

Econometric Game 2011

Case A

Identifying and estimating the causal effect of maternal alcohol intake during pregnancy on cognitive outcomes of the child, using a genetic marker as instrument.

Background

There has been a long and ongoing debate whether alcohol consumption by the mother during pregnancy has a beneficial or detrimental effect on the development of the foetus. The ALSPAC (Avon Longitudinal Study of Parents and Children) data set contains information about the mother's alcohol intake during pregnancy and the children's school test scores at age 14. A simple regression analysis may not identify the causal effect of alcohol intake due to unobserved factors that affect both maternal alcohol intake and the human capital formation of the child.

The ALSPAC data set is unique, however, in the sense that we have information for the mother on the presence of a genetic marker for alcohol consumption: *ADH1B*. This genetic marker has been robustly shown to be associated with alcohol consumption and may therefore be a good candidate to use as an instrumental variable (IV) to identify/estimate a causal effect of maternal alcohol consumption on the child's cognitive ability by means of instrumental variables estimation techniques. "Mendelian Randomisation" is the term used to describe the random allocation of genes from parents to offspring and forms the rationale for using genetic markers as instruments.

Research Question

After critically assessing the use of genetic markers as instruments to identify causal effects, estimate the effect of maternal alcohol consumption on the child's schooling outcomes, as summarised in the variable *KS2*, by IV estimation techniques, using the genetic marker associated with alcohol intake as an IV. Interpret your results.

Data

The data are a 90% random sample from the ALSPAC data set. They contain binary indicators of maternal alcohol intake at three different time points, information on the number of units consumed, and the type of drink (wine or beer). The child's test scores *KS2* are standardised to have a mean of 100 and standard deviation of 10. The genetic marker *ADH1B* is a binary variable indicating the presence of at least one rare allele. The data further contain information on some observable child and family characteristics.

Modelling Issues

The genetic marker may well be a weak predictor of the mother's alcohol consumption during

pregnancy. Discuss in detail the problem of IV estimation in the presence of weak instruments and assess whether this problem occurs in your analysis. Do this for the standard Two-Stage Least Squares estimator and the Fuller(1) estimator and discuss the relative merits of these two estimators.

Investigate separately the effects of the binary indicators of alcohol consumption; the number of units consumed; and whether consumption is wine or beer. Discuss for every case the modelling/identification issues involved.

References

von Hinke Kessler Scholder, S. et al. (2010). "Genetic Markers as Instrumental Variables: An Application to Child Fat Mass and Academic Achievement." *CMPO Working Paper 10/229*

Lawlor, D. et al. (2008). "Mendelian Randomization: Using Genes as Instruments for Making Causal Inferences in Epidemiology." *Statistics in Medicine* 27: 1133-63.

Stock, J.H., Wright, J.H. and M. Yogo (2002). "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments." *Journal of Business & Economics Statistics* 20, 518-529.

Stock, J.H. and M. Yogo (2005). "Testing for Weak Instruments in Linear IV Regression", in D.W.K. Andrews and J.H Stock (eds.) *Identification and Inference for Econometric Models*, Cambridge University Press, New York.