The impossibility of using highest achieved level of education to estimate the probability of passing a transition unconditional on having passed previous transitions*

Maarten L. Buis

m.buis@fsw.vu.nl

June 3, 2007

A common model for estimating the inequality of educational opportunity for children of different social background assumes that each level of education can be reached through one and only one path through the education system. Because of this assumption one only need the highest achieved level of education to know which transitions a person passed or failed, and at which transitions the individual was no longer at risk of passing. In this model one estimates for each transition the effect of social background on passing the transition given that one is at risk. Many names exist for this model: sequential response model (Maddala, 1983), continuation ratio logit (Agresti, 2002), model for nested dichotomies (Fox, 1997), and the Mare model (Shavit and Blossfeld, 1993). The resulting estimates are conditional on being at risk. Sometimes one wants the unconditional effects. In many models this can be achieved by controlling for all observed and unobserved variables. This is obviously a hard task, but some models exist that get close (Cameron and Heckman, 1998; Holm, 2007). The purpose of this note is to show that even if one could estimate such models one would still not get the effects of social background *un*conditional on being at risk.

This can be seen in a stylized model. The educational system in this model consists of two transitions. An individual has a propensity of passing transition k (y_k^*) which depends

*This note is the result of a discussion with Anders Holm and Mike Hout at the RC28 spring meeting in Brno on 26 May 2007

on observed (x_k) and unobserved (e_k) variables:

$$
\begin{aligned}
y_1^* &= \beta_1 x_1 + e_1 \\
y_2^* &= \beta_2 x_2 + e_2
\end{aligned}
$$

The data we have consists of whether or not someone passed the levels of education (y_k) and the xs. Someone passes a level if his propensity is larger than zero. Also, someone can only pass transition 2 if he has already passed transition 1. In other words we don't know whether a person who hasn't passed transition 1 would pass transition 2 if he where to try.

$$
\begin{aligned}
y_1 &= \begin{cases} 1 & \text{if } \beta_1 x_1 + e_1 \geq 0 \\ 0 & \text{if } \beta_1 x_2 + e_1 < 0 \end{cases} \\
y_2 &= \begin{cases} 1 & \text{if } \beta_2 x_1 + e_2 \geq 0 \ \& \ y_1 = 1 \\ 0 & \text{if } \beta_2 x_2 + e_2 < 0 \ \& \ y_1 = 1 \\ \text{NA} & \text{if } y_1 = 0 \end{cases}
\end{aligned}
$$

The aim is to estimate the βs using this data. Say we are interested in the effect of family SES, that that variable is the only observed variable, and that the effect of SES is 1 and the constant is -1. The problems occur in the second transition (and any subsequent transitions if they exist), so from now on only these are shown. The data would be generated in the following way:

$$
y_2 = \begin{cases} 1 & \text{if } -1 + 1SES + e_2 \geq 0 \ \& \ y_1 = 1 \\ 0 & \text{if } -1 + 1SES + e_2 < 0 \ \& \ y_1 = 1 \\ \text{NA} & \text{if } y_1 = 0 \end{cases}
$$

Now assume that those who have passed transition 1 actually learned something at school that will help them passing transition 2. Say that the effect of having passed transition 1 is 1 and that the constant for those that haven't passed is -2. Then the data would be generated in the following way:

$$
\begin{aligned}
y_2 &= \begin{cases} 1 & \text{if } -2 + 1y_1 + 1SES + e_2 \geq 0 \text{ \& } y_1 = 1 \\ 0 & \text{if } -2 + 1y_1 + 1SES + e_2 < 0 \text{ \& } y_1 = 1 \\ \text{NA} & \text{if } y_1 = 0 \end{cases} \\
&= \begin{cases} 1 & \text{if } -1 + 1SES + e_2 \geq 0 \text{ \& } y_1 = 1 \\ 0 & \text{if } -1 + 1SES + e_2 < 0 \text{ \& } y_1 = 1 \\ \text{NA} & \text{if } y_1 = 0 \end{cases}
\end{aligned}
$$

Now assume that the effect of SES differs between those kids that have passed level 1 and those that have not. Assume that the interaction effect is .5 and that the effect of SES for those who haven't passed level 1 is .5. Then the data would be generated in the following way:

$$
\begin{aligned}
y_2 &= \begin{cases} 1 & \text{if } -2 + 1y_1 + .5SES + .5SESy_1 + e_2 \geq 0 \text{ \& } y_1 = 1 \\ 0 & \text{if } -2 + 1y_1 + .5SES + .5SESy_1 + e_2 < 0 \text{ \& } y_1 = 1 \\ \text{NA} & \text{if } y_1 = 0 \end{cases} \\
&= \begin{cases} 1 & \text{if } -1 + 1SES + e_2 \geq 0 \text{ \& } y_1 = 1 \\ 0 & \text{if } -1 + 1SES + e_2 < 0 \text{ \& } y_1 = 1 \\ \text{NA} & \text{if } y_1 = 0 \end{cases}
\end{aligned}
$$

Notice that in all three scenarios the data is generated in exactly the same way. Furthermore we can create an infinite number of scenarios that will also lead to exactly the same data generation mechanism, by just varying the constant, the main effect of passing transition 1, the interaction effect, and the effect of SES. Claiming that once one controls for e_1 and e_2 one would get the effects on the probability of passing transition 2 *un*conditional on having passed the previous transition is equivalent to saying that students have learned nothing at transition 1 that might help them to pass transition 2. Once one wants to relax this assumption the effects on the unconditional probability become unidentified, even if we could correctly control for e_1 and e_2. However, if we could correctly control for e_1 and e_2 (a big if), the effects on the conditional probability are identified.

References

Agresti, Alan. 2002. *Categorical Data Analysis.* 2nd ed. Hoboken, NJ: Wiley-Interscience.

Cameron, Stephen V. and James J. Heckman. 1998. "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males." *The Journal of Political Economy* 106(2):262–333.

Fox, John. 1997. *Applied Regression Analysis, Linear Models, and Related Methods.* Thousand Oaks: Sage.

Holm, Anders. 2007. "The effects of social background on educational progression; the importance of unobserved heterogeneity." presented at the RC28 in Brno on 26 May 2007.

Maddala, G.S. 1983. *Limited Dependent and Qualitative Variables in Econometrics.* Cambridge: Cambridge University Press.

Shavit, Yossi and Hans-Peter Blossfeld. 1993. *Persistent Inequality: Changing Educational Attainment in Thirteen Countries.* Boulder: Westview Press.