On the attitudes toward voluntary HIV antibody counselling and testing in Africa: Revisited

Econometric Game 2010, April 14$^{\text{th}}$

Abstract

In the first paper on the topic, we investigated the relationship between acceptance of voluntary HIV antibody counselling and testing (VCT) and individual's observables, in order to recognise traits that characterise the "type" of persons that are willing to be tested. Two assumptions were implicitly made: that of independence between intra-family outcomes and that of a (fully) parametric specification of the model.

In this follow-up we explore the validity of these assumptions. (i) We test the independence assumption by using the regressive logit of Bonney (1987). (ii) We investigate the parametric assumptions previously made by generalising the logistic distribution to the Burr-10 (Scobit model), and by employing single index semi-parametric model by Klein and Spady (1993). For the latter we compare the prediction success rate to that of the parametric specification.

Our findings are that the head of the household and the spouse do not make their decisions to participate in the VCT independently. Regarding the functional form of the link function we cannot reject the appropriateness of the logistic CDF when testing against a Scobit alternative. When comparing the logit to the semi-parametric approach the conclusions are less clear – the estimated link function does not resemble the logistic CDF, but the semi-parametric classifier does not perform significantly better than the logistic.

Introduction

Voluntary HIV antibody counselling and testing (VCT) is believed to play a critical role in the ongoing struggle against the HIV/AIDS epidemic in Africa. Due to an apparent reluctance to testing we, in the previous paper, investigated which characteristics influenced the individual choice to participate in the blood test. The results suggested that religion was important in explaining the willingness to participate for males and education was important in participating for females. The results, however, was based on individual choices ignoring that the choices are (potentially) interdependent within the household. Several reasons motivate this assumption could potentially be problematic; examples are peers and social interactions. Norms and beliefs are most likely transmitted within the household such that if a strong reluctance to participate is present in one member of the family the choices of the other members reflect this. Gardner and Steinberg (2005) show that risky decision making is related to peers in that peers tend to increase the willingness to take risky decisions. Becker (1974) address' the issue of social interactions among family members. He defines a family with a head as a highly interdependent organization where the head is defined as the member who transfers general purchasing power to all other members because he cares about their welfare. Utility is thus maximized within the family and not on an individual basis. Although Becker focuses on income decisions it is highly likely that other decisions such as willingness to participate in a HIV antibody blood test is dependent on the total utility of the household.

This paper investigates whether the decision to participate in a voluntary anonymous blood test depends on the choices within the household, more specifically whether dependence exists between the head and spouse of the household when deciding whether to participate or not. The estimations are based on an extensive survey data set conducted in an anonymous African country. The data set contains a wide variety of demographic-, socio economic-, personality and medical variables. Dimension reduction on personality and risk attitude variables are conducted using principle component factor analysis from which factor scores are obtained using the regression method. The validity of the factors are tested by Kaiser-Meyer-Olkin (KMO) and Cronbach's Alpha test statistics. The factors used showed good internal validity.[1]

1. Investigation of the independence assumption

1.1. Social interaction and peer effects

The choice to participate in the blood test is likely to depend on the choice of the other members of the household. In the econometric literature this is referred to as 'peer effects' ((Manski, 1993)). This can have important policy implications, if peer effects are strong the government can focus on e.g. persuading the head of the household to take the test.

The identification of peer effects is difficult, and one has to consider many issues. Who influences whom? The spouse or the head, or is group behavior

[1]For further information see previous paper.

		spouse		
		no test	test	total
head	no test	287	296	583
	test	190	795	985
	total	477	1091	1568

Table 1: Test decisions of spouses and heads

just an aggregate of individual behavior? In the literature one distinguishes between three overall effects. 1) Endogenous effects. The tendency to take the test depends on the tendency of the group, thus causing an endogeneity problem. 2) Exogenous (contextual) effects. As shown in case A personal, medical and socio-economic variables influence the decision to take the test, and this is the case for all members of the household. 3) Correlated effects. Unobserved individual characteristics are generally found to be similar among the members of the household, and this creates correlation within the household inducing biased estimates.

In this paper the group is the household, and we chose to focus on the interdependence between the head and the spouse (defined as the first wife). This simplifies the analysis considerably and we can model this by a regressive logistic model.

1.2. Examining dependence

The total sample consists of 5393 observations. After deleting individuals below the age of 18 and respondents diagnosed with HIV/AIDS the sample size reduces to 2752 observations consisting of 1390 households. As we are interested in the interdependence between head and spouse we delete households only consisting of either a head or a spouse reducing the sample size to only 1895 observations distributed on 783 households.

Table (1) shows the number of respondents who choose to take the test or not, dependent on whether their partner takes the test or not. Notice that the number of observations are smaller than the 1895 observations in the data set. This is due to the fact that the data set also consists of other relatives living in the household. As is visible from the table there seem to be a tendency that heads and spouses make the same decision. A simple (initial) test for dependence between their decision is the Pearson's χ^2-testor. Under the null hypothesis of independence between the two categories, the simultaneous probability of e.g. the spouse and the head who choose not to take the test is equal to the product of the marginal

		spouse		
		no test	test	total
head	no test	177	406	583
	test	300	685	985
	total	477	1091	1568

Table 2: Expected test decisions of spouses and heads under the hypothesis of independence.

probabilities. We get a test statistic of 156 which asymptotically is χ^2 distributed with one degree of freedom. Hence, we can clearly reject the null.

Calculating the correlation between spouses' and heads' willingness to participate in the blood test further indicates that there is dependence between their decisions (corr=0.3145). Overall the simple statistics indicates a dependence in the decision making within the household. This dependence should be taken into account when investigating the importance of different individual characteristics on individuals' willingness to participate in the blood testing scheme. Failure to account properly for this dependence gives inconsistent parameter estimates.

1.3. The regressive logistic model

In order to model the potential dependence in the decision of the head and spouse to participate in the voluntary blood test we employ the regressive logistic model of Bonney (1987). Letting $Y = (Y_H, Y_S)$ where $Y_H = 1$ if the head of the household wants to participate in the test and 0 otherwise. Similarly, $Y_S = 1$ and 0 otherwise, denotes the outcome variable for the spouse. Then, letting X denote the matrix of covariates the joint probability can be decomposed into

$$(1) \qquad P(Y_H, Y_S|X) = P(Y_S|Y_H, X)P(Y_H|X)$$

As mentioned above the relevant case to consider for African households, is the influence of the head on the spouse's decision to participate. The other case is straightforward to estimate as well, and would in general imply a different model (Bonney, 1987). We assume that the covariates are exogenous among household members, e.g. the education of the head does not influence the spouse's decision to take the HIV test. Of course one ought to verify this, and correct potential problems. However, this is beyond the purpose of this exercise. Assuming that the probability to take the test follows the logistic distribution we may write the likelihood as

$$(2) \qquad P(Y_H, Y_S|X) = \prod_{i=H,S} \frac{\exp(Y_i\theta_i)}{1 + \exp(\theta_i)}$$

where $\theta_i = \alpha + \beta_i' X_i + \gamma Z_{i,H}$. Bonney suggests to take the following functional form for Z_{iH}, but others could also be considered

$$(3) \qquad Z_{iH} = \begin{cases} 0 & \text{if } i = H \\ 1 & \text{if } i = S \text{ and } Y_H = 1 \\ -1 & \text{if } i = S \text{ and } Y_H = 0 \end{cases}$$

Since Z is measurable wrp Y_H this transformation is in line with (1). We set $Z_{S,H} = -1$ in order to allow for different effects for spouses whose husband does not participate than for their husband (setting $Z_{S,H} = 0$ would impose equal effects).

Hence, the potential dependence between the head's and the spouse's decision may be modeled by estimating the parameters by maximum likelihood and then testing the the null that the decisions are mutually independent, which is the case if $\gamma = 0$. If one rejects this hypothesis it is concluded that spouse's willingness to participate in the blood test is influenced by her husbands choice.

1.4. Results

After having removed insignificant variables the results of the estimation of the regressive logit are as given in Table 2. The first thing we notice is that the spouses are much more willing to take the test if the head of the household is(`z`). Furthermore one is *more likely* to take the test if

- the household size is large (`hhsize`)
- household consumption is high (`log_cons`)
- educated people are more positive towards the test (`primary, secondary`)
- persons are in a generally bad health condition (`chro_dis, visit_doc`)

Table 2

Average marginal effects Number of obs = 1556
Model VCE : OIM

Expression : Pr(test), predict()
dy/dx w.r.t. : hhsize log_cons female primary secondary chro_dis visit_doc open per_risk_beh z

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
hhsize	.0222581	.0068059	3.27	0.001	.0089187	.0355975
log_cons	.0282508	.0205551	1.37	0.169	-.0120365	.0685381
female	.0701392	.0233281	3.01	0.003	.0244169	.1158615
primary	.10073	.0298648	3.37	0.001	.042196	.1592639
secondary	.1441806	.0298178	4.84	0.000	.0857388	.2026223
chro_dis	.1380661	.0417878	3.30	0.001	.0561635	.2199687
visit_doc	.0396575	.0241618	1.64	0.101	-.0076988	.0870138
open	-.0222769	.01115	-2.00	0.046	-.0441305	-.0004234
per_risk_beh	-.0319481	.0118532	-2.70	0.007	-.0551799	-.0087162
z	.1381297	.0157098	8.79	0.000	.1073391	.1689203

Logistic model for test

Classified	True D	True ~D	Total
+	924	378	1302
-	117	137	254
Total	1041	515	1556

Classified + if predicted Pr(D) >= .5
True D defined as test != 0

Sensitivity	Pr(+| D)	88.76%
Specificity	Pr(-|~D)	26.60%
Positive predictive value	Pr(D| +)	70.97%
Negative predictive value	Pr(~D| -)	53.94%
False + rate for true ~D	Pr(+|~D)	73.40%
False - rate for true D	Pr(-| D)	11.24%
False + rate for classified +	Pr(~D| +)	29.03%
False - rate for classified -	Pr(D| -)	46.06%
Correctly classified		68.19%

Model with a constant only

Correctly classified	65.33%

These marginal effects are all as expected. What is also worth noticing are which variables turned out insignificant (and have hence been removed from the model)

- age
- of the big four only openness (`open`) is relevant

We also notice that the regressive logit is better at classifying the individuals attitude towards the test than a simple model with a constant only. In particular the fraction of correctly classified are 68.19% as opposed to 65.33% with model including the constant only.

From the analysis in this section we conclude that the decion to take the test is highly interdependent between the head of the household and the spouse. In particular, the Pearson test as well as the regressive logit rejected this hypothesis.

2. Investigation of the parametric assumptions

In this section we return to the setting of Case A where independence among the household members is assumed. We now investigate two alternatives to the logit. (i) The Scobit which uses the Burr-10 CDF instead of the logistic CDF. The logistic CDF is nested in the Burr-10 which allows us to test for the appropriateness of the logit. (ii) The fixed form of the link function, for example, by logistic CDF is rarely justified by the context of the observed data but is often motivated by numerical convenience and by reference to “standard practice”. Hence, the single index semi-parametric model of (Klein and Spady, 1993), which leaves the link function unspecified and instead lets the data speak, is implemented. The estimate of the link function can then be compared to the logistic CDF and the Burr-10.

2.1. Possible skewness: the Scobit model

In the logit it is implicitly assumed that individuals with a probability of $p = 0.5$ of choosing either of two alternatives are most sensitive to changes in the independent variables. This assumption is imposed by the estimation technique because both the logistic function is symmetric about zero. However, the assumption can be tested by estimating the Scobit model of Nagler (1994). Here,

$$P(Y = 1|X) = \frac{1}{(1 + \exp(-x'\beta))^{\alpha}} \tag{4}$$

where $\alpha > 0$ is a parameter to be estimated which allows for the marginal effects to achieve there maximum at other points than $p = 0.5$. In particular, the appropriateness of the logit may be investigated by testing whether $\alpha = 1$.

2.2. *Results from the Scobit*

We estimate the model including only the variables which were found to be relevant in case A. The results from the estimated Scobit model are given below. What is of interest to us is that we cannot reject that $\alpha = 1$. However, it should be mentioned that the standard error of α is very large giving the test low power. In this regard, the logit is a suitable model. The classification table of the corresponding logit will be shown and compared to the semi-parametric results in the next section.

Scobit
Skewed logistic regression — Number of obs = 2679
Zero outcomes = 855
Log likelihood = -1602.768 — Nonzero outcomes = 1824

test	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
hhsize	-.1257967	.0827267	-1.52	0.128	-.287938	.0363447
hhsizesq	.0154057	.0092274	1.67	0.095	-.0026798	.0334911
log_cons	.1683315	.0910815	1.85	0.065	-.0101849	.346848
female	.2684907	.1331135	2.02	0.044	.0075929	.5293884
primary	.3426943	.1587079	2.16	0.031	.0316325	.6537562
secondary	.4578922	.1945646	2.35	0.019	.0765527	.8392318
chro_dis	.4553358	.2242274	2.03	0.042	.0158581	.8948134
visit_doc	.2843336	.1370697	2.07	0.038	.0156819	.5529853
open	-.1265655	.0570689	-2.22	0.027	-.2384185	-.0147125
per_risk_beh	-.1177675	.0664249	-1.77	0.076	-.247958	.0124229
_cons	-2.968875	1.746238	-1.70	0.089	-6.391438	.4536876
/lnalpha	1.031083	2.100098	0.49	0.623	-3.085034	5.1472
alpha	2.804101	5.888888			.0457285	171.9494

Likelihood-ratio test of alpha=1: chi2(1) = 0.51 Prob > chi2 = 0.4740

3. A semi-parametric approach to the binary response estimation

In this section we employ the efficient single index model of Klein and Spady (1993) to fit the model

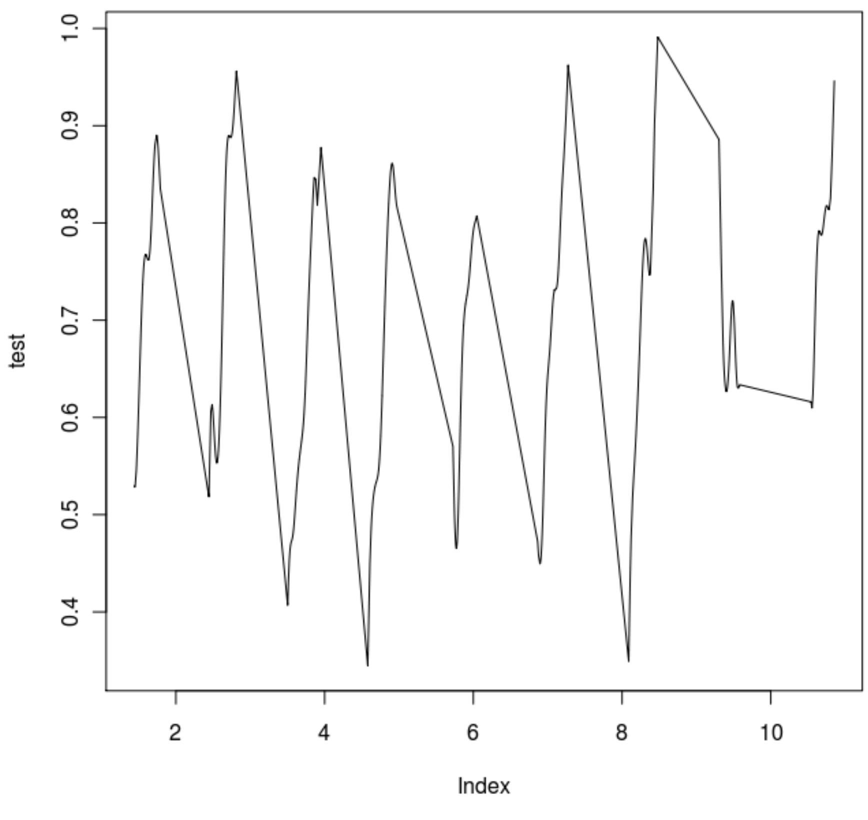
$$(5) \qquad Y_i = f(\boldsymbol{x}_i'\boldsymbol{\beta}),$$

where $f(\cdot)$ is left unspecified. In the logit model, $f(\cdot)$ is the logistic function. We include the parsimonious model specification with only the variables significant in the parametric models previously estimated. This implicitly assumes that those variables that help explain the index in the logit/scobit model are also those that explain the unknown link function. This should be investigated more thoroughly; yet, since time-constrained we abstain from this.

Under the assumptions discussed below the Klein-Spady estimator is $\sqrt{N}$ consistent, asymptotically normal, and achieves the lower efficiency bound for semi-parametric estimators (Theorem 2-4 in (Klein and Spady, 1993)) The estimator requires (among other things) that $\{Y_i, X_i\}$ is a random sample (assumption C1 inKlein and Spady (1993)), and thus the violated indepdce assumption poses a problem. This should make us cautious in drawing to drastic conclusions from this model. Under the stated assumptions in problem text we proceed estimating this model. It should be mentioned that other assumptions are required in order to establish the asymptotics (Assumptions C2-C9). Some of these are standard regularity conditions in semi-parametric estimation, while others are harder to verify, and beyond the scope of this paper. The output from the estimation is presented below along with the classification table for the logit (Notice that the tables are rotated due to the use of different software). The predictive ability of the two procedures is almost identical – the (Klein and Spady, 1993) having a slight edge. Both models have the problem that they classify too many negatives as positives. We also plot the predicted values of the (Klein and Spady, 1993) model against the index $x'\beta$. It is seen that the graph by no means has the sigmoidal shape of the logistic CDF. This indicates that the logit might not be appropriate. More formally, one could test the logit (known link finction) against a semi-parametric alternative (unknown link function) using the test of (H "ardle et al., 1997). Alternatively it could be investigated whether a multiple index model would be more appropriate than the single index.

```
> source("kleinspady.R")

Single Index Model
Regression Data: 2679 training points

Bandwidth: 0.03460392
Kernel Regression Estimator: Local-Constant

Confusion Matrix
      Predicted
Actual    0    1
     0   55  800
     1   36 1788

Overall Correct Classification Ratio:  0.6878267
Correct Classification Ratio By Outcome:
         0          1
0.06323185 0.98026316

McFadden-Puig-Kerschner performance measure:  0.5984062


        ***LOGIT CLASSIFICATION TABLE***

              -------- True --------
Classified |         D            ~D  |      Total
-----------+--------------------------+-----------
     +     |      1735           764  |       2499
     -     |        89            91  |        180
-----------+--------------------------+-----------
   Total   |      1824           855  |       2679

Classified + if predicted Pr(D) >= .5
True D defined as test != 0
--------------------------------------------------
Sensitivity                     Pr( +| D)   95.12%
Specificity                     Pr( -|~D)   10.64%
```

```
Positive predictive value       Pr( D| +)   69.43%
Negative predictive value       Pr(~D| -)   50.56%
--------------------------------------------------
False + rate for true ~D        Pr( +|~D)   89.36%
False - rate for true D         Pr( -| D)    4.88%
False + rate for classified +   Pr(~D| +)   30.57%
False - rate for classified -   Pr( D| -)   49.44%
--------------------------------------------------
Correctly classified                        68.16%
--------------------------------------------------
```



4. Conclusion

In this paper we investigated two questions regarding the validity of the methods employed in Case A for describing the attitudes towards the VCT. Firstly, we looked into whether the attitude towards the test is independent across the head of the household and the spouse. This was investigated by means of a Pearson test and the regressive logit. Both procedures clearly rejected that the head of the household and the spouse make their decisions independently of each other. Secondly, we investigated the appropriateness of the logistic CDF as a link function. When testing this against a Scobit alternative, this could not be rejected. However, this does not mean that the logit is the right link function. In particular the estimated link function using the (Klein and Spady, 1993) procedure did not resemble a sigmoidal function. The classification rate of the semi-parametric procedure was not markedly better than that of the logit. Hence, there exist two problems which seem hard to solve simultaneously – dependence of decisions and finding an appropriate unknown link function. Our proposed direction for further research is two investigate multiple index models for known and unknown link functions.

References

Barrick, M., Mount, M., 1991. The Big Five personality dimensions and job performance: A meta-analysis. Personnel psychology 44 (1), 1–26.

Becker, G., 1974. A theory of social interactions. Journal of Political Economy 82 (6).

Bonney, G., 1987. Logistic regression for dependent binary observations. Biometrics 43 (4), 951–973.

Cameron, A., Trivedi, P., 2005. Microeconometrics: methods and applications. Cambridge Univ Pr.

Gardner, M., Steinberg, L., 2005. Peer Influence on Risk Taking, Risk Preference, and Risky Decision Making in Adolescence and Adulthood: An Experimental Study. Developmental Psychology 41 (4), 625–635.

H
"ardle, W., Spokoiny, V., Sperlich, S., 1997. Semiparametric single index versus fixed link function modelling. The Annals of Statistics 25 (1), 212–243.

Johnson, R., Wichern, D., 2007. Applied multivariate statistical analysis. Prentice Hall Englewood Cliffs, NJ.

Kawaguchi, D., 2004. Peer effects on substance use among American teenagers. Journal of Population Economics 17 (2), 351–367.

Klein, R., Spady, R., 1993. An efficient semiparametric estimator for binary response models. Econometrica: Journal of the Econometric Society 61 (2), 387–421.

Manski, C., 1993. Identification of endogenous social effects: The reflection problem. Review of Economics Studis 60 (3), 531–536.

Mount, M., Barrick, M., 1998. Five reasons why the" Big Five" article has been frequently cited. Personnel Psychology 51 (4), 849–858.

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

Schmitt, D. P., 2004. The Big Five related to risky sexual behaviour across 10 world regions: Differential personality associations of sexual promiscuity and relationship infidelity. European Journal of Personality 18 (4), 301–319.

Van Buuren, S., Brand, J., Groothuis-Oudshoorn, C., Rubin, D., 2006. Fully conditional specification in multivariate imputation. Journal of Statistical Computation and Simulation 76 (12), 1049–1064.