ECONOMETRIC GAME 2010: CASE B

April 12, 2010

1. Introduction

HIV/AIDS is widely recognized as a serious public health problem in Africa. The estimated adult HIV prevalence in Sub-Saharan Africa Africa is 6.1%, with 24.5m people infected and 2.0m deaths in 2005 according to an UN report. In some countries such as South Africa the rate of infection may be up to 20%. With the dissemination of the HIV virus progressively assuming epidemic dimensions, health authorities and aid organizations are directing substantial resources to voluntary HIV antibody counseling and testing (VCT) in an attempt to curb the rate of growth of infections. Despite these efforts, the success of these programs has been hampered by very low participation rates. Kalichman and Simbayi (2003) notes that only one in five South Africans aware of VCT have been tested.

In this paper we seek to better understand why this is so. Can we identify the most relevant determinants for the willingness to participate in the blood test? Possible causes may include religious reasons, reluctance to know one's own HIV status due to AIDS related stigmas, misinformation about the HIV virus, distrust of the testing services, etc. To study this question, we turn to a survey data set from an undisclosed African country which asks individuals whether they are willing to participate in a free anonymous blood test (which among other medical conditions tests for HIV) and collects a large amount of information regarding socioeconomic characteristics, HIV attitudes, risk attitudes, and personality traits.

In a predictive analysis we find that family structure and relative relationships in the household are important determinants of willingness to participate in VCT. Furthermore, our results indicate that education and HIV awareness predict important increases in the willingness to participate in blood testing. Personality traits and risk factors are dubious, even though the openness personality trait is negatively related to willingness to take the blood test.

In particular, we find that interactions within the household strongly influence the propensity of different members to participate in VCT testing. In particular, we find that the participation of non head of household members depends strongly upon the participation of the male head of household. Using an IV approach, we find that this effect is stronger on their wives but less important on their sons. In contrast, the propensity of female head of households to participate in VCT testing shows strong positive spillovers on their sons. These effects would suggest that policies targeting female spouses must find a way to circumvent male decisions. It also confirms the positive effects women head of households have on the education and health of their children

2. Data

In this study we examine a national survey conducted in an undisclosed African country. The sample consists of 2711 thousand participants of diverse background. Of direct interest for our purposes is the adult population, so that we have excluded subjects with less than 18 years from the original sample of 5939 observations. 67% of the subjects reported willingness to take the blood test. The questions included in the survey collect information on demographic and income characteristics, risk attitudes, possible personality traits, HIV testing history, HIV risk behavior history, HIV prevention knowledge, HIV testing attitudes and AIDS stigmas. The reader is referred to Appendix A for a detailed description of the question groups.

Despite the wealth of information provided by the 131 variables included in the data set, a major limitation of our sample is that for many variables the number of missing observations is substantial. In some cases, there are too few actual reported values for the variable to be useful in a general analysis. Some of the most relevant variables that are subject to this problem is the current level in school, questions regarding AIDS stigmas, the variables related to medical history in the past year and, how often the subject uses condoms with casual partners and whether the subject has exchanged sex for money.

2.1. **Variables.**

2.1.1. *Socioeconomic Characteristics.* One may suppose that participation varies substantially with education and income levels. In poor countries these variables are often unavailable or measured with error. Since formal income figures tend to ignore revenues from informal streams we make use of per capita aggregate consumption (in log form) as a proxy for relative wealth within each household. Furthermore, since education levels are unavailable, we simply introduce a dummy if the individual obtained any form of education.

Religious beliefs may also strongly affect the willingness to participate in blood testing. 78% of the population in our sample are muslim, a group which is of special concern. Pious muslims follow the writings of the Coran which forbids the contact of a blade to the flesh. For many muslims, blood testing is impossible to the use of syringes. Also, in predominantly muslim countries, family structures may also strongly affect participation of married women in blood testing and responsiveness to questions related to sexual behavior. Family structures are therefore explicitly introduced into the estimation, as explained in the previous paragraph.

2.1.2. *Personality Traits.* Even though there is extensive evidence in the economics and psychology literature showing that cognitive ability is powerful predictor of economic and social outcomes, only more recently the intuition that other traits beside the capacity for abstract thought, reasoning and problem solving are relevant found its way into the study of socioeconomic outcomes by economists. Underlying this trend are the advances in the personality psychology literature over the last two decades, which has produced a collection of stylized facts about how people think, feel and act.

While economists have been typically reluctant to use preferences to explain behavior, the robustness of the evidence from the personality psychology has motivated the use of personality traits as meaningful constraints on agent choice behavior. Moreover, conventional economic preference parameters may be interpreted as consequences of these constraints: for example, high rates of time preference may come from the inability of agents to delay gratification. For a comprehensive discussion of the interface between personality psychology and economics see Borghans et al. (2008).

Personality traits are defined as patterns of thought, feelings, and behavior. With a pragmatic focus, economists have begun to use measurement systems for personality traits developed by psychologists. Most prominent is the "big five" psychology inventory. Srivastava (2010) offers an introductory discussion of the Big Five personality factors. Our data set includes 33 personality related questions (each of which could be answered on a scale from one to five) designed to measure four personality factors. These are:

(1) Openness to experience: captures receptivity to novel experiences and ideas.
(2) Conscientiousness: measures motivation and diligence.
(3) Extraversion: preference for personal interaction, activity level and cheerfulness.
(4) Neuroticism: indexes the propensity to experience negative emotions like anxiety, anger and depression.

In Section 3 we detail how we use a factor analysis methodology to construct measures for the four personality factors from the personality related questions present in the sample. This is done to summarize the information contained in the personality questions in a way that is parsimonious and can be meaningfully interpreted.

2.1.3. *Risk Attitudes & HIV knowledge.* Risk Attitudes and knowledge of disease and health issues may strongly correlate to an individuals propensity to get tested for health problems. Those people with a greater knowledge of possible threats to their health are more likely to obtain regular check ups. These factors are introduced through several variables. In a first step, we introduce an indicator for people who have ever heard of HIV-AIDS, since knowledge of health related dangers may strongly influence willingness to get tested. We then introduce risk related variables through several measures: (i) the age at first sexual relationships and (ii) we use factor analysis to create a generalized risk measure based on questions related to an individuals sexual behavior, alcohol consumption, driving habits among others. Our motivation for using these different risk related measures is to account for possible measurement errors and unobserved heterogeneity in risk related responses.

3. Modeling Framework

The choice of an individual i, for $i = 1, 2, \ldots, n$, to participate voluntary in a free HIV test is modeled as a Bernoulli experiment:

$$y_i = \begin{cases} 1 & \text{with probability } \pi_i \\ 0 & \text{with probability } 1 - \pi_i \end{cases} . \tag{1}$$

We model the probability π_i by

$$\pi_i = \mathbb{P}\left[y_i = 1 | \theta_i\right] = F(\theta_i), \tag{2}$$

where $F(\cdot)$ is a cumulative distribution function to ensure that $0 \leq \pi_i \leq 1$. Signal θ_i is build up from the following components:

$$\theta_i = \gamma f_i + x_i \beta + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}\left(0, \sigma^2_{\varepsilon_i}\right), \tag{3}$$

where γ is a $1 \times p$ parameter vector, f_i is a $p \times 1$ vector of latent factors that represent the big four and the risk behavior factors. In Section 3.2 we discuss how the latent factors are constructed. x_i is a $1 \times m$ vector of explanatory variables, which are measured by $m \times 1$ parameter vector β. In this paper we model the probability π_i by a well known logit model and the more complex Scobit model proposed by Nagler (1994).

$$F\left(\theta_i\right) = \frac{e^{\theta_i}}{1 + e^{-\theta_i}} \tag{4}$$

$$F\left(\theta_i; \alpha\right) = \frac{1}{\left(1 + e^{-\theta_i}\right)^{\alpha}}, \tag{5}$$

where cumulative distribution function (5) is the Burr-10 distribution.
The logit specification imposes the assumption that individuals with a probability of 0.5 of choosing either of two alternatives are most sensitive to changes in independent variables. This assumption is imposed by the estimation technique because the logistic density functions are symmetric about zero. This might be a restrictive assumption, therefore we include the skobit (skewed-logit) specification, which include the parameter $\alpha > 0$ to measure possible skewness.
We summarize fixed and unknown parameters in the vector $\Psi = (\gamma, \beta', \sigma^2_{\varepsilon}, (\alpha))'$. The factors f_i are determined by factor analysis after which the model is estimated by maximum likelihood, as we detail in the following subsections.

3.1. **Maximum Likelihood Estimation.** We estimate the general binary outcome model by exact maximum likelihood. The logarithm of the probability density function of y_i conditional on the transformed probability θ_i is given for the logit specification (4) by

$$\log p(y_i|\theta_i) = y_i\theta_i - \log(1 + \exp(\theta_i)), \tag{6}$$

and for the Scobit specification (5) by

$$\log p(y_i|\theta_i) = -y_i\alpha \log(1 + e^{-\theta_i}) + (1 - y_i) \log(1 - (1 + e^{-\theta_i})^{-\alpha}). \tag{7}$$

We assume independence for both specifications for all $i \neq j$. This leads to the following log likelihood

$$l(\Psi) = \sum_{i=1}^{n} \log p(y_i|\theta_i), \tag{8}$$

which can be maximized by an arbitrary quasi-newton optimization routine. In the results we only mention the marginal effects of change in a regressor on the conditional probability that y_i equals one. For our model these effects are given by

$$\frac{\partial p(y_i|\theta_i)}{\partial x_{ij}} = F(\theta_i)\beta_j, \tag{9}$$

for the explanatory variables and

$$\frac{\partial p(y_i|\theta_i)}{\partial f_{ij}} = F(\theta_i)\gamma_j, \tag{10}$$

for the factors.

The independence assumption mentioned is rather restrictive. We are dealing with households. Within these households there is a lot of social interaction. We expect that an individual living in a certain household is influenced by the other members of the household. To account for this dependence we will follow the general approach of **?**, while taking into account some simplifying restrictions. We will assume that the decision of whether to take a free HIV test for the household members is influenced only by the head of the household and the head of the household is not influenced. This seems a reasonable assumption because 1010 heads of households are male of which 799 are Muslim (In contrast to 321 female heads of which 240 are Muslim). It is often observed in the Muslim culture that the male/female head of household makes all the decisions for the other household members. We denote a specific household by k, for $k = 1, 2, \ldots, 1331$. The head of the household is indicated by i_k^*. The head of the household is assumed not to be influenced by the other individuals, therefore the specification in (3) remains the same. For the other members of the household the specification changes

$$\theta_{i_k} = \gamma f_i + x_i\beta + \psi \mathbf{1}_{i_k^*} + \phi x^{i_k^*} + \varepsilon_i^k, \quad \varepsilon_i^g \sim \mathcal{N}\left(0, \sigma^2_{\varepsilon_i^k}\right), \tag{11}$$

where the indicator $\mathbf{1}_{i_k^*}$ indicates whether the head of the household took the free HIV test. The parameter ψ measure the additional effect of the decision of the head of the household. The explanatory variables belonging to the head of the household are $x^{i_k^*}$ and are measured by ϕ. Now we assume independence again and model θ_i as a logit specification for which the likelihood becomes (6). The parameter vector of this specification $\Psi^* = (\gamma, \beta', \sigma_\varepsilon^2, \psi, \phi)'$.

3.2. **Factor Analysis.** Let z_{ij} be the answer of respondent i to question j, where $i = 1, \ldots, n$ and $j = 1, \ldots, k$. The k questions used here are the relevant questions for the personality traits. We have the following model

$$z_i = \Lambda f_i + \delta_i, \quad \delta_i \sim NID(0, \Sigma_{\delta_i}), \tag{12}$$

where z_i is a $k \times 1$ vector of observed survey responses, Λ is a $k \times p$ matrix of factor loadings which are estimated as byproduct, f_i is a $p \times 1$ latent factor ($p < k$) and δ_i is a $k \times 1$ vector of measurement errors. In order to lose the subscript i we stack equation (12) for the different respondents

$$Z = F\Lambda' + \Delta, \tag{13}$$

where Z is a $n \times k$ matrix of observed survey responses, F is a $n \times p$ matrix of latent factors, Λ' is the transpose of the $k \times p$ matrix of factor loadings and Δ is a $n \times k$ matrix of measurement errors. To make the factor analysis model tractable we constrain the covariance structure in the following fashion:

$$\mathbb{E}(\Delta' F) = \mathbf{0}, \tag{14}$$

$$\mathbb{E}(\Delta' \Delta) = \mathbf{\Theta}_\delta, \tag{15}$$

$$\mathbb{E}(F' F) = \mathbf{\Phi}, \tag{16}$$

$$\mathbb{E}(Z' Z) = \mathbf{\Sigma}. \tag{17}$$

The covariance structure of equation (13) is given by

$$\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}' + \mathbf{\Theta}_\delta. \tag{18}$$

To estimate the model we need to place the proper identification constraints on the matrices. More information of these constraints can be found in Jackman (2005). One of the constraints is that $\mathbf{\Phi}$ is an identity matrix. The model parameters are estimated via principal components by using the decomposition

$$\mathbf{\Sigma} = \mathbf{W}'\mathbf{D}\mathbf{W}, \tag{19}$$

where D is a diagonal matrix containing the eigenvalues of Σ in decreasing order and W is a $k \times k$ matrix of orthogonal eigenvectors. The idea of the principal component analysis is to find a parsimonious representation of the structure underlying the Z variables. This is accomplished by not using the full k dimensional decomposition in equation (20). But only by using the first p eigenvectors. This results in

$$\mathbf{\Sigma} = \mathbf{Z}'_{(p)}\mathbf{Z}_{(p)} + \mathbf{\Theta}_{\delta}, \quad (20)$$

where $\mathbf{Z}_{(p)}$ is the $k \times p$ matrix containing the first p eigenvectors of Σ. The representation in equation (20) is not unique. The factor loadings are identified only up to orthogonal rotation. It is possible to pick a particular rotation based on different criteria. We use the varimax rotation described in Kaiser (1967).

3.2.1. *Results: Factor Analysis.* Before estimating the various logit and skobit models we need to calculate the factors. This is done by the principal component analysis described in previous section. We include factors for the four broad dimensions of personality traits (extraversion, conscientiousness, neuroticism and openness to experience) and the risk taking behavior.

One possible pitfall is that the evidence on the importance of these major personality traits is based on studies conducted in western countries. However, these findings have not been analyzed in the context of African countries. Do these same dimensions describe the universe of human traits in Africa well? It is fundamental to check first whether we are indeed measuring what we intend to measure.

Therefore we perform two separate factor analyses. The first series of factor analyses separates the questions for each personality trait. For instance we estimate a factor for extraversion from the questions that are in western society related to extraversion. We present the factor loadings for each of these four factor analyses in the second column of Table 1 and 2. We only included the factor loadings from the factor which belonged to the largest eigenvalue. The usual rule of thumb selects the eigenvalues that are above one, however the eigenvalue results for all four factor analyses showed two eigenvalues above one. The second factor analysis we performed is on the full set of personality related questions. From this set we extracted the factors that corresponded to eigenvalues that were above one. We found four of such factors. We are interested in determining that these factors are actually strongly related to the specific personality traits.

Therefore we regress the factors, found in the second factor analysis, on the questions that relate to the specific personalty trait that each factor is supposed to represent. We have chosen the best matching combination of factors with the question set. The results of these regressions are presented in column 3 of Table 1 and 2. We see that the factor loadings coming from the first individual factor analysis do not match very well with the estimated regression coefficients.

We conclude that the survey conducted is not very suitable to extract measurements for some personality traits. The personality traits openness and extraversion seem to be reasonably well measurable from the second factor analysis. The estimated factor loadings of the individual factor analyses are reasonably close to the regression coefficient estimated after the second factor analysis. The other two, neuroticism and conscientiousness, are not measured accurately. Despite these results

we still use the four factors coming from the second factor analysis in the various logit and skobit regressions.
We also estimate factors for the risk taking behavior. There are many questions in the survey that attempt to measure this risk taking behavior. The questions are related to eating unhealthy food, drinking alcohol, engaging in unprotected sex, driving without seatbelt, waling in unsafe areas and sleeping without protection against hazardous insects. The factor analysis applied to the answers of these six questions came up with two factors that had eigenvalues larger than one. We use both of these factors in the logit and skobit regressions.

3.3. **Semi parametric Extension.** Instead of specifying a logit or scobit specification for $F(\theta_i; \alpha)$ we could also use a semi-parametric model. The advantage of this approach might be that they provide additional modeling flexibility. In this paper we propose to use the maximum rank correlation estimator of Han (1987). This is an unusual choice for a semi-parametric estimator, because it does not require the use of smoothing parameters such as the choice of bandwidth. This and the fact that the estimator is $\sqrt{n}$ consistent and asymptotically normal, makes it an attractive estimator. The estimator choses ψ to maximize

$$Q_n^{MRC}(\Psi) = \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j<i}}^{n} \mathbf{1}(y_i > y_j)\mathbf{1}(\theta_i > \theta_j) + \mathbf{1}(y_i < y_j)\mathbf{1}(\theta_i < \theta_j), \tag{21}$$

This estimator is called the maximum rank estimator because $Q_n^{MRC}(\Psi)$ has a multiple of Kendall's rank correlation coefficient between y_i and θ_i. We maximize this function by using a simplex algorithm implemented with starting values obtained by the parameter estimates from the logit specification. Standard errors for the estimates are usually obtained by non-parametric bootstrap.

4. Predictive Analysis

In this section we assume that the decision of taking the blood test is independent among members of the households. What can we learn from this modeling strategy? Do more sophisticated methods generate any new insights? Our main objective here is to better understand which variables predict willingness to participate in a blood test well for adults. In particular, we are concerned with the robustness of these variables to different specifications. As usual this type of analysis do not allow us to make causal statements, so that our results should be interpreted with caution.

We start the analysis by looking at the estimation and marginal effect results for the Logit and Scobit models in Tables 3 and 4. The first message that come out of the results is that we should be judicious in taking implications out of this analysis since the fit of the models is modest: in classifying subjects we are only able to do slightly better than an unconditional average. However, we should keep in mind that even small effects may represent real and important welfare gains for the (probably few) individuals at the margin. If the objective is to help formulate policy small effects may still be well worth aiming for in a cost-benefit analysis (even though we not derive any policy implications here).

Income (as approximated by log household per capita consumption) has a very positive effect in the two specifications. From the marginal effects tables, a 10% increase in consumption is associated with a 0.5% additive increase in the probability of taking the test. The younger cohort (18-25 years) is significantly more likely to take the test (belonging to this category is estimated to increase the probability of taking the test by 0.08). As discussed before, the effect of household structure is substantial with female subjects and female head of families meaningfully increasing the chances of taking the test. For example, a female head of household increases the chances of taking the test by 15 percentage points even after controlling for the gender of the subject. Other meaningful variables are whether the participant has attended school (with a positive effect, as expected), awareness of AIDS and the age of first sexual intercourse (which has a small positive effect). On the other hand, household size and muslim background do not seem to matter much.

The impact of the personality and risk factors, however, are difficult to interpret and only generate a couple of interesting results. We might expect that conscientiousness would be a relevant factor, but this hypothesis is not supported by our results. The only personality factor that is significant across the specifications is the openness factor, which puzzlingly *decreases* willingness to take the test. One plausible but speculative (ex post) explanation is that open people are more superstitious and hence less confident in their ability (or the ability of the health authorities) to do anything against AIDS. In any case, the effect in this case is substantive: a unit increase in the openness factor (measured from 1 to 5) decreases the probability of taking the test by 0.05. Finally, one of the risk factors has a significant and substantive effect on decisions. However, without a more solid understanding of the risk factors (something which is out of the scope of this report) the interpretation of this result is problematic.

Turning to the model specification, the introduction of skewness via the Scobit model does not seem to generate important modeling gains, with the skewness parameter α failing to be significantly from the nested Logit benchmark at the 5% level. In Figures 1 and 2 we show the estimated probabilities as a function of the parameters and covariates in the two cases, confirming the modest impact of skewness. Finally, the last column of Table 3 displays the parameter estimates in the Maximum Rank Correlation method. We do not report the standard errors since extremely time consuming nonparametric bootstrap is needed to calculate them. The results are in general coherent with the previous models. However, the effect of the openness personality trait essentially disappears.

4.1. **Identification of Social Interaction Within Households.** Underlying the social interaction literature, pioneered by Manski (1993 a,b, 1997), is the problem of identification. The researcher must be able to distinguish three effects due to social spillovers. We can discuss these three effects in the context of VCT testing:

(1) Endogenous Effects: captures the changes in an individuals propensity to take blood test depending upon whether the head of the household received blood testing.
(2) Exogenous effects: captures the changes in an individuals propensity to take a blood test depending upon the characteristics of the head of the household.
(3) Correlated Effects: captures the changes in an individuals propensity to take a blood test due to individuals facing similar institutional characteristics.

A basic identification in our model is that the participation in VCT testing of the head of household influences the decision of all other members in his family only. This rules out the reflection problem whereby members of the non-head members of the household will influence the choice of the head. This assumption seems well justified in the context of our study. Given the predominantly muslim population which tends to be male dominated, we expect little feedback effects from wives or daughters on VCT participation of male heads of household. Furthermore, we notice that all households with a female as the head have this female as the single parent. This rules out any feedback from a male husband on the female head of household decisions.

In order to separate the endogenous effects of interest from exogenous and correlated effects, we simply make use of a set of control family and individual characteristics. We assume that there are correlated effects depending upon household size, the fraction of males in the household, religion and the number of wives present in the household. The fraction of males in the household may introduce stronger social pressure for the females to follow the decision of their male counterparts. The number of wives may in turn correlate strongly to the degree of piousness. Indeed in a predominantly muslim society one can expect that men who choose to have multiple wives hold more traditional values (after controlling for income levels through household aggregate consumption). We also account for exogenous characteristics of the head of household through an indicator for whether he received education and his knowledge of HIV. We decide to leave out other head of household personality and risk characteristics since they suffer from endogeneity issues or report insignificant effects. We do however still include these personality, risk and demographic characteristics on an individual level.

4.2. **Estimation results of Social Interaction Within Households.**

4.2.1. *General head of household spillovers.* Table 5 presents the results for social interaction effects. For ease of comparison and interpretation, we present marginal effects for all non linear parameter estimates. The logit model constitutes our baseline specification. Assuming exogeneity of VCT participation of the head of household, the results indicate that VCT participation of male head of households increases the probability of VCT participation in his family by approximately 5% with a slightly smaller effect for female heads of households.

In terms of the exogenous interaction effects, we see that the probability of taking a blood test decreases with a higher fraction of males and a higher number of wives in a household. This would suggest hat more pious muslim household members, holding endogenous spillover effects constant, are less likely to take a blood test.

The second and third columns of Table 5 present robustness checks for these results with probit and linear specifications. As expected, the results change very little in the probit specification. The linear estimates seem to differ which justifies the non-linear underlying structure of our model.

4.2.2. *Spouse and Son specific spillovers.* General results of head of household spillovers may overlook targeted spillovers. Table 6 explores this possibility by focusing on spillover effects from one spouse to another and from the head of household to his son. Of course, the identification of the spillover effects of the head of household to his wife or from the wife to the head of household due to each others choice of participation in VCT testing may be endogenous. Indeed, we relax our assumptions here and allow for unobserved interactions which may strongly biased our estimates of the *endogenous effects.* To account for this possibility we use an IV approach where we use the spouses education and knowledge of HIV as exclusion restrictions to informally capture intelligence and health awareness characteristics. This assumption claims that intelligence and health awareness characteristics of the head (wife) will affect his (her) propensity to participate in VCT testing directly but will only affect his wives (her husbands) propensity to participate in VCT testing via her husband (his wives) participation in VCT testing. This is simply the fundamental assumption underlying identification in an IV model.

The results in the first row of Table 6 report the effect of male head of household VCT participation on his wife (wives) propensity to participate in VCT testing. We notice that the naive Logit and Probit estimates severely underestimate the endogenous interaction effect. This relationship does not apply the other way around. This is evident in the second row which shows that the IV estimations do not seem to vary the endogenous interaction effects estimates. This result show that male heads of households have far spillover effects on their wives. Thus, in a development optic, there is a fundamental need to target married women and to find ways to educate them independent of the influence of their husbands.

In our base model, we also noticed that the propensity to participate in VCT testing for female heads of households tended to be higher than their male counterparts. We may want to know if the spillover effects of head of household on their children differ between male and female head of households. The results in the last two rows of Table 6 clearly show that female head of household participation in VCT testing strongly influence their sons propensity to participate. The results are consistent with the body of literature in development which shows that female head of households are also more reliable in redistributing household income towards the education and health of their children

References

Borghans, L., A. Duckworth, J. Heckman, and B. ter Weel (2008): "The Economics and Psychology of Personality Traits," *Journal of Human Resources*, 43.

Jackman, S. (2005): "An Introduction to Factor Analysis," Mimeo.

Kaiser, H. (1967): "The Varimax Criterion for Analytic Rotation in Factor Analysis," *Psychometrica*, 23, 443–482.

Kalichman, S. and L. Simbayi (2003): "HIV testing attitudes, Aids Stigma, and Voluntary HIV couseling and testing in a black township in Cape Town, South Africa," *Sex Transm Infect*, 79, 442–447.

Nagler, J. (1994): "Scobit: An Alternative Estimator to Logit and Probit," *American Journal of Political Science*, 38.

Srivastava, S. (2010): "Measuring the Big Five Personality Factors," From the web.

Appendix A. Variable Groups

(i) Demographic characteristics: Participants were asked their sex, age, years of education completed, marital status and religion.

(ii) Personality traits: a number of questions were included with the object of acquiring information about the personality traits of subjects.

(iii) Risk attitudes: Participants were asked they engaged in a myriad of risky behaviors, such as eating unhealthy food, heavily drinking alcohol, walking alone at an unsafe area, etc.

(iv) HIV testing history: The survey included questions concerning history of HIV antibody testing. Participants who reported having been tested for HIV indicated the results of their most recent test as positive, negative, or that did not know the results. Furthermore, participants also reported the number of times they had ever been tested for HIV.

(v) HIV risk behavior history: To assess HIV risk history, participants indicated the number of sex partners they had in previous 12 months, whether they ever exchanged sex for money, how often they use a condom with regular partner, how many casual partners and how often use condom with casual partner.

(vi) HIV prevention knowledge: A 12 item test was used to assess HIV risk and prevention related knowledge. Items reflected information about HIV transmission, condom use and AIDS related knowledge and were responded to as yes, no or don't know.

(vii) HIV testing attitudes: The survey contained five HIV testing attitude item. Two items reflected positive outcomes from testing, two assessed adverse outcomes and one item reflected HIV testing avoidance. Items were responded to as either agree or disagree.

(viii) AIDS stigmas: Seven AIDS stigma items were used. The AIDS stigma items reflected negative beliefs about people living with AIDS and the endorsement of social sanctions against people with HIV. These items were also responded to as either agree or disagree.

Appendix B. Tables

Table 1. Principal Component Analysis for openness and extraversion.

Factor 1: openness		
Questions	Factor Loading	Regr.Coef. (Std.Err)
original	0.235	0.145 (0.016)
curious	0.195	0.060 (0.012)
ingenious	0.207	0.103 (0.014)
activeimagin	0.236	0.134 (0.015)
inventive	0.205	0.089 (0.012)
artistic	0.109	-0.108 (0.010)
reflect	0.200	0.098 (0.012)
artknowledge	0.048	-0.131 (0.010)
routinework	-0.201	-0.532 (0.011)
artnotlike	-0.041	-0.164 (0.010)
		Adjusted $R^2 = 0.6472$
Factor 4: extraversion		
Questions	Factor Loading	Regr.Coef. (Std.Err)
talkative	0.285	0.202 (0.010)
energy	-0.130	0.009 (0.011)
enthusiasm	-0.214	-0.216 (0.012)
assertive	-0.188	-0.181 (0.011)
outgoing	-0.024	0.014 (0.010)
reserve	0.326	0.008 (0.011)
quiet	0.350	0.205 (0.011)
shy	0.297	0.151 (0.010)
		Adjusted $R^2 = 0.5316$

[1] The full sample: $n = 2704$.

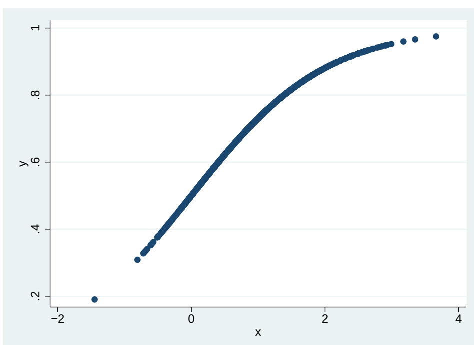

Figure 1. Logit CDF

Table 2. Principal Component Analysis for conscientiousness and neuroticism.

Questions	Factor Loading	Regr.Coef. (Std.Err)
Factor 3: conscientiousness		
thorough	0.201	-0.042 (0.024)
reliable	0.195	-0.069 (0.023)
perseveres	0.199	0.039 (0.022)
efficient	0.214	-0.070 (0.025)
planfollows	0.191	0.131 (0.022)
careless	0.153	0.045 (0.017)
disorganized	0.178	0.007 (0.018)
lazy	0.176	0.008 (0.016)
distracted	0.120	-0.047 (0.016)
		Adjusted $R^2 = 0.0202$
Factor 2: neuroticism		
depressed	0.237	0.066 (0.009)
tense	0.228	0.063 (0.009)
worries	0.264	0.004 (0.009)
moody	0.263	0.047 (0.009)
nervous	0.178	0.096 (0.008)
relax	0.247	-0.854 (0.009)
emstable	0.234	0.152 (0.009)
calm	0.186	0.183 (0.009)
		Adjusted $R^2 = 0.7827$

[1] The full sample: $n = 2704$.

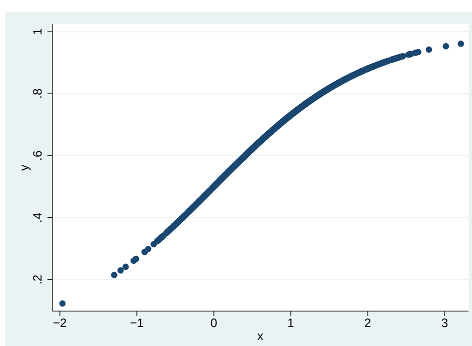

Figure 2. Scobit CDF

TABLE 3. Estimation results for the Logit, Scobit and Maximum Rank Correlation models

Variables	Logit	scobit	MRC
attended school	0.3713 (0.002)	1.1850 (0.307)	0.6484
has ever heard of AIDS	0.5458 (0.000)	1.7782 (0.294)	1.4743
household size	-0.0461 (0.704)	-0.2749 (0.596)	-0.0483
household size squared	0.0170 (0.128)	0.0662 (0.396)	0.0282
Fractions of males in household	0.0630 (0.847)	-0.2655 (0.824)	0.0796
Number of spouses in household	-0.2323 (0.038)	-0.7165 (0.352)	-0.3327
head of house is female	0.8722 (0.000)	2.7474 (0.290)	0.1073
wife of the head	0.5208 (0.000)	1.6126 (0.291)	1.0410
son of the head	-0.8077 (0.005)	-1.9961 (0.241)	-0.2771
daughter of the head	1.1619 (0.070)	4.4645 (0.380)	1.4174
age 18 to 25	0.4911 (0.014)	1.5944 (0.338)	0.7181
age 26 to 35	-0.0711 (0.608)	-0.2952 (0.557)	-0.0686
age 36 to 45	-0.0999 (0.453)	-0.2925 (0.534)	-0.1148
log consumption	0.2295 (0.009)	0.8197 (0.347)	0.3344
Is Muslim	-0.0805 (0.537)	-0.3308 (0.529)	-0.0791
age of first time sex	0.0466 (0.000)	0.1401 (0.295)	-0.1108
factor1 : openness	-0.1990 (0.000)	-0.6688 (0.316)	0.0085
factor2 : neuroticism	-0.0274 (0.574)	-0.1468 (0.522)	-0.0311
factor3: conscientiousness	-0.0628 (0.273)	-0.2063 (0.438)	-0.0726
factor4: extraversion	0.0233 (0.696)	-0.0119 (0.950)	-0.0014
factor5: risk taking behavior	-0.0271 (0.567)	-0.1183 (0.537)	-0.0496
factor6: risk taking behavior	0.2171 (0.000)	0.7544 (0.316)	0.1315
Constant	-3.4759 (0.002)	-8.1832 (0.329)	
α		-1.5766 (0.115)	
log likelihood	-1339.791	-1338.125	
Pseudo R^2	0.0626	–	
% Correctly Classified	69.13	68.86	

Table 4. Marginal parameter estimation results on logit and scobit models

Variables	Logit	scobit
attended school	0.076 (0.025)	0.069 (0.022)
has ever heard of AIDS	0.117 (0.023)	0.111 (0.022)
household size	-0.010 (0.025)	-0.016 (0.024)
household size squared	0.004 (0.002)	0.004 (0.002)
Fractions of males in household	0.013 (0.068)	-0.016 (0.064)
Number of spouses in household	-0.048 (0.023)	-0.042 (0.020)
head of house is female	0.155 (0.035)	0.132 (0.031)
wife of the head	0.105 (0.023)	0.091 (0.021)
son of the head	-0.188 (0.072)	-0.142 (0.075)
daughter of the head	0.180 (0.066)	0.173 (0.063)
age 18 to 25	0.093 (0.034)	0.082 (0.030)
age 26 to 35	-0.015 (0.029)	-0.018 (0.026)
age 36 to 45	-0.021 (0.028)	-0.017 (0.025)
log consumption	0.048 (0.018)	0.048 (0.016)
Is Muslim	-0.017 (0.026)	-0.019 (0.025)
age of first time sex	0.010 (0.002)	0.008 (0.002)
factor1 : openness	-0.041 (0.011)	-0.039 (0.010)
factor2 : neuroticism	-0.006 (0.010)	-0.009 (0.009)
factor3: conscientiousness	-0.013 (0.012)	-0.012 (0.010)
factor4: extraversion	0.005 (0.012)	-0.001 (0.011)
factor5: risk taking behavior	-0.006 (0.010)	-0.007 (0.008)
factor6: risk taking behavior	0.045 (0.012)	0.044 (0.012)

Table 5. Parameter estimation result including outcome head of household

Variables	Logit - Marginal	Probit - Marginal	Linear Regression
Indicator male head took blood test	0.550 (0.022)	0.565 (0.020)	0.608 (0.016)
Indicator female head took blood test	0.341 (0.018)	0.353 (0.015)	0.777 (0.037)
attended school	0.066 (0.026)	0.071 (0.027)	0.050 (0.018)
has ever heard of AIDS	0.100 (0.026)	0.097 (0.026)	0.067 (0.016)
household size	-0.070 (0.028)	-0.074 (0.028)	-0.044 (0.018)
household size squared	0.006 (0.003)	0.006 (0.003)	0.004 (0.002)
Fractions of males in household	-0.231 (0.075)	-0.218 (0.076)	-0.132 (0.049)
Number of spouses in household	-0.011 (0.024)	0.003 (0.025)	-0.002 (0.017)
head of house is female	-0.251 (0.111)	-0.149 (0.083)	-0.107 (0.041)
wife of the head	0.126 (0.023)	0.112 (0.025)	0.103 (0.018)
son of the head	-0.219 (0.096)	-0.226 (0.081)	-0.138 (0.047)
daughter of the head	0.118 (0.064)	0.149 (0.072)	0.052 (0.068)
age 18 to 25	0.108 (0.028)	0.107 (0.033)	0.086 (0.028)
age 26 to 35	-0.001 (0.030)	-0.002 (0.031)	-0.002 (0.021)
age 36 to 45	-0.033 (0.031)	-0.031 (0.031)	-0.025 (0.020)
log consumption	0.004 (0.019)	-0.001 (0.019)	-0.003 (0.013)
Is Muslim	0.006 (0.029)	-0.001 (0.029)	0.008 (0.019)
age of first time sex	0.007 (0.003)	0.008 (0.003)	0.005 (0.002)
factor1 : openness	-0.027 (0.011)	-0.030 (0.011)	-0.018 (0.008)
factor2 : neuroticism	-0.016 (0.011)	-0.017 (0.011)	-0.012 (0.007)
factor3: conscientiousness	-0.034 (0.013)	-0.035 (0.013)	-0.022 (0.009)
factor4: extraversion	-0.010 (0.013)	-0.010 (0.013)	-0.008 (0.009)
factor5: risk taking behavior	-0.011 (0.011)	-0.009 (0.011)	-0.010 (0.007)
factor6: risk taking behavior	0.041 (0.012)	0.043 (0.012)	0.025 (0.007)
constant	-	-	0.254 (0.169)
log likelihood	-834.65	-842.51	-

Table 6. Parameter estimation results for model I, II and III

Model I				
Variables	Logit	Probit	IV probit	2SLS regression
Indicator for spouse equals 1 if male head took blood test	0.365 (-0.025)	0.350 (0.024)	0.722 (0.026)	0.750 (0.105)
Model II				
Variables	Logit	Probit	IV probit	2SLS regression
Indicator for male head equals 1 if spouse took blood test	0.358 (0.025)	0.356 (0.025)	0.383 (0.097)	0.357 (0.087)
Model III				
Variables	Logit	Probit		
indicator for son equals 1 if male head took blood test	0.009 (0.003)	0.007 (0.003)		
indicator for son equals 1 if female head took blood test	0.228 (0.082)	0.199 (0.069)		