Econometric Game 2010

Modelling the Willingness to take part in a voluntary HIV test

Case B

1. Introduction

AIDS has been called one of the main development emergencies which need an urgent act. Although the history of its origin can be traced back to 19^{th} century, it is still one of the most challenging problems which human is facing, especially in some specific regions of the world like Sub-Saharan Africa where the share of HIV infected individuals in the population (the HIV prevalence rate) is around 5% (see UNAIDS (2009)).

According to different studies, one of the main factors in controlling AIDS is HIV testing which has been a successful policy for both prevention and treatment of the disease (Boysen et. al, 2009). Many explanatory variables like gender, age, religion, stigmatization, family matters, infrastructure and sex related knowledge and behaviors have been tested and the results have been used for making more effective policies for control AIDS around the world (Matovu and Makumbi 2007; Gage and Ali, 2007; Rou et.al, 2009 and our first submitted report).

In this paper we have investigated some new factors associated with HIV testing in an anonymous country in Africa according to the data provided by Econometric Game 2010. The first modelling exercise attempts to explore the affects of family position of a participant on their test uptake decision. The second model will investigate and relax assumptions made about the error term in binary choice models (potentially allowing better estimation of the parameters of interest).

According to our knowledge the literature around the first subject is not very noteworthy. For instance, Gage and Ali (2010) have recorded the “relationship to the head of household” but have not used it in their analysis. Other studies have also considered the family position of their participant in order to discover the effects of the partner consent on testing (Homsey et. al, 2010) or for considering the financial implications of HIV/AID (Mahal, 2004).

2. Investigation of the effects of partner decisions on HIV testing uptake

As the objective of the exercise is to determine the effect of household head/spousal decisions, the modelling exercise is to be split into two components. Firstly, the latent propensity of household head i to take the test for HIV, $y_i{}^*$, is assumed to be determined by some linear function that is dependent on some variables, x_i, as well as the (binary) decision of their spouse, z_i.

$$y_i{}^* = x_i^{'}\beta_1 + z_i^{'}\gamma + \varepsilon_i$$

The choice of outcome, y_i, is then expressed as

$$y_i = \begin{cases} 0 & if\ y_i{}^* \le 0 \\ 1 & if\ y_i{}^* > 0 \end{cases}$$

Then, under the assumption that $\varepsilon \sim N(0,1)$, this leads to the following log-likelihood function for the model:

$$\ell(\beta)_i = y_i ln\Phi(x_i^{'}\beta + z_i^{'}\gamma) + (1 - y_i)ln\,[1 - \Phi(x_i^{'}\beta + z_i^{'}\gamma)]$$

This model will hereon be referred to as model 1. A similar model, estimated separately, considers the decision for each spouse. Assuming the latent propensity equation for spouse i, $z_i{}^*$,

$$z_i{}^* = w_i^{'}\beta_2 + y_i^{'}\delta + u_i$$

where w_i possibly, but not necessarily, includes the same variables as x_i. The choice of outcome, y_i, is then expressed as

$$z_i = \begin{cases} 0 & if\ z_i{}^* \le 0 \\ 1 & if\ z_i{}^* > 0 \end{cases}$$

Then, under the assumption that $u \sim N(0,1)$, this leads to the following log-likelihood function for the model:

$$\ell(\beta)_i = y_i ln\Phi(w_i^{'}\beta + y_i^{'}\delta) + (1 - y_i)ln\,[1 - \Phi(w_i^{'}\beta + y_i^{'}\delta)]$$

This model will hereon be referred to as model 2. Of course, for the first model we assume that the spouses' decisions are not dependent on the household head's decision, while the second model assumes the household head's decision is not dependent on the spouse's decision. However, in reality this is unlikely to be true. If the household head makes its decision taking into account the propensity of his spouse to take the test, and the spouse make its decision taking into account the propensity of the household head's propensity, then,

$$y_i{}^* = x_i^{'}\beta_1 + z_i{}^*\gamma + \varepsilon_i$$

$$z_i{}^* = w_i^{'}\beta_2 + y_i{}^*\delta + u_i$$

Therefore,

$$y_i^* = x_i'\beta_1 + (w_i'\beta_2 + y_i^*\delta + u_i)\gamma + \varepsilon_i$$

$$y_i^* = x_i'\beta_1 + w_i'\eta + y_i^*\theta + u_i\gamma + \varepsilon_i$$

$$y_i^*(1-\theta) = x_i'\beta_1 + w_i'\eta + u_i\gamma + \varepsilon_i$$

$$y_i^* = \frac{x_i'\beta_1}{(1-\theta)} + \frac{w_i'\eta}{(1-\theta)} + \frac{u_i\gamma + \varepsilon_i}{(1-\theta)}$$

Therefore, as y_i is a monotonic function of y_i^*, assuming that $\gamma \neq 0$, the error terms are correlated, and the equations are endogenous. In order to deal with this, a bivariate probit model could be used, taking into account the correlation of the error terms.

$$y_i^* = x_i'\beta_1 + z_i\gamma + \varepsilon_i$$

$$z_i^* = w_i'\beta_2 + y_i\delta + u_i$$

$$corr(\varepsilon_1, u_i) = \rho$$

This implies the following log-likelihood function:

$$\ell(\beta) = yzln\Phi_2\left(x_i'\beta_1 + z_i\gamma, w_i'\beta_2 + y_i\delta, \rho\right) + (1-y)(z-1)ln\Phi_2\left(-x'\beta, -w_i'\beta_2, \rho\right)$$

$$+y(1-z)ln\Phi_2\left(x_i'\beta_1 - w_i'\beta_2 - y_i\delta, -\rho\right) + (1-y)zln\Phi_2\left(-x_i'\beta_1 - z_i\gamma, w_i'\beta_2, -\rho\right)$$

Of course, exclusion restrictions apply for the bivariate probit case, so now w_i and x_i are not constructed using exactly the same variables. Unfortunately, estimation of the bivariate probit could not be performed using Stata (we suspect because of data issues). As a result, three stage least squares was used to approximate the problem. That is, a linear probability model is assumed for each decision maker. This has the obvious flaws of unbounded probabilities, heteroskedasticity and potentially inappropriate marginal effects. However, it often gives close results to other binary choice models, and is therefore considered the best feasible method of dealing with the endogeneity, given the time constraints.

Method

In order to save time, data and additional variables, including imputations, used in the previous day of competition were utilised. Although a full explanation of how variables were derived is not possible, given the time constraints, a brief summary of some important variables manipulations, performed on the previous day is as follows:

- Simplification of many categorical variables to ensure parsimonious estimation;
- Construction of a household log-income variable from individual incomes;
- Random regression imputation (Gelman & Hill, 2006) of missing observations from many variables, including stigma-related variables, income and the index of HIV understanding;

- Indices for four of the "big five" personality factors (Srivastava, 2010), constructed using principle components analysis;
- An index summarising knowledge about HIV transmission to children, again constructed with principle components analysis; and,
- An index summarising factors that may imply a stigma towards HIV and AIDS.

Further data manipulation that was required in order to deal with this problem included the following. Data was cleaned such that only household heads and their spouses were considered. The first step involved removing all observations for which the individual was not a household head or a spouse. This decision required the assumption that out of sample members did not affect household head and spousal decisions, which we considered a reasonable assumption, due to the fact that couples often make joint decisions that do not depend on the preferences of their adult children.

The next step was to remove observations for which there was only a household head in the household, and households for which there was only a spouse. The latter may represent the case where the household head is under 18, but the spouse is an adult. This procedure was motivated by our interest only in the observations for which there was a household head and spouse. However, it is clear to see that if we would like to get unbiased and consistent results, the decisions of single household heads may need to be taken into account. This potential problem was ignored for this analysis, due to time constraints. The model estimated, however, will only be valid conditional on the household having a spouse. Removing the observations was achieved by splitting the data into a household head data set, and a spousal data set. The datasets were then merged according to household id, such that if any household id's did not at least one spouse and one household head it was removed.

There 16 households (out of 806) where the number of spouses was equal to "2". For these cases, an indicator variable was generated, taking the value of "1" if both spouses agreed to take the test, and "0" otherwise. This ensured the number of overall spousal decision observations (793) was equal to the number of household head observations in model 1. It should be noted that in model 2 all 809 observations were utilised.

Variable choice

As the variable of interest for these models is the effect of family members participation decisions on others' participation decisions, we have selected the control variables in the following manner.

- Economic relationships and statistical fit have been used to select a range of initial variables, based on the factors deemed influential in the first round of the competition.
- The models were then run, and the statistical significance of each variable was considered.
- Variables that showed a lower degree of significance, and for which the economic motivation was debateable, were removed from the model.
- Information criteria, including the Akaike information criterion (AIC) and Bayesian information criterion (BIC), were compared to ensure a better model fit was obtained,

given the number of variables. This ensured more parsimonious estimation, and implied marginal gains in efficiency.

Results

Estimation results from the separate individual choice models, as well as the bi-varying model estimated by 3SLS, can be found in appendices one to five. For the separately estimated models of individual choice, the influence of both partner's decisions appear to affect each other, even (asymptotically) significant at the 1% level. This seems to be consistent with the 3SLS model results. However, the partner variables in the household head and spouse models appear less significant than those in the 3SLS case. This may imply that accounting for endogeneity does have a positive effect on the model, and is therefore the correct choice. While it would be nice to investigate the bias of coefficient estimates, this is not possible due to the different functional form of the 3SLS model to the probit estimations.

While the overall measures of fit for the 3SLS model and the probit models should not be compared directly, it is clear from examining the hit-miss tables[1] in appendix 2 and 4, that the probit models predict the data to a better extent than the naive model. This suggests that the models are a good fit. However, these results may be biased by the model's endogenous nature. When comparing to the 3SLS model results, it should be noted that the probability bound problem arose, as probabilities below zero and above one. If this is ignored, however, an interesting result if found. While the choices made by household heads are well predicted (69.7% correct vs. 63.2% for the constant model), the model only made marginal gains over the constant model when predicting the choices of spouses. Therefore, the high rate of correct prediction of spousal decisions in the second model, may have been a result of endogenous factors, rather than good model results. Overall, however, it appears that the decision of an individual's partner does influence the individual's decision to be tested for HIV.

[1] Note that for these tables and outcome is predicted if the predicted probability of observing that outcome is greater than 0.5. Although other thresholds could have been applied, this was considered outside the scope of the study.

3. Binary choice error distribution assumption investigation

The next area of investigation consists of looking at the effect of different distributional forms on the effectiveness of a simple binary choice model. The two binary choice models that are commonly used are the probit and logit models, which imply assumptions of normal and logistic distributed error terms. However, assumptions that the error terms are symmetric may not necessarily be appropriate for some sets of data. In order to test this, this research looks at the implementation of a skewed error distribution assumption, as well as specifying a semi-parametric error distribution.

An extension of the logit model that allows for skewness in the functional form for $P(Y = 1 \mid X=x=F(x'\beta)$ is the scobit model, which uses the Burr-10 CDF:

$$F(x'\beta, \alpha) = \frac{1}{(1 + e^{-x'\beta})^{\alpha}}$$

If the additional parameter α is equal to one, the scobit model is equivalent to the logit model, as the CDF's are the same (Nagler, 1994). Thus, a likelihood ratio test of $H_0: \alpha = 1$ is to test the null that the density is not skewed (effectively, a test of the logit specification versus the scobit specification). The likelihood function for the scobit model is the same as for the probit and logit models, but with the Burr-10 CDF in place of the normal or logistic CDF.

All three the probit, logit and scobit models were estimated as per appendix 6. There were only marginal differences between the coefficient estimates from the probit and logit models. However, none of the scobit model parameters were found to be significant. This may be an indication that the scobit model was not appropriate for the data, and that the true CDF may not be skewed.

Overall measures of model fit appear to agree with this, with the hit-miss table indicating small differences in the predictive ability of the models. While the scobit obtained better results than the logit, an LR test indicates that the scobit model would not be preferred over the simple logit (p = 0.46).

Overall, the probit managed to predict the highest proportion of correct results. In addition, while the maximised log-likelihood of the scobit estimator is slightly higher than in the probit case, comparison of the AIC and BIC for each of these models indicates that the gain in log-likelihood is not worth the estimation of the additional parameter, α (ie: assuming a burr distribution for the error term is not justified by the improvement in solution).

As a result of this, we have extended to a semi-nonparametric specification based on the Gaussian distribution. This is the estimator proposed by Gallant & Nychka (1987). This is the product of a normal distribution and a polynomial square term. The functional form is approximated by:

$$f_K(\varepsilon) = \frac{1}{\theta}\left(\sum_{k=0}^{K} \gamma \varepsilon^k\right)^2 \phi(\varepsilon)$$

where the term:

$$\theta = \int_{-\infty}^{\infty} (\sum_{k=0}^{K} \gamma \varepsilon^k)^2 \phi(\varepsilon)\, d\varepsilon$$

ensures that the density integrates to 1. The normal density function was specified, in order to compare the results of the model estimation with those from the best binary choice model predicted earlier (the probit model).

The order of the polynomial K is specified by the user; we have specified a third order polynomial, in order to allow for some skewness and kurtosis, while ensuring the model remained parsimonious. Thus coefficients $\gamma_1, \gamma_2, \gamma_3$, needed to be estimated, as well as the original explanatory variable parameters. Due to the nested normal distribution, a likelihood ratio test can be performed to compare this semi-nonparametric specification to a simple probit model.

Estimation of the semi-parametric model obtained the results shown in appendix 7. Formally testing for this generalised form of the model results in the null hypothesis (the parametric probit model being true) not being rejected (p-value = 0.61). That is, there is insufficient evidence to suggest that the increased flexibility in the error distribution is justified, and a probit model would be more suitable in this case.

4. Conclusions

In this case we have investigated whether the decision to take a HIV test is independent of a partner's decision to take the test. In addition, we have analyzed the fit of the parametric specifications of the model, and compared to a non-parametric specification.

Modelling whether the effects of whether a partner takes a HIV test is difficult, because it involves estimating simultaneous equations, or otherwise making unreasonable assumptions about the direction of the effects. Although a simultaneous equation bivarite probit would be ideal, given the time constraints faced, we estimated a linear three stage least squares model to approximate the effects; a linear model is often a good approximation of binary choice equations. Both models we estimated showed that a 'spouse' taking a HIV test has a significant positive effect on the 'household head' also taking the test. These results also held in the opposite direction, although the effect was less strong.

For analyzing the fit of the data, a scobit model was estimated to check for skewness. A likelihood ratio test as well as AIC and BIC showed that this model did not improve the fit of the results. The semi-nonparametric model estimated was compared to the binary probit model; in this case, a likelihood ratio test also showed that we could not reject the null of a probit specification. Thus, our conclusion is that allowing for skewness and estimating semi-nonparametric models does not significantly improve the fit of results.

Reference List

Gallent, A., & Nychka, D. N. (1987). Semi-nonparametric maximum likelihood estimation. *Econometrica , 55*, 363-390.

Gelman, A., & Hill, J. (2006). *Chapter 25: Missing-data imputation.* New York: Cambridge University Press.

Nagler, J. (1994). Scobit: an alternative estimator to logit and probit. *American Journal of Political Science , 38* (1), 230-255.

Srivastava, S. (2010). *Measuring the Big Five Personality Factors.* Retrieved April 12, 2010, from http://www.uoregon.edu/~sanjay/bigfive.html

Mahal, A. (2004). "Economic Implications of Inertia on HIV/AIDS and Benefits of Action." Economic and Political Weekly 39(10): 1049-1063.

Matovu, J. K. B. and F. E. Makumbi (2007). "Expanding access to voluntary HIV counselling and testing in sub-Saharan Africa: alternative approaches for improving uptake" Tropical Medicine & International Health 12(11): 1315-1322.

Boyer, S., F. Marcellin, et al. (2009). "Financial barriers to HIV treatment in Yaoundé, Cameroon: first results of a national cross-sectional survey." Bulletin of the World Health Organization 87: 279-287.

Homsy, J., R. King, et al. (2007). "The Need for Partner Consent Is a Main Reason for Opting Out of Routine HIV Testing for Prevention of Mother-to-Child Transmission in a Rural Ugandan Hospital." JAIDS Journal of Acquired Immune Deficiency Syndromes 44(3): 366-369 10.1097/QAI.0b013e31802f1303.

Rou, K., J. Guan, et al. (2009). "Demographic and Behavioral Factors Associated With HIV Testing in China." J Acquir Immune Defic Syndr 50(4): 432-434.

Gage, A. J. and D. Ali (2005). "Factors associated with self-reported HIV testing among men in Uganda." AIDS Care: Psychological and Socio-medical Aspects of AIDS/HIV 17(2): 153 - 165.

Appendix 1: Test participation decision probit model (Household heads, First estimation)

	Coefficients	Marginal Effects
Spouse Test	0.886	0.336
	(8.24)**	(8.26)**
Age 18-25	0.155	0.055
	(0.43)	(0.43)
Age 26-35	0.001	0.001
	(0.01)	(0.01)
Age 36-50	Reference	Reference
	-	-
Age 51-65	0.156	0.057
	(1.17)	(1.18)
Age 65+	-0.026	-0.007
	(0.16)	(0.12)
Muslim	-0.217	-0.078
	(1.57)	(1.57)
HH size = 2	-0.050	-0.019
	(0.24)	(0.25)
HH size = 3	-0.160	-0.060
	(1.05)	(1.05)
HH size = 4	Reference	Reference
	-	-
HH size = 5	0.138	0.050
	(0.90)	(0.89)
HH size = 6	0.226	0.080
	(1.35)	(1.35)
HH size = 7	0.344	0.121
	(1.83)+	(1.88)+
HH size = 8 or more	0.383	0.131
	(1.84)+	(1.84)+
Has Attended School	-0.231	-0.087
	(1.89)+	(1.92)+
Has Chronic Disease	0.709	0.223
	(3.58)**	(3.57)**
Has Had Health Problem	-0.414	-0.162
	(1.74)+	(1.75)+
Health Much Better	0.009	0.004
	(0.05)	(0.07)
Health Better	-0.013	-0.005
	(0.07)	(0.06)
Health Worse	0.072	0.027
	(0.20)	(0.20)

Blood Pressure	0.139	0.052
	(1.12)	(1.15)
Diabetes Test	-0.069	-0.026
	(0.26)	(0.26)
Pregnant	-0.279	-0.097
	(0.72)	(0.73)
Imputed LOG income	-0.037	-0.014
	(0.62)	(0.63)
Imputed Age of first sex	0.019	0.007
	(1.64)	(1.64)
Imputed no Partners	-0.058	-0.022
	(0.81)	(0.81)
Imputed Comdom	0.014	0.005
	(0.10)	(0.10)
Imputed Stigma	-0.082	-0.030
	(1.33)	(1.32)
Imputed Knowledge	-0.082	-0.030
	(1.26)	(1.25)
Extraversion	-0.076	-0.028
	(1.02)	(1.03)
Conscientiousness	-0.015	-0.006
	(0.21)	(0.24)
Neuroticism	0.071	0.026
	(0.96)	(0.97)
Openness	-0.040	-0.015
	(0.52)	(0.51)
Weight	-0.005	-0.002
	(1.09)	(1.11)
Constant	0.387	783
	(0.34)	
Observations	782	
Absolute value of z statistics in parentheses + significant at 10%; * significant at 5%; ** significant at 1%		

Appendix 2: Test participation decision probit model (Household Heads, parsimonious estimation)

	Coefficients	Marginal Effects
Spouse Test	0.865	0.328
	(8.26)**	(8.26)**
Muslim	-0.195	-0.070
	(1.47)	(1.47)
HH size = 2	-0.011	-0.004
	(0.05)	(0.05)
HH size = 3	-0.166	-0.063
	(1.11)	(1.11)
HH size = 4	Reference	Reference
	-	-
HH size = 5	0.134	0.049
	(0.90)	(0.90)
HH size = 6	0.228	0.082
	(1.39)	(1.39)
HH size = 7	0.344	0.120
	(1.90)+	(1.90)+
HH size = 8 or more	0.361	0.125
	(1.82)+	(1.82)+
Has Attended School	-0.187	-0.069
	(1.76)+	(1.76)+
Has Chronic Disease	0.692	0.221
	(3.91)**	(3.91)**
Has Had Health Problem	-0.352	-0.136
	(1.50)	(1.50)
Imputed LOG income	-0.039	-0.014
	(0.68)	(0.68)
Imputed Age of first sex	0.018	0.007
	(1.68)+	(1.68)+
Imputed Knowledge	-0.077	-0.028
	(1.25)	(1.25)
Imputed Stigma	-0.086	-0.032
	(1.45)	(1.45)
Constant	-0.085	788
	(0.12)	
Observations	788	
Absolute value of z statistics in parentheses + significant at 10%; * significant at 5%; ** significant at 1%		

Household Head Probit Model Hit-Miss Table

		Predicted		
		No	Yes	Total
Actual	No	128	162	290
	Yes	75	422	497
	Total	203	584	787

Correctly predicted: 69.9%

Household Head Constant Model Hit-Miss Table

		Predicted		
		No	Yes	Total
Actual	No	0	290	290
	Yes	0	497	497
	Total	0	787	787

Correctly predicted: 63.2%

Appendix 3: Test participation decision probit model (Spouses, First estimation)

	Coefficients	**Marginal Effects**
Head Test	0.852	0.292
	(8.20)**	(8.20)**
Age 18-25	0.536	0.152
	(2.71)**	(2.71)**
Age 26-35	0.030	0.010
	(0.22)	(0.22)
Age 36-50	Reference	Reference
	-	-
Age 51-65	-0.132	-0.044
	(0.90)	(0.90)
Age 65+	-0.317	-0.112
	(1.03)	(1.03)
Muslim	0.057	0.019
	(0.41)	(0.41)
HH size = 2	0.154	0.048
	(0.69)	(0.69)
HH size = 3	-0.179	-0.061
	(1.12)	(1.12)
HH size = 4	Reference	Reference
	-	-
HH size = 5	0.126	0.040
	(0.78)	(0.78)
HH size = 6	0.113	0.036
	(0.64)	(0.64)
HH size = 7	0.002	0.001
	(0.01)	(0.01)
HH size = 8 or more	0.185	0.058
	(0.91)	(0.91)
Has Attended School	-0.252	-0.081
	(1.96)*	(1.96)*
Has Chronic Disease	0.323	0.096
	(1.70)+	(1.70)+
Has Had Health Problem	0.021	0.007
	(0.07)	(0.07)
Health Much Better	-0.050	-0.016
	(0.30)	(0.30)
Health Better	0.031	0.010
	(0.15)	(0.15)

Health Worse	-0.444	-0.162
	(0.92)	(0.92)
Blood Pressure	-0.025	-0.008
	(0.22)	(0.22)
Diabetes Test	-0.174	-0.060
	(0.62)	(0.62)
Pregnant	-0.038	-0.012
	(0.09)	(0.09)
Imputed LOG income	0.028	0.009
	(0.46)	(0.46)
Imputed Age of first sex	0.024	0.008
	(1.52)	(1.52)
Imputed no Partners	-0.069	-0.023
	(0.33)	(0.33)
Imputed Comdom	-0.265	-0.087
	(1.45)	(1.45)
Imputed Stigma	0.033	0.011
	(0.45)	(0.45)
Imputed Knowledge	0.001	0.000
	(0.02)	(0.02)
Extraversion	0.136	0.045
	(1.78)+	(1.78)+
Conscientiousness	-0.099	-0.032
	(1.41)	(1.41)
Neuroticism	-0.062	-0.020
	(0.81)	(0.81)
Openness	0.011	0.004
	(0.14)	(0.14)
Weight	0.008	0.003
	(1.93)+	(1.93)+
Constant	-0.061	798
	(0.05)	
Observations	798	
Absolute value of z statistics in parentheses + significant at 10%; * significant at 5%; ** significant at 1%		

Appendix 4: Test participation decision probit model (Spouses, parsimonious estimation)

	Coefficients	**Marginal Effects**
Head Test	0.856	0.294
	(8.43)**	(8.43)**
Age 18-25	0.457	0.133
	(2.47)*	(2.47)*
Age 26-35	0.012	0.004
	(0.09)	(0.09)
Age 36-50	Reference	Reference
	-	-
Age 51-65	-0.123	-0.041
	(0.88)	(0.88)
Age 65+	-0.404	-0.146
	(1.38)	(1.38)
Has Attended School	-0.228	-0.074
	(1.85)+	(1.85)+
Chronic Disease	0.298	0.090
	(1.62)	(1.62)
Imputed LOG income	0.011	0.004
	(0.18)	(0.18)
Imputed Age of first Sex	0.023	0.008
	(1.50)	(1.50)
Imputed Condom	-0.247	-0.081
	(1.40)	(1.40)
Extraversion	0.123	0.040
	(2.04)*	(2.04)*
Conscientiousness	-0.116	-0.038
	(1.88)+	(1.88)+
Weight	0.008	0.003
	(1.95)+	(1.95)+
Constant	0.075	800
	(0.08)	
Observations	800	
Absolute value of z statistics in parentheses + significant at 10%; * significant at 5%; ** significant at 1%		

Spouse Probit Model Hit-Miss Table

		Predicted		
		No	Yes	Total
Actual	No	89	146	235
	Yes	58	510	568
	Total	147	656	803

Correctly predicted: 74.6%

Spouse Constant Model Hit-Miss Table

		Predicted		
		No	Yes	Total
Actual	No	0	235	235
	Yes	0	568	568
	Total	0	803	803

Correctly predicted: 70.4%

Appendix 5: 3SLS model for household dependence

	(1)	(2)
	Head Test	Spouse Test
Spouse Test	0.516	
	(3.23)**	
Head Test		0.502
		(3.07)**
Common Variables		
HH size = 2	-0.005	0.044
	(0.08)	(0.67)
HH size = 3	-0.044	-0.033
	(0.83)	(0.64)
HH size = 4	Refrence	Reference
	-	-
HH size = 5	0.029	0.022
	(0.56)	(0.46)
HH size = 6	0.065	0.008
	(1.18)	(0.15)
HH size = 7	0.100	-0.037
	(1.64)	(0.62)
HH size = 8 or more	0.095	0.025
	(1.45)	(0.38)
Imputed LOG income	-0.016	0.021
	(0.84)	(1.19)
Household Head Variables		
Age 18-25	-0.083	
	(0.80)	
Age 26-35	-0.021	
	(0.52)	
Age 36-50	Reference	
	-	
Age 51-65	0.035	
	(0.95)	
Age 65+	0.015	
	(0.30)	
Has Attended School	-0.065	
	(1.83)+	
Has Chronic Disease	0.144	
	(3.06)**	
Has Had Health Problem	-0.063	
	(0.97)	
Imputed Age of first sex	0.006	

	(1.81)+	
Imputed Knowledge	-0.016	
	(0.95)	
Imputed Stigma	-0.032	
	(1.84)+	
Sopuse Variables		
Age 18-25		0.137
		(2.84)**
Age 26-35		-0.008
		(0.22)
Age 36-50		Reference
		-
Age 51-65		-0.058
		(1.46)
Age 65+		-0.162
		(1.85)+
Has Attended School		-0.034
		(1.09)
Chronic Disease		0.086
		(1.65)+
Imputed Age of first sex		0.005
		(1.22)
Imputed Condom		-0.043
		(1.21)
Extraversion		0.023
		(1.51)
Conscientiousness		-0.029
		(1.85)+
s_weight		0.002
		(1.67)+
Constant	0.297	0.125
	(1.25)	(0.47)
Observations	782	782
Absolute value of t statistics in parentheses + significant at 10%; * significant at 5%; ** significant at 1%		

Household Head 3SLS Hit-Miss Table

		Predicted		
		No	Yes	Total
Actual	No	141	147	288
	Yes	90	404	494
	Total	231	551	782

Correctly predicted: 69.7%

Household Head Constant Hit-Miss Table

		Predicted		
		No	Yes	Total
Actual	No	0	288	288
	Yes	0	494	494
	Total	0	782	782

Correctly predicted: 63.2%

Spouse 3SLS Hit-Miss Table

		Predicted		
		No	Yes	Total
Actual	No	133	99	232
	Yes	131	419	550
	Total	264	518	782

Correctly predicted: 70.6%

Spouse Constant Hit-Miss Table

		Predicted		
		No	Yes	Total
Actual	No	0	232	232
	Yes	0	550	550
	Total	0	782	782

Correctly predicted: 70.3%

Appendix 6: Logit and Probit Estimation of Individual Participation Decision

	Probit	Logit	Scobit
Age 18-25	0.339	0.572	0.378
	(2.95)**	(2.93)**	(1.50)
Age 26-35	0.040	0.047	0.061
	(0.48)	(0.35)	(0.65)
Age 36-50	References	References	References
	-	-	-
Age 51-65	0.003	-0.004	0.013
	(0.04)	(0.03)	(0.14)
Age 65+	-0.280	-0.461	-0.325
	(2.81)**	(2.80)**	(1.67)
Muslim	-0.072	-0.130	-0.076
	(0.94)	(1.00)	(0.71)
Female	0.461	0.774	0.516
	(5.18)**	(5.15)**	(1.71)
Married or with Partner	-0.050	-0.075	-0.064
	(0.56)	(0.51)	(0.63)
Household Head	0.294	0.498	0.325
	(3.24)**	(3.24)**	(1.50)
HH size = 1	0.083	0.138	0.096
	(0.65)	(0.64)	(0.63)
HH size = 2	0.101	0.155	0.122
	(0.91)	(0.83)	(0.94)
HH size = 3	-0.252	-0.415	-0.293
	(2.73)**	(2.72)**	(1.68)
HH size = 4	References	References	References
	-	-	-
HH size = 5	0.116	0.180	0.139
	(1.20)	(1.12)	(1.18)
HH size = 6	0.078	0.128	0.094
	(0.77)	(0.76)	(0.76)
HH size = 7	-0.018	-0.032	-0.018
	(0.17)	(0.18)	(0.15)
HH size = 8 or more	0.302	0.507	0.341
	(2.53)*	(2.50)*	(1.50)
Has Attended School	-0.250	-0.415	-0.287
	(3.48)**	(3.48)**	(1.76)
Chronic Disease	0.416	0.698	0.466
	(4.23)**	(4.11)**	(1.68)
Health Problem	0.400	0.663	0.453
	(2.38)*	(2.27)*	(1.54)

Blood Pressure	0.148	0.251	0.170
	(2.30)*	(2.31)*	(1.42)
Diabetes Test	-0.171	-0.292	-0.199
	(1.11)	(1.12)	(0.95)
Imputed LOG income	0.008	0.014	0.006
	(0.25)	(0.25)	(0.16)
Imputed Age Sex	0.030	0.049	0.035
	(4.20)**	(4.11)**	(1.93)
Imputed Comdom	-0.152	-0.257	-0.169
	(2.10)*	(2.01)*	(1.38)
Conscientiousness	-0.054	-0.094	-0.057
	(1.83)	(1.93)	(1.05)
Weight	0.005	0.008	0.005
	(2.03)*	(2.05)*	(1.24)
Imputed Knowledge of Sex	-0.025	-0.041	-0.030
	(0.66)	(0.66)	(0.66)
Constant	-0.192	-0.326	-1.809
	(0.37)	(0.37)	(0.50)
Observations	2352	2352	2352
Absolute value of z statistics in parentheses *** significant at 5%; ** significant at 1%**			

Logit Model Hit-Miss Table

		Predicted		
		No	Yes	Total
Actual	No	99	665	764
	Yes	98	1581	1679
	Total	197	2246	2443

Correctly predicted: 68.77%

Constant Model Hit-Miss Table

		Predicted		
		No	Yes	Total
Actual	No	0	759	759
	Yes	0	1679	1679
	Total	0	2438	2438

Correctly predicted: 68.86%

Scobit Model Hit-Miss Table

		Predicted		
		No	Yes	Total
Actual	No	95	669	764
	Yes	91	1588	1679
	Total	186	2257	2443

Correctly predicted: 68.89%

Probit Model Hit-Miss Table

		Predicted		
		No	Yes	Total
Actual	No	98	661	759
	Yes	94	1585	1679
	Total	192	2246	2438

Correctly predicted: 69.03%

Appendix 7: Semi-nonparametric Model Estimation

	participate
Age 18-25	0.328
	(0.80)
Age 26-35	0.055
	(0.49)
Age 36-50	Reference
	-
Age 51-65	0.013
	(0.15)
Age 65+	-0.283
	(0.77)
Muslim	-0.065
	(0.58)
Female	0.449
	(0.81)
Married or with Partner	-0.056
	(0.49)
Household Head	0.282
	(0.80)
HH size = 1	0.084
	(0.52)
HH size = 2	0.108
	(0.63)
HH size = 3	-0.255
	(0.77)
HH size = 4	Reference
	-
HH size = 5	0.122
	(0.69)
HH size = 6	0.081
	(0.56)
HH size = 7	-0.016
	(0.14)
HH size = 8 or more	0.296
	(0.78)
Has Attended School	-0.250
	(0.79)
Chronic Disease	0.405
	(0.81)
Health Problem	0.395
	(0.78)

Blood Pressure	0.147
	(0.76)
Diabetes Test	-0.171
	(0.65)
Imputed LOG income	0.006
	(0.17)
Imputed Age Sex	0.031
	(0.80)
Imputed Comdom	-0.147
	(0.77)
Conscientiousness	-0.049
	(0.74)
Weight	0.005
	(0.75)
Imputed Knowledge	-0.026
	(0.52)
Observations	2352
Absolute value of z statistics in parentheses *** significant at 5%; ** significant at 1%**	

Appendix 8: Model Information Criteria

	AIC	BIC
Model 1	960.2164	1114.058
Model 1 parsimonious	943.2147	1017.906
Model 2	897.5202	1052.03
Model 2 parsimonious	867.6058	935.1903
Probit	2791.197	2946.798
Logit	2791.514	2947.116
Scobit	2792.963	2954.328