilde{h}_2} = \beta \cdot \ln(\varphi) \cdot \varphi^{x_{us}} \cdot \left[\rho \cdot u_h(c_2, h_{high}) + (1-\rho) \cdot u_h(c_2, h_{low})\right] < 0 \quad \text{(A1.8)}$$

$$U_{x_s\tilde{h}_2} = 0 \quad (A1.9)$$

$$U_{S\tilde{h}_2} = \beta \cdot \varphi^{x_{us}} \cdot \left[\rho \cdot u_{c_2h}\left(c_2, h_{high}\right) + (1-\rho) \cdot u_{c_2h}(c_2, h_{low})\right] > 0 \quad (A1.10)$$

$$U_{x_{us}\rho} = \beta \cdot \ln(\varphi) \cdot \varphi^{x_{us}} \cdot \left[u\left(c_2, h_{high}\right) - u(c_2, h_{low})\right] < 0 \quad (A1.11)$$

$$U_{x_s\rho} = 0 \quad (A1.12)$$

$$U_{S\rho} = \beta \cdot \varphi^{x_{us}} \cdot \left[u_{c_2}\left(c_2, h_{high}\right) - u_{c_2}(c_2, h_{low})\right] > 0 \quad (A1.13)$$ [36]

$$U_{x_{us}Y_2} = \beta \cdot \ln(\varphi) \cdot \varphi^{x_{us}} \cdot \left[\rho \cdot u_{c_2}\left(c_2, h_{high}\right) + (1-\rho) \cdot u_{c_2}(c_2, h_{low})\right] < 0 \quad (A1.14)$$

$$U_{x_sY_2} = 0 \quad (A1.15)$$

$$U_{SY_2} = \beta\varphi^{x_{us}}\left[\rho \cdot u_{c_2c_2}\left(c_2, h_{high}\right) + (1-\rho) \cdot u_{c_2c_2}(c_2, h_{low})\right] < 0 \quad (A1.16)$$

$$U_{x_{us}\varphi} = \beta \cdot \varphi^{x_{us}-1} \cdot \left[\rho \cdot u\left(c_2, h_{high}\right) + (1-\rho) \cdot u_c(c_2, h_{low})\right] \cdot [1 + x_{us} \cdot \ln(\varphi)] \quad (A1.17)$$

$$U_{x_s\varphi} = 0 \quad (A1.18)$$

$$U_{S\varphi} = x_{us} \cdot \beta \cdot \varphi^{x_{us}-1} \cdot \left[\rho \cdot u_{c_2}\left(c_2, h_{high}\right) + (1-\rho) \cdot u_{c_2}(c_2, h_{low})\right] > 0 \quad (A1.19)$$

[36] Since $u_{c_2h_2} > 0$

Further, let us define:

$$\left[U_{x_s x_s} \cdot U_{SS} - \left(U_{x_s S}\right)^2\right] = D_{x_{us}} > 0 \quad \text{(A1.20)}$$

[37]

$$\left[U_{x_{us} x_s} U_{x_s S} - U_{x_s x_s} U_{x_{us} S}\right] = D_S$$
$$= -\beta \cdot \ln(\varphi) \cdot \varphi^{x_{us}} \cdot \frac{\partial E[u(c_2, h_2)]}{\partial S} \cdot \left(q^2 u_{cc} + \alpha^2 \cdot g_{x_s x_s}\right) \quad \text{(A1.21)}$$
$$< 0$$

The derivatives in *proposition 1a-1e* are given by:

$$\frac{\partial x_{us}}{\partial \gamma} = \frac{-U_{x_{us}\gamma} \cdot D_{x_{us}} - U_{S\gamma} \cdot D_S}{D} > 0 \quad \text{(A1.22)}$$

$$\frac{\partial x_{us}}{\partial \tilde{h}_2} = \frac{-U_{x_{us}\tilde{h}_2} \cdot D_{x_{us}} - U_{S\tilde{h}_2} \cdot D_S}{D} < 0 \quad \text{(A1.23)}$$

$$\frac{\partial x_{us}}{\partial \rho} = \frac{-U_{x_{us}\rho} \cdot D_{x_{us}} - U_{S\rho} \cdot D_S}{D} < 0 \quad \text{(A1.24)}$$

$$\frac{\partial x_{us}}{\partial Y_2} = \frac{-U_{x_{us}Y_2} \cdot D_{x_{us}} - U_{SY_2} \cdot D_S}{D} \lessgtr 0 \quad \text{(A1.25)}$$

$$\frac{\partial x_{us}}{\partial \varphi} = \frac{-U_{x_{us}\varphi} \cdot D_{x_{us}} - U_{S\varphi} \cdot D_S}{D} \lessgtr 0 \quad \text{(A1.26)}$$

The signs of equation (A1.22) to equation (A1.26) follow from equation (A1.2) – (A1.21)

[37] By the second order condition for a maximum

Appendix A2

Variables related to sexual risk taking	
rel_type	Type of relationship to last sex partner (1; spouse/married, 2; Someone loved but not married to, 3; Someone known well but not loved, 4; Someone not known well)
timesex	Time before sexual intercourse with last sex partner (1; More than a year, 2; Less than a year but more than 1 month, 3; less than a month but more than 2 days, 4; less than 2 days)
multisex	Whether or not young adult had other sex partners during last sexual relationship: (1; definitely no, 2; unsure, 3; definitely yes)
sexp12	Number of sex partners last 12 months
Sexp_life	Number of sex partners in life
sexp_rel_age	Number of sex partners relative to age group (sexp_life/average sexp_life in age group *j*), categorical variable where each group represent a quartile of the sample
Condom use	Ordinal variable for condom use with most recent sex partner (1; always, 2; usually, 3; sometimes, 4; rarely)
Condom_dic	Dichotomous variable for condom use with most recent sex partner (0; always, 1; less than always)
Independent variables	
lifeEXP	Life expectancy of young adult (30, 40, 50, 60, 70, 80)
Variance in lifeEXP	Variance in subjective life expectancy at sub-place level
Mean lifeEXP	Mean of subjective life expectancy at sub-place level
Parent lifeEXP	Parent subjective life expectancy
Age	Age in wave 1 (2002)
Age^2	Age in wave 1 squared
Education	Educational level (1; primary, 2; secondary or higher)
HIVknow	Dummy variable for relationship to an individual infected by or died of HIV/AIDS
pcinc	Household per capita income
infHIV	How often young adult discuss HIV with parent or step parent (1=never, 2=rarely, 3= sometimes, 4=often)
childEXP	Dummy variable for whether the young adult expected to have a child in three years time in 2002 (wave 3 took place in 2005)
HIVprotct_abstain	Dummy variable for whether young adult mentioned abstaining from sex as a method for preventing an HIV infection
HIVprotct_limit	Dummy variable for whether young adult mentioned limiting number of sex partners as a method for preventing an HIV infection
HIVprotct_condom	Dummy variable for whether young adult mentioned using condoms as a method for preventing an HIV infection

Appendix A3

The description in this section is to a large extent based on Jöreskog (2005), and Jöreskog and Moustaki (2001) (see also: e.g. Jöreskog, 1994). The CFA for ordinal variables described in section 4.2.1 was conducted as follows. In our dataset, we have extracted a set of manifest variables related to sexual behavior. Let us call these variables y_1, y_2, y_3 and y_4. Our hypothesis is that these variables are related to a latent variable; ξ, in terms of sexual risk taking. The latent variable is assumed to account for the dependencies between the manifest variables, such that if the latent variable was to be held fixed, the manifest variables would be independent. The basic model that we want to estimate can be represented by a *path diagram*, as depicted in *Figure A1*:

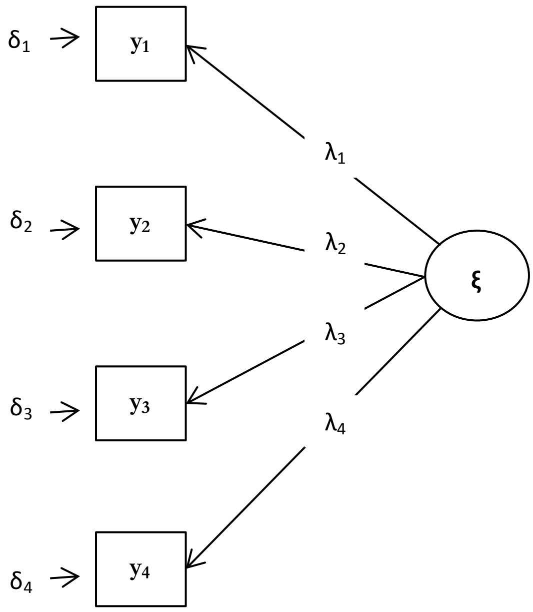

Figure A1: Path diagram

Observed variables represented by rectangular boxes and latent factors by elliptic boxes.

Classic factor analysis is based on the assumption that both manifest and latent variables are continuous and normally distributed with zero mean and unit variance, and suggests the following model for *k* latent factors:

$$E[y_i|\xi_1, \xi_2, \dots, \xi_k] = \lambda_{i1} \cdot \xi_1 + \lambda_{i2} \cdot \xi_2 + \dots + \lambda_{ik} \cdot \xi_k \quad \text{(A3.1)}$$

$$E\left[y_i y_j|\xi_1, \xi_2, \dots, \xi_k\right] = 0 \quad i \neq j \quad \text{(A3.2)}$$

Assuming independence between the latent factors, the above model implies that the correlations between the observed variables are given by:

$$\rho_{ij} = \sum_{l=1}^{k} \lambda_{il} \cdot \lambda_{jl} \quad \text{(A3.3)}$$

However, in our case the observed variables are of ordinal scale. Since ordinal variables do not have metric properties, the correlations and covariance of ordinal variables lacks meaning. In contrast to classical factor analysis, ordinal factor analysis is based on the *polychoric* correlation between the observed variables. This correlation is obtained by assuming that each observed ordinal variable has an underlying normally distributed continuous variable y_i^* with range $-\infty$ to $+\infty$. The underlying variable y_i^* assigns a metric to the ordinal variable and can be used for CFA. For an ordinal variable with *m* categories, the relationship between y_i and y_i^* is

$$y_i = i \quad \Rightarrow \quad \tau_{i-1} < y_i^* < \tau_i, \qquad i = 1,2, \ldots, m \tag{A3.4}$$

where

$$-\infty = \tau_0 < \tau_1 < \tau_2 < \cdots < \tau_{m-1} < \tau_m = +\infty \tag{A3.5}$$

are the thresholds. The underlying variables are assumed to have a standard normal distribution $\phi(u)$ and a distribution function $\Phi(u)$.[38] The probability of an answer in category i is:

$$\pi_i = Pr[y = i] = Pr[\tau_{i-1} < y^* < \tau_i\ ] = \int_{\tau_{i-1}}^{\tau_i} \phi(u)du = \Phi(\tau_i) - \Phi(\tau_{i-1}) \tag{A3.6}$$

This implies that $\tau_i = \Phi^{-1}(\pi_1 + \pi_2 + \cdots + \pi_i)$, where Φ^{-1} is the inverse of the standard normal distribution function, and the quantity $(\pi_1 + \pi_2 + \cdots + \pi_i)$ is the probability of an answer in category i or lower. π_i can be consistently estimated by the percentage of answers in category i,

$$\hat{\tau}_i = \Phi^{-1}(p_1 + p_2 + \cdots + p_i) \quad i = 1, 2, .., m - 1 \tag{A3.7}$$

where the quantity $(p_1 + p_2 + \cdots + p_i)$ is the proportion of cases in the sample with answers in category i or less. The polychoric correlation is estimated by assuming that the underlying variables are standard bivariate normal with a correlation ρ. The polychoric correlation is thus the correlation of the underlying variables. For two ordinal variables, the polychoric correlation can be estimated by maximizing the log likelihood of the multinominal distribution,

[38] Although the univariate marginal normality of the underlying variables is not testable, the bivariate normality of two underlying variables can be tested in LISREL.

$$\ln L = \sum_{i=1}^{m_1}\sum_{j=1}^{m_2} n_{ij} \cdot \log\left(\pi_{ij}(\boldsymbol{\theta})\right) \tag{A3.8}$$

where n_{ij} is the number of cases in category *i* on variable 1 and in category *j* on variable 2, and where

$$\pi_{ij}(\boldsymbol{\theta}) = Pr[y_1 = i, y_2 = j] = \int_{\tau_{i-1}^{(1)}}^{\tau_i^{(1)}} \int_{\tau_{j-1}^{(2)}}^{\tau_j^{(2)}} \phi_2(u, v) du dv \tag{A3.9}$$

Finally, $\phi_2(u, v) = \frac{1}{2\pi\sqrt{(1-\rho^2)}} \cdot e^{-\frac{1}{2(1-\rho^2)} \cdot (u^2 - 2\rho uv + v^2)}$ is the standard bivariate normal density with correlation ρ and $\boldsymbol{\theta} = \left(\tau_1^{(1)}\tau_2^{(1)}, \dots, \tau_{m_1-1}^{(1)}, \tau_1^{(2)}\tau_2^{(2)}, \dots, \tau_{m_2-1}^{(2)}, \rho\right)$. Maximizing $\ln L$ is the same as minimizing the function,

$$F(\boldsymbol{\theta}) = \sum_{i=1}^{m_1}\sum_{j=1}^{m_2} p_{ij} \cdot \left(\ln(p_{ij}) - \ln\left(\pi_{ij}(\boldsymbol{\theta})\right)\right) \tag{A3.10}$$

where p_{ij} are the sample proportions.

The PRELIS program in LISREL estimates the threshold values from the univariate marginal distributions, $\hat{\tau}_i = \Phi^{-1}(p_1 + p_2 + \cdots + p_i)$ and then estimates the polychoric correlations and their asymptotic covariance matrix from the bivariate marginal distributions by minimizing $F(\boldsymbol{\theta})$. The hypothesis that the observed variables are indicators of a one-dimensional variable is tested by estimating the model with maximum likelihood with robust standard errors and chi-square corrected for non-normality using the asymptotic covariance matrix of the polychoric correlations (Jöreskog, 2005).

If the model fit is adequate, the estimation procedure in LISREL provides us with consistent estimates of the factor loading associated with each variable in combination with each variable's unique variation.

Appendix A4

Table A4 *Heteroscedasticity corrected Tobit without selection*

	B	std.err	z	P[|Z|>z]
Constant	-3.273	1.227	-2.667	0.008
Black	1.023	1.004	1.019	0.308
Colored	0.844	0.975	0.865	0.387
Female	1.563	0.317	4.926	0.000
Education	0.24	0.282	0.853	0.394
infHIV	-0.431	0.097	-4.457	0.000
HIVknow	0.428	0.216	1.986	0.047
LifeEXP	-0.001	0.05	-0.022	0.982
Variance lifeEXP	0.016	0.009	1.717	0.086
Income	-0.534	0.172	-3.1	0.002
HIVprotct_limit	0.917	0.315	2.91	0.004
HIVprotct_condom	1.209	0.426	2.836	0.005
Heteroscedasticity term				
Black	-0.016	0.334	-0.049	0.961
Colored	-0.155	0.313	-0.496	0.62
Female	-0.328	0.142	-2.308	0.021
Education	-0.006	0.172	-0.033	0.974
infHIV	0.053	0.049	1.081	0.28
HIVknow	0.015	0.131	0.117	0.907
LifeEXP	0.011	0.03	0.377	0.706
Variance lifeEXP	0.666D-05	0.005	0.001	0.999
Income	0.055	0.042	1.332	0.183
HIVprotct_limit	0.009	0.158	0.059	0.953
HIVprotct_condom	-0.064	0.188	-0.341	0.733
Disturbance standard deviation				
Sigma	3.251	1.55	2.097	0.036

6. References

Abdool Karim, Q., Abdool Karim, S. S., Soldan, K., & Zondi, M., (1995), Reducing the risk of HIV infection among South African Sex workers: socio-Economic and gender barriers, *American Journal of Public Health 85(11),* 1521-1525

Adih, W. K. & Alexander, C. S., (1999), Determinants of condom use to prevent HIV infection among youth in Ghana, *Journal of Adolescent Health 24(1)*, 63-72

Anderson, K. G. & Beutel, A. M., (2007), HIV/AIDS Prevention Knowledge among Youth in Cape Town, South Africa, *Journal of Social Sciences 3(3),* 143-151

Anderson, K. G., Beutel, A. M. & Maughan-Brown, B., (2007), HIV Risk Perceptions and First Sexual Intercourse among Youth in Cape Town, South Africa, *International Family Planning Perspectives 33(3),* 98-105

Anderson, T. W. & Rubin, H., (1956), Statistical inference in factor analysis, in *Proceedings of the Third Berkeley Symposium V,* University of California Press

Angrist J. D, & Pischke, J. S. (2009), *Mostly Harmless Econometrics - An empiricist's companion*, Princeton University Press

Baum, C. F., M. E. Schaffer, S. Stillman, and V. Wiggins (2006). 'overid: Stata module to calculate tests of overidentifying restrictions after ivreg,ivreg2, ivprobit, reg3', *http://ideas.repec.org/c/boc/bocode/s396802.html.*

Benz, S., (2005), The Killing of the Fittest – A Quantitative Analysis of HIV/AIDS and Conflict, Mimeo, University of Konstanz

Bloom, S. S., Banda, C., Songolo, G., Mulendema, S., Cunningham, A. E. and Boema, T. J., (2000), Looking for Change in Response to the AIDS Epidemic: Trends in AIDS Knowledge and Sexual Behavior in Zambia, 1990 Through 1998, *Journal of Acquired Immune Deficiency Syndromes 25,* 77-85

Buvé, A., Carael, M., Hayes, R. J, Auvert, B., Ferry, B., Robinson, N. J., Anagonou, S., Kanhonou, L., Laourou, M, Abega, S., Akam, E., Zekeng, L., Chege, J., Kahindo, M., Rutenberg, N., Kaona, F., Musonda, T., Sukwa, T., Morison, L., Weiss, H. A., Laga, M., For the study group on heterogeneity of HIV Epidemic in African Cities, (2001), The multicentre study on factors determining the differential spread of HIV in four African cities: summary and conclusion", *Journal of Acquired Immune Deficiency Syndromes 15,* 127-131

Caldwell, J. C., (1999) Reasons for limited sexual behavioural change in the sub-Saharan AIDS epidemic, and possible future intervention strategies, Health Transition Centre, *Resistances to Behavioural change to reduce HIV/AIDS Infection*, The Australian National University, Canberra

Caldwell, J. C., (2000), Rethinking the African AIDS Epidemic, *Population and Development Review 26(1),* 117-135

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),* 273-281

Connolly, C., Colvin, M., Shishana, O. & Stoker, D., (2004), Epidemiology of HIV in South Africa – results of a national, community-based survey, *South African Medical Journal 94(9)*, 776-781

Czarnitzki, D. & Kraft, K., (2004), An empirical test of the asymmetric models on innovative activity: who invests more into R&D, the incumbent or the challenger?, *Journal of Economic Behavior & Organization 54*, 153-173

D'Hombres, B., Rocco, L., Schurcke, M., & Mckeel, M., (2010), Does Social Capital Determine Health? Evidence from Eight Transition Countries, *Health Economics 19*, 56–74

Diclimente, R. J., Durbin, M., Siegel, D., Krasnovsky, F., Lazarus, N. & Comacho, T., (1992), Determinants of Condom Use Among Junior High School Students in a Minority, Inner-City School District, *Pedriatics 89(2)*, 197-202

Dinkelman, T., Lam, D. and Leibbrandt, M., (2007), Household and community income, economic shocks and risky sexual behaviour of young adults: evidence from the Cape Area Panel Study 2002 and 2005, *Journal of Acquired Immune Deficiency Syndrome 21(s7)*, S49-S56

Dow, W. H., Philipson, T. J., Sala-i-Martin, X., (1999), Longevity Complementarities under Competing Risks, *American Economic Review*, *89(5),* 1358-1371

Gertler, P., Shah, M., and Bertozzi, S. M., (2005) Risky business: the Market for Unprotected Commercial Sex, *Journal of Political Economy 113(3),* 518-550

Greene, W., (1981), Sample Selection Bias as a Specification Error: A Comment, *Econometrica 49(3)*

Greene, W., (2007), *Limdep Version 9.0, Econometric Modeling Guide*, Econometric Software Inc. New York.

Hallman, K., Quisumbing, A., Ruel, M., & de la Brière, B., (2005), Mothers' Work and Child Care: Findings from the Urban Slums of Guatemala City, *Economic Development and Cultural Change 53,* 855-885

Hearst, N. & Chen, S., (2004), Condom Promotion for AIDS Prevention in the Developing World: Is It Working?, *Studies in Family Planning 35(1),* 39-47

Jöreskog, K. G., (2000), Latent variable scores and their uses, Scientific Software International, INC. (SSI)

Jöreskog, K. G., (2005), Structural Equation Modeling with Ordinal Variables using LISREL, Scientific Software International, INC. (SSI)

Jöreskog, K. G. & Moustaki, I., (1994), On the estimation of polychoric correlations and their asymptotic covariance matrix, *Psychometrica 59*, 381-389

Jöreskog, K. G. & Moustaki, I., (2001), Factor Analysis of Ordinal Variables: A Comparison of Three Approaches, *Multivariate Behavioral Research, 36(3)*, 347-387

Jöreskog, K. G. & Sörbom, D., (1979), *Advances in Factor Analysis and Structural Equation Models*, Abt Books, Cambridge, Massachusetts

Jöreskog, K. G., Sörbom, Dag, & Wallentin, F. Y., (2006), Latent Variable Scores and Observational Residuals, Scientific Software International, INC. (SSI)

Kaplan, D. (2000), *Structural Equation Modeling - foundations and Extensions*, in Advanced Quantitative Techniques in the Social Sciences Series 10, Sage Publications, Thousand Oaks

Kremer, M., (1996), Integrating Behavioral Choice into Epidemiological Models of AIDS, *Quarterly Journal of Economics 111(2)*, 549-573

Lagarde, E., Carael, M., Glynn, J. R., Kanhonou, L., Abega, S-C., Kahindo, M., Musonda, R., Auvert, B. & Buvé, A., (2001), Educational level is associated with condom use within non-spousal partnerships in four cities of sub-Saharan Africa *Journal of Acquired Immune Deficiency Syndrome 15,* 1399-1408

Lam, D., Ardington, C., Branson, N., Case, A., Leibbrandt, M., Menendez, A., Seekings, J. & 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 2008

MacPhail, C. and Campbell, C., (2001), 'I think condoms are good but, aai, I hate those things': condom use among adolescents and young people in a Southern African township", *Social Science & Medicine 52*, 1613-1627

Macintyre, K., Brown, L. & Sosler, S., (2001), 'It's Not What You Know, But Who You Knew': Examining the Relationship Between Behavior Change and AIDS Mortality in Africa", *AIDS Education and Prevention 13(2)*, 160-174

Mannberg, A., (2010), Risky Sex in a Risky World – Sexual behavior in a HIV/AIDS environment, *Scandinavian Journal of Economics* (forthcoming)

Meekers, D. & Klein, M., (2002), Determinants of Condom Use among Young People in Urban Cameroon, *Studies in Family Planning 33(4),* 335-346

Miller, K. S., Levin, M. L., Whitaker, D. J., & Xu, X., (1998), Patterns of condom use among adolescents: the impact of mother-adolescent communication, *American Journal of Public Health 88(10)*, 1542-1544

Mwaluko, G., Urassa, M., Isingo, R., Zaba, B. & Boerma, J. T., (2003), Trends in HIV and sexual behaviour in a longitudinal study in a rural population in Tanzania, 1994-2000, *Journal of Acquired Immune Deficiency Syndrome 17*, 2645-2651

Oster, E., (2005), Sexually Transmitted Infections, Sexual Behavior, and the HIV/AIDS epidemic, *Quarterly Journal of Economics 120(2)*, 467-515

Oster, E., (2007a), HIV and Sexual Behavior Change: Why not Africa? Preliminary Draft, University of Chicago and NBER

Oster, E., (2007b), Routes of Infection: Exports and HIV Incidence in Sub-Saharan Africa University of Chicago and NBER

Pettifor, A. E., Beksinska, M. E. & Rees, H. V., (2000), High Knowledge and High Risk Behaviour: A Profile of Hotel-Based Sex Workers in Inner-City Johannesburg, *African Journal of Reproductive Health 4(2),* 35-43

Pettifor A. E., Rees H. V., Steffenson, A., Hlongwa-Madikizela, L., MacPhail, C., Vermaak, K. & Kleinschmidt, I., (2004), HIV and sexual behavior among young South Africans: a national survey of 15-24 year olds, Johannesburg: Reproductive Health Research Unit, University of Witwatersrand

Philipson, T. J. & Posner, R. A., (1993), *Private Choices and Public Health, the AIDS Epidemic in an Economic Perspective*, Harvard University Press, Cambridge

Philipson, T. J. & Posner, R. A., (1995), The Microeconomics of the AIDS Epidemic in Africa, *Population and Development Review 21(4)*, 835-848

Piasani, E., (2008), *The Wisdom of Whores – Bureaucrats, Brothels and the Business of AIDS,* W. W. Norton & Co.

van der Pligt, J., (1996), Risk Perception and Self-Protective Behavior, *European Psychologist 1(1)*, 34-43

Quigley, J. M. & Raphael, S., (2005), Regulation of the High Cost of Housing in California, *American Economic Review 95 (2),* 323-328.

Quinn, T. C., Wawer, M. J., Wewankambo, N., Serwadda, D., Li, C., Wabire-Mangen, F., Meehan, M. O., Lutalo, T. & Gray, R. H., (2000), Viral Load and Heterosexual Transmission of Human Immunodeficiency Virus Type 1 *New England Journal of Medicine*, *342(13),* 921-929

Rao, C., Lopez, A. D., & Hemed, Y., (2006), Causes of Death, Jamison, D. T., Feachem, R. G., Makgoba, M. W., Bos, E. R., Baingana, F. K., Hofman, K. J., Rogo, K. O., (Eds), *Disease and Mortality in sub-Saharan Africa*, The World Bank, , Washington D C

Robinson, J. & Yeh, E., (2008), Transactional Sex as a Response to Risk in Western Kenya, MPRA Paper No. 7405

Schroeder, K. D., & Rojas, F. G., (2002) A Game Theoretical Analysis of Sexually Transmitted Disease Epidemics, *Rationality and Society 14(3),* 353-383

Simbayi, L. C., Chauveau, J. & Shisana, O., (2004), Behavoural Responses of South African youth to the HIV/AIDS epidemic: A nationwide survey, *AIDS Care – Psychological and Socio-medical Aspects of AIDS/HIV 16(5)*, 605-618

Swart-Kruger, J. & Richter, L. M., (1997), AIDS-related Knowledge, Attitudes and Behaviour among South African Street Youth: Reflections on Power, Sexuality and the Autonomous Self, *Social Science & Medicine*, *45(5),* 957-966

UNAIDS, (2008), Report on the global AIDS epidemic 2009, Geneva

UNAIDS, (2009), Report on the global AIDS epidemic 2009, Geneva

UNAIDS & IOM, (2003), *Mobile Populations and HIV/AIDS in the SOUTHERN AFRICAN REGION – Desk Review and Bibliography on HIV/AIDS and Mobile Populations,* IOM-UNAIDS Reports on Mobile Populations and HIV/AIDS, Geneva

UNAIDS/WHO, (2002), AIDS Epidemic Update, Geneva

VanLandingham, M. J., Suprasert, S., Sittiitrai, W., Vaddhanaphuti, C., Grandjean, N., (1993), Sexual Activity Among Never-Married Men in Northern Thailand, *Demography 30(3),* 297-313

van der Velde, F. W., Hoykaas, C. & van der Pligt, J., (2002), Conditional verus unconditional risk estimates in models of AIDS-related behavior, Marks, D. F (Ed), (2002), *The Health Psychology Reader*, Sage Publications, London

van der Velde, F. W. & van der Pligt, J., (1991), AIDS-Related Health Behavior: Coping, Protection Motivation, and Previous Behavior, *Journal of Behavioral Medicine 14(5),* 429-451

van der Velde, F. W., van der Pligt, J. & Hooykaas, C., (1994), Perceiving AIDS-Related Risk: Accuracy as a Function of Differences in Actual Risk, *Health Psychology 13(1)*, 25-33

Varga, C. A., (1997), Sexual Decision-making and negotiation in the midst of AIDS: youth in KwaZulu-Natal, South Africa, *Health Transition Review 7(S3)*, 45-67

Williams, B. G., Taljaarnd, D., Campbell, C. M., Gouws, E., Ndhlovu, L., van Dam, J., Caraël, M. & Auvert, B., (2003), Changing patterns of knowledge, reported behaviour and sexually transmitted infections in a South African gold mining community, *Journal of Acquired Immune Deficiency Syndrome 17*, 2099-2107

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