Appendix for "Estimation of Educational Borrowing Constraints using Returns to Schooling"

by Stephen Cameron and Christopher Taber

Downward Bias of Discount Rate with Foregone Earnings

To demonstrate the intuition as to how discount rate bias can actually bias the estimate downward, we follow the methodology of Imbens and Angrist (1994). Suppose schooling is binary with S equal to either zero or one (loosely college or high school) and that local labor market conditions are binary with R equal to zero or one. We consider $R=1$ to represent a period of temporary bad labor market conditions (i.e. a recession). Using notation similar to Imbens and Angrist, we let $d(R)$ be an individual specific dummy variable equal to 1 if this individual would choose schooling level 1 in labor market condition R. That is for each student we define $d(1)$ to indicate whether they would attend college during poor local labor market conditions, and $d(0)$ indicate whether they would attend during a good labor market.

Let W_S denote the wage the student would receive if he chooses schooling level S, and W denote the unconditional wage. Assume that R is a legitimate instrument so that it is uncorrelated with $(W_1, W_0, d(1), d(o))$. The Wald estimate takes the form,

$$
\begin{aligned}
&\frac{E\left(W \mid R=1\right)-E\left(W \mid R=0\right)}{\Pr\left(S=1 \mid R=1\right)-\Pr\left(S=1 \mid R=0\right)} \\
=&\frac{E(W_1 \mid R=1, d(1)=1)\Pr(d(1)=1)+E(W_0 \mid R=1, d(1)=0)\Pr(d(1)=0)}{\Pr(d(1)=1)-\Pr(d(0)=1)} \\
&-\frac{E(W_1 \mid R=0, d(0)=1)\Pr(d(0)=1)+E(W_0 \mid R=0, d(0)=0)\Pr(d(0)=0)}{\Pr(d(1)=1)-\Pr(d(0)=1)} \\
=&\frac{E(W_1 \mid d(1)=d(0)=1)\Pr(d(1)=d(0)=1)+E(W_1 \mid d(1)>d(0))\Pr(d(1)>d(0))}{\Pr(d(1)=1)-\Pr(d(0)=1)} \\
&+\frac{E(W_0 \mid d(1)<d(0))\Pr(d(1)<d(0))+E(W_0 \mid d(1)=d(0)=0)\Pr(d(1)=d(0)=0)}{\Pr(d(1)=1)-\Pr(d(0)=1)} \\
&-\frac{E(W_1 \mid d(1)=d(0)=1)\Pr(d(1)=d(0)=1)+E(W_1 \mid d(1)<d(0))\Pr(d(1)<d(0))}{\Pr(d(1)=1)-\Pr(d(0)=1)} \\
&-\frac{E(W_0 \mid d(1)>d(0))\Pr(d(1)>d\left(0\right))+E(W_0 \mid d(1)=d(0)=0)\Pr(d(1)=d(0)=0)}{\Pr(d(1)=1)-\Pr(d(0)=1)} \\
=&\frac{E\left(W_1-W_0 \mid d(1)>d(0)\right)\Pr(d(1)>d(0))-E\left(W_1-W_0 \mid d(1)<d(0)\right)\Pr\left(d(1)<d(0)\right)}{\Pr(d\left(1\right)>d\left(0\right))-\Pr(d\left(1\right)<d\left(0\right))}.
\end{aligned}
$$

Given that we find that enrollment increases in bad times, Imbens and Angrist's (1994) monotonicity assumption would imply that people are only induced to increase schooling during a recession. That is there are no individuals who would attend school during a boom, but not during a recession ($\Pr(d(1) < d(0)) = 0$). Under these conditions, the Wald estimate would be equal to $E(W_1 - W_0 \mid d(1) = 1, d(0) = 0)$. The argument in the text suggests that this parameter is not likely to be strongly influenced by "discount rate bias."

Now consider relaxing the monotonicity assumption. Suppose that there are some students from borrowing constrained families that can not afford to send their families to schooling during a recession but would send their children to college during a boom. In this case monotonicity would be violated because for these students $(d(1) = 0, d(0) = 1)$.Thus $\Pr(d(1) < d(0))$ will be positive. To be consistent with the observation in the data that schooling rises during bad times, it must be the case that $\Pr(d(1) = 1, d(0) = 0) > \Pr(d(1) = 0, d(0) = 1)$. Under the discount rate bias hypothesis, students from these borrowing constrained families have high marginal returns to college. This means that $E(W_1 - W_0 \mid d(1) < d(0))$,would be large. Notice however that this term enters the expression for the Wald estimate negatively. Thus in this case discount rate bias will bias this estimate *downward.*

Identification of Structural Model

Nonlinear models that impose linear index assumptions can often avoid the use of exclusion restrictions, so to focus on the variation arising from the exclusion restriction we consider nonparametric identification of the model Specifically we consider identification of the model,

$$\begin{aligned} S &= 1(\log(e^{g_1(X)+\varepsilon_1} + g_2(Z_1)) + g_3(X) + g_4(Z_2) + \varepsilon_2 > 0) \\ \log(w_{1t}) &= \gamma_S + g_1(X) + g_t(Z_{\ell t}) + E_t\beta_E + E_t^2\beta_{E^2} + u_1 \\ \log(w_{0t}) &= \gamma_S + g_1(X) + g_t(Z_{\ell t}) + E_t\beta_E + E_t^2\beta_{E^2} + u_0, \end{aligned}$$

where Z_2 represents the exclusion restriction from local labor market conditions, Z_1 represents the exclusion restrictions that represent direct costs of schooling, $Z_{\ell t}$ represents local labor market variables at time t, and X represents other values regressors that influence wages and the tastes for schooling.[1]

Before showing identification, there are some normalizations that must be imposed. First notice that we can always normalize $g_3(0) = g_4(0) = 0$ by modifying the location of the error term ε_{2i}. For g_2 we need a somewhat different normalization. As long as $g_2(0) \neq 0$,we can normalize $g_2(0) = 1$. To see why the scale of g_2 needs to be normalized notice that for any τ we can multiply g_2 by e^τ,add τ to ε_{1i},and subtract τ rom ε_{2i} without changing the expression. Under this assumption and standard assumptions about support conditions, the model is identified.

[1]The one type of variation we have not included in this specification is the local labor market variables that influence college choice. For this exercise we can fix them to a constant and only consider variation from the variables that influence local labor markets during college. We are assuming that Z_2 help predict local labor market levels $Z_{\ell t}$ during college.

Theorem 1 *Normalize* $g_3(0) = g_4(0) = 0$ *and* $g_2(0) = 1$. *Assume that (a) the support of* X *does not depend on* (Z_1, Z_2), *(b)* $(g_2(Z_1), g_4(Z_2))$*has support* $\Re^2$,*the error terms* $(\varepsilon_1, \varepsilon_2, u_1, u_2)$ *is independent of the observables* $(X, Z_2, Z_2, Z_{\ell t})$. *Then* $g_1, g_2, g_3,$*and* g_4 *and the joint distribution of* $(\varepsilon_{1i}, \varepsilon_{2i}, u_{1i})$ *are identified.*

Proof: First notice that we can use a standard identification at infinite argument (see e.g. Heckman, 1990) to identify g_1. That is send $g_4(Z_2) \to \infty$,and we can identify the form of the wage equation.

Now suppose the model is not identified so that there are two distinct models that explain the data, the true model

$$1(\log(e^{g_1(X)+\varepsilon_1} + g_2(Z_1)) + g_3(X) + g_4(Z_2) + \varepsilon_2 > 0)$$

and an alternative model,

$$1(\log(e^{g_1(X)+\varepsilon_1^*} + g_2^*(Z_2)) + g_3^*(X) + g_4^*(Z_3) + \varepsilon_2^* > 0).$$

We will show these two models are equivalent.

Taking g_4 to the other side of the inequliy sign and exponentiating both sides one can see that the distribution of,

$$\omega_1 = e^{g_1(X)+g_3(X)+\varepsilon_1+\varepsilon_2} + g_2(Z_1)e^{g_3(X)+\varepsilon_2}$$

must be the same as the distribution of,

$$\omega_1^* = f\left(e^{g_1(X)+g_3^*(X)+\varepsilon_1^*+\varepsilon_2^*} + g_2^*(Z_1)e^{g_3^*(X)+\varepsilon_2^*}\right)$$

when f is defined so that $e^{g_4(X)} = f\left(e^{g_4^*(X)}\right)$. Notice that

$$\log\left(\frac{\partial E\left(\omega_1 \mid X, Z_1, Z_2\right)}{\partial Z_1}\right) = \log(g_2^{'}(Z_1)) + g_3(X) + \log(E(e^{\varepsilon_2}))$$

which is separable in Z_1 and X. The only way that $\log\left(E\left(\omega_1 \mid X, Z_1, Z_2\right)\right)$ can take this separable form is if f is linear. The fact that $g_4(X)$ has full support and that $g_4(0)$ is normalized to zero implies that $g_4 = g_4^*$.

Since

$$\log(g_2^{'}(Z_2)) + g_3(X) + \log(E(e^{\varepsilon_2})) = \log(g_2^{*'}(Z_1)) + g_3^*(X) + \log(E(e^{\varepsilon_2^*}))$$

g_3 and g_3^* can only differ by a location parameter, so since $g_3(0)$ is normalized to zero, g_3 is identified.

Since $g_1$and g_3 are identified it must be that

$$g_2^*(Z_1) = \frac{E(e^{\varepsilon_2^*})}{E(e^{\varepsilon_1})} g_2\left(Z_{\cdot}\right)$$

But since $g_2(0) = g_2\,(0) = 1$, $g_2 = g_2^*$. Thus g_1, g_2, g_3,and g_4 are all identified.

Now consider identification of the joint distribution of the error terms. Fixing $X = 0$, $Z_1$to a particular value z_1, and varying $g_4(Z_2)$ we can identify the joint distribution of $(-\log(e^{\varepsilon_1} + g_2(z_1)) - \varepsilon_2, u_1)$from

$$\Pr(s_i = 1, u_1 < y) = \Pr\left(-\log(e^{\varepsilon_1} + g_2(z_1)) - \varepsilon_2 < g_4(Z_2), u_{1i} < y\right).$$

Thus for any $(t_1, t_2) \in \Re^2$ we can identify,

$$\begin{aligned} E\left(\exp\left\{i\left(t_1 e^{\log(e^{\varepsilon_1}+g_2(z_1))+\varepsilon_2} + t_2 u_1\right)\right\}\right) &= E\left(\exp\left\{i\left(t_1 e^{\varepsilon_1+\varepsilon_2} + t_1 g_2(z_1) e^{\varepsilon_2} + t_2 u_1\right)\right\}\right) \\ &= \varphi(t_1, t_1 g_2(z_1), t_2), \end{aligned}$$

where φ characteristic function of $\left(e^{\varepsilon_{1i}+\varepsilon_{2i}}, e^{\varepsilon_{2i}}, u_{1i}\right)$. By varying t_1, t_2,and $g_2(z_1)$we can identify the characteristic function and thus the joint distribution of $\left(e^{\varepsilon_1+\varepsilon_2}, e^{\varepsilon_2}, u_1\right)$. From this we can identify the distribution of $\left(\varepsilon_1, \varepsilon_2, u_1\right)$. ■