

# Econometric Game 2011

## Case B

### Instrumental Variables Estimation for Binary Outcomes

#### Background

We will initially be answering the same question about the effect of maternal alcohol intake during pregnancy on the child's cognitive outcomes, but we now have as the outcome a binary indicator, whether the child attained level 5 in their key-stage 2 test scores for science, denoted *KS2\_lev5*.

We will also try to determine the causal effect of BMI (Body Mass Index, defined as weight in kg divided by height in metres squared) or of being overweight on a binary indicator whether the child has peer problems at the age of almost 10 years old. There are two genetic markers that have been shown to be associated with adiposity, or fat mass. These are called *FTO* and *MC4R*.

#### Research Question

1. Establish the causal effects of maternal alcohol consumption during pregnancy on the binary educational outcome variable of the child.
2. Establish the causal effects of the child's BMI and the binary indicator of being overweight on whether the child has peer problems.

For both cases, use genetic markers as IVs, addressing the issues of identification and estimation of models for binary outcomes.

#### Data

The data are from the same sample of the ALSPAC data set. The peer problem indicator is derived from a set of questions about the child answered by the mother, e.g. "Child is rather solitary, tends to play alone" (see codebook for further details).

#### Modelling Issues

A review is given in Clarke and Windmeijer (2010). In IV models for binary outcomes, the causal effect is not point identified under minimal assumptions on the data generating process, unlike in the linear model for continuous outcomes, and only bounds can be derived on the causal parameters for these cases. Identification can of course be achieved by increasing the number of modelling assumptions, with at the extreme end of the scale a full likelihood approach. With a continuous endogenous regressor, like BMI, the control function approach has frequently been used.

Within a potential outcomes framework, and allowing for heterogeneous treatment effects, Imbens and Angrist (1994) introduce the notion of *monotonicity* and Local Average Treatment Effects (LATE), see also Angrist, Imbens and Rubin (1996). A linear IV would then estimate a LATE, but also consider estimating a Local Risk Ratio (LRR) from an exponential model as in Mullahy (1997). For the latter estimation you could use the *gmm* function or command in various statistical packages.

When multiple and/or multivalued instruments are available, assess the various ways of utilising this information. Also, within the LATE framework, use the results of Imbens and Angrist (1994) to assess what effects are being estimated in these cases.

#### References

Angrist, J.D., Imbens, G.W. and Rubin, D.B. (1996), Identification of causal effects using instrumental variables, *Journal of the American Statistical Association* 91, 444-455.

Clarke, P. and Windmeijer, F. (2010), Instrumental variable estimators for binary outcomes, CMPO Working Paper 10/239, Centre for Market and Public Organisation, University of Bristol, UK.

Imbens, G.W. and Angrist, J. (1994), Identification and estimation of local average treatment effects, *Econometrica* 62, 467-476.

Mullahy J., 1997, Instrumental variable estimation of Poisson regression models: application to models of cigarette smoking behavior, *Review of Economics and Statistics* 79, 586-593.