UNIVERSITEIT VAN AMSTERDAM

Amsterdam School of Economics

Econometric Game 2010

Case B (14 April 2010): Modelling the willingness to take part in a voluntary HIV test

Motivation and background

The background and dataset are the same as for Case A. Today we will investigate dependence between decisions of family members (in particular head and spouse) and look at alternative functional specifications of the model.

Research Questions

Up until now we have implicitly assumed that the choices of individuals to take part in the blood test are made independently across the sample. Of course this is a simplifying assumption, since the choices within households may be strongly dependent on each other.

1. Investigate whether the decisions to participate in the blood test of the head and spouse of the family are dependent given the values of the explanatory variables. To simplify matters, you may analyse this question in the single index Probit/Logit framework. Explicitly state the model and statistical assumptions that you have used.

If you have time left, please consider the second question.

For the second question, we return to the models of Case A assuming independence among the household members for simplicity. However, we now explicitly consider parametric and semi-parametric specifications to model $P(Y=1|X=x)=f(x)$, where $Y=1$ denotes the event that one is tested. There are various ways to specify $f(x)$:

- Parametric single index model: $f(x)=F_\theta(x'\beta)$, where θ and β are finite dimensional parameters. For instance, if $F_\theta(\cdot)$ denotes the Normal/Logit cumulative distribution function (CDF), we obtain the Probit/Logit model. However, you can also specify an asymmetric CDF such as the Burr-10 CDF leading to a Scobit model; see Nagler (1994).
- Semi-parametric single index model: $f(x)=F(x'\beta)$, where β is a finite dimensional parameter and F is a CDF that is to be estimated non-parametrically.

Given sufficient data, a possible advantage of semi-parametric models is that they provide additional model flexibility. If the assumed parametric specification for F deviates substantially from the "true" CDF, one might expect the semi-parametric specification to outperform a parametric specification.

2a. To start with, investigate whether skewness plays a role by using a parametric single index model.

2b. Estimate a semi-parametric model for the choice of individuals to take part in the blood test or not. Carefully report and motivate the choices made in the specification and the estimation of the model. Compare the results obtained by your semi-parametric model to those obtained by your best parametric model. [Note: there are many packages and libraries available for semi-parametric model estimation. For the freely available *R* programming language, for instance, one might consider the function `gam` from `library(gam)`, `npindex` from `library(np)` or `locpoly` from `library(KernSmooth)`.]

References

Fox, J. (2005), Introduction to Nonparametric Regression, Lecture notes, ESRC Oxford Spring School: http://socserv.socsci.mcmaster.ca/jfox/Courses/Oxford-2005/slides-handout.pdf.

Greene, W.H. (1998), Gender Economics Courses in Liberal Arts Colleges: Further Results, *The Journal of Economic Education*, Vol. 29, No. 4, 291-300.

Hayfield, T. and J. S. P. Racine (2008). Nonparametric econometrics: The np package, *Journal of Statistical Software*, Vol. 27, 1-32: http://www.jstatsoft.org/v27/i05/paper.

Klein, R. W. and R. H. Spady (1993), An efficient semiparametric estimator for binary response models, *Econometrica*, 61, 387-421.

Maddala, G. S. (1983), *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge, U.K.: Cambridge University Press (Section 5.7).

Nagler, J. (1994), Scobit: An Alternative Estimator to Logit and Probit, *American Journal of Political Science*, Vol. 38, No. 1, 230-255.

Prentice, R.L. (1976), A Generalization of the Probit and Logit Methods for Dose Response Curves, *Biometrics*, Vol. 32, No. 4, 761-768.

Sherman, R.P. (1993), The Limiting Distribution of the Maximum Rank Correlation Estimator, *Econometrica*, Vol. 61, No. 1, 123-137.