Education and social capital: empirical evidence from microeconomic analyses
Huang, J.

Citation for published version (APA):
Huang, J. (2010). Education and social capital: empirical evidence from microeconomic analyses Amsterdam: Thela Thesis

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 4

Endogeneity Models for Single Treatment Evaluation with a Binary Outcome

4.1 Introduction of the single treatment model

The one-factor model, where a linear relationship between years of schooling and level of individual social capital is assumed to be rational, was applied in the previous chapter to examine the effect of schooling on the development of social trust and social participation. Later in Chapters 5 and 6, a single treatment model will be applied to assess the average treatment effect \( (ATE) \) of college education or higher education on social trust and social participation at the individual level. Before presenting the empirical results, a description of the model is essential for the average treatment effect and the application of the single treatment evaluation with a binary outcome.

The average treatment effect is an econometric measure used to compare treatments in medical trials and policy evaluation. The average treatment effect measures the average causal difference in outcomes under the treatment and the control. The expression “treatment effect” refers to the causal effect of a given medical treatment or policy (for example, the administering of a drug or training program for disadvantaged workers) on an outcome variable of scientific or policy interest (for example, the health of the patient, income of disadvantaged workers, or unemployment spell of unemployed workers). The average treatment effect is the average of the individual treatment effects across the whole population of interest. In this dissertation, it denotes the average expected effect of college education or higher education relative to lower education on individual social trust and individual social participation.

The current approaches to causal inference in treatment evaluation stem from the statistical analysis of randomized experiments and potential outcomes. In the simplest binary framework, there are two outcomes \((Y_1, Y_0)\), which correspond to the treatment dummy \( T \) \((T = 1 \text{ if an individual chooses the treatment, and } T = 0 \text{ otherwise})\). The outcome observed for the individual is hence defined as:

\[
Y = T Y_1 + (1 - T) Y_0
\]
This is the famous Roy (1951)-Rubin (1974) model, or switching model, and the gain of participating in the treatment is \( \Delta = Y_i - Y_o \). This chapter aims to assess the treatment effect of higher education or college education relative to lower levels of education (\( T_i = 1 \) if individual \( i \) undertakes higher/college education, and \( T_i = 0 \) otherwise) on the social participation or social trust outcome (i.e. \( y_i = 1 \) if individual \( i \) is a member of at least one social group, and \( y_i = 0 \) otherwise). In a binary treatment framework where both the outcome and the treatment are a binary response variable:

\[
T_i = 1( T_i^*(Z_i, \nu_i) > 0) \\
y_i = 1( y_i^*(T_i, X_i, \eta_i) > 0)
\]

where \( T_i^* \) and \( y_i^* \) are the latent variables. \( T_i^* \) depends on observed covariates \( Z_i (Z_i = (X_i, z_i)) \)^9, and an unobserved factor \( \nu_i \); \( y_i^* \) depends on education choice \( T_i \), observed covariates \( X_i \), and an unobserved factor \( \eta_i \). Assuming additive separability between observables and unobservables for both latent variables, and a cumulative standard normal distribution for the conditional probability in each equation, a standard bivariate specification is obtained as follows:

\[
\Pr(T_i = 1) = \Phi(f(X_i, z_i) + \nu_i) \\
\Pr(y_i = 1) = \Phi(m(X_i, T_i) + \eta_i) \\
(\nu_i, \eta_i) \sim N(0, 0, 1, 1, \rho_{\nu\eta})
\]

where \( \rho_{\nu\eta} \) is a constant correlation matrix between the unobservable components in treatment and outcome equations\(^10\). Define \( m(X_i, T_i) = b_0 + m_o(X_i) + \beta(X_i)T_i \), and the population ATE, given characteristics \( x \), is directly obtainable:

---

9 Observed covariates \( Z_i (Z_i = (X_i, z_i)) \) include exogenous variable set \( x \), and excluded variable \( z_i \).

10 In a general framework of treatment evaluation, the unobservable component in the outcome equation comprises the random coefficients representing the heterogeneous relationship between treatment choice and outcome. It is difficult, however, to introduce the individual specific random coefficients into a binary response model. Therefore, this analysis focuses on omitted-variable bias instead of selection bias (which contains biases arising from omitted-variable and individual specific marginal returns), and only constant \( \rho_{\nu\eta} \) is considered.
Endogeneity Models for Single Treatment Evaluation with a Binary Outcome

\[ ATE = E[Y_i \mid x] - E[Y_o \mid x] \]
\[ = E[\Phi(b_0 + m_0(X_i) + \beta(X_i))] - E[\Phi(b_0 + m_0(X_i))] \]
\[ = E(\beta(X_i)) \quad (6) \]

When \( \rho_{\nu \eta} \) is non-zero, there would be endogeneity bias in the estimate of \( \beta(X_i) \). The general two-step procedure methods, such as two-stage probit or 2SLS, are not sufficient to provide a consistent estimate for \( \beta(X_i) \), and, consequently, \( ATE \) in the binary response model. The bivariate probit (BVP) method is considered to be more appropriate to handle the endogeneity problem (Wooldridge, 2002; Bhattacharya et al., 2006).

The BVP model has been widely used in medical evaluation to reduce the bias due to the endogeneity in the treatment choice. It is a simultaneous equation model that controls for the endogeneity in the likelihood of four joint sets of the treatment and outcome distribution. Take a joint set \((y_i = 1, T_i = 1)\), for example. The likelihood of this joint set, \( P[y_i = 1, T_i = 1 \mid Z_i, X_i] \), can be written as \( P[y_i = 1 \mid T_i = 1, Z_i, X_i] \cdot P[T_i = 1 \mid Z_i, X_i] \), where the first term \( P[y_i = 1 \mid T_i = 1, Z_i, X_i] \) is expressed as:

\[ P[y_i = 1 \mid T_i = 1, Z_i, X_i] = \int_{-f(x_i,z_i)}^{\infty} \phi((t_i + \frac{m_0(X_i) + \beta(X_i) + \rho_{\nu \eta} \nu_i}{\sqrt{1 - \rho_{\nu \eta}^2}}) \cdot \phi(v_i) \cdot \phi(Z_i, X_i) dv_i \quad (7) \]

The likelihood of the second term is simply a probit likelihood. Combining the first term likelihoods (for all four joint sets of \((y_i, T_i)\)), along with the probit model for the treatment \( T_i \), and taking the log, gives the log-likelihood function for maximum likelihood analysis (Wooldridge, 2002). The bivariate probit imposes a constant \( \rho_{\nu \eta} \) in its implementation, and thus there are no individual specific marginal returns, although it allows for observable heterogeneities of the independent variables \( X \). Under this assumption, the model provides consistent estimates for coefficient \( \beta(X_i) \) and an endogeneity test for the existence of non-zero correlation \( \rho_{\nu \eta} \). On the subject of the endogeneity problem, Bhattacharya et al. (2006) present an inclusive comparison of the performances of the probit, two-stage probit, and bivariate probit models. The results from their Monte Carlo simulations suggest that the bivariate probit model is the only method to produce a consistent estimator when there is an endogenous treatment.

The control functions probit (CFP) method also provides comparable estimates to the
bivariate probit method in a binary response setup. The CFP is an application of the control functions (CF) method in a probit specification. The CF method is generally applied to correct for selection problems in the study of a treatment effect on a continuous outcome variable. Since the probit specification can be derived from a model involving the latent variable $y_i^*$ with a linear expression, the application of the CF method in a probit specification will produce a good approximation of the true $ATE$ in a binary response setup.

The principle inspiring the CF method is to evaluate the treatment effects by controlling directly for the correlation between the treatment choice and the unobservable heterogeneity in the outcome equation (see, e.g., Heckman, 1978; Jimenez and Kugler, 1987; Heckman et al., 2004; Blundell et al., 2005). The CF method allows for outcome unobservables $\eta_i$ to depend on the treatment $T_i$, and it models this dependence. The control functions probit (CFP) applies the same idea to identify the treatment effect on the binary outcome variable. Under joint normality of $\nu_i$ and $\eta_i$ in the treatment and outcome equations and a constant $\rho_{\nu\eta}$ between the unobservable components in the treatment and outcome choices:

$$y_i^* = b_0 + m_0(X_i) + \beta(X_i)T_i + \rho_{\nu\eta}(1-T_i)\lambda_{0i} + \rho_{\nu\eta}T_i\lambda_{ui} + \delta_i \tag{8}$$

A consistent estimator of $\beta(X_i)$ is achievable in equation (8) with a continuous dependent variable, where $\lambda_{0i}$ and $\lambda_{ui}$ are the standard inverse Mills ratios such that:

$$\lambda_{0i} = -\frac{\phi(v_i)}{1 - \Phi(Z_i, X_i)} \tag{9}$$
$$\lambda_{ui} = \frac{\phi(v_i)}{\Phi(Z_i, X_i)} \tag{10}$$

In the binary response model, the transformed error term $\delta_i$ in equation (8) does not generally follow a standard normal distribution. Moreover, the introduction of the corrected functions $(\rho_{\nu\eta}(1-T_i)\lambda_{0i} + \rho_{\nu\eta}T_i\lambda_{ui})$ would lead to a change of the mean and index functions, so that the estimate obtained from the CFP approach is merely an approximation of the true treatment effect. Nevertheless, our simulation practice in the following sections demonstrates that the CFP approach provides a rather precise $ATE$ estimate, which can be comparable to the estimate obtained from the BVP approach, under the assumption of standard bivariate
normality. Compared with the BVP approach, which has a messy and time consuming, though doable, maximum likelihood, the CFP approach has a considerably lower calculation cost, especially when it comes to the estimation of the confidence interval for the treatment effect that involves Monte Carlo simulation. Meanwhile, the CFP method produces a more accurate estimate relative to that from the two-stage probit. The CFP allows the ATE to be identified when individuals select on the basis of the unobservables, and it is possible to test for the presence of treatment endogeneity by a test of the null hypothesis that $\rho_{\eta}$ equals zero. In general, an exclusion restriction is required in the CFP and the BVP approaches. The identification of the estimates will be troublesome if $X_i = Z_i$, especially when there is not much variation in the observable characteristics $X_i$.

### 4.2 Simulation design

The previous section gives a simple review of the key assumptions and procedures of the single treatment framework with a binary outcome. The BVP method produces a consistent estimator in this framework, while the CFP method produces an approximation of the BVP estimator at a considerably lower computational cost. These two methods are major approaches in this evaluation to reduce omitted-variable bias in the assessment of the education effect on individual social capital in adulthood.

To support the arguments that the CFP method does considerably better than the 2SLS and the probit in estimating the ATE, and that it provides an approximate estimate of the ATE that can be comparable to the estimate from the BVP method, the Monte Carlo simulations will be performed here as they were applied by Bhattacharya et al. (2006). Their analysis demonstrates the limitations of the two-step procedure, such as 2SLS and the two-step probit, and they argue in favor of using the bivariate probit rather than the two-step or linear probability model estimators.

In the simulation exercise, a large random data set (5000 observations) is drawn according to a simple data generating process, and then the four different estimators, 2SLS, TSP, BVP

---

11 In the application of the CFP method in the binary response model, the correlation matrix between the unobservable components is also restricted to be constant. In the continuous outcome model, however, a general CF method allows selection on unobserved or omitted ‘ability’ and selection on unobserved marginal returns to treatment. The CFP approach provides comparable estimates to the BVP approach under the assumption of constant correlation and standard bivariate normality, according to the results from Monte Carlo simulations, which follows the same design applied by Bhattacharya et al. (2006). The simulation exercises in this dissertation show that the CFP approach does considerably better than the probit and two-stage probit approaches in the identification of ATE, and it produces an approximate estimate of the true ATE, while the BVP approach produces a consistent estimator.
and CFP estimators, are applied to the same random data set. For the simulation, this step is repeated 1000 times, and the average bias for the four estimators is reported. The simple probit is not considered in the simulation exercise because its inadequacy in handling endogenous treatment has been heavily exploited, and the bias of its estimate obscures the scale in the comparison figures.

**Model specification and detailed steps**

\[
T_i^* = \gamma_0 + x_\gamma x_i + \nu_i \\
T_i = 1(T_i^* > 0) \\
y_i^* = m_\mu + \beta T_i + b_z x_i + \eta_i \\
y_i = 1(y_i^* > 0) \\
(\nu_i, \eta_i) \sim N(0,0,1,1, \rho_{\nu\eta})
\]

where \( T_i^* \) represents the index function generating the treatment \( T_i \); \( x_i \) represents the other repressor; \( z_i \) represents the instrument; \( \nu_i \) and \( \eta_i \) represent the error term in the treatment equation and the outcome equation, respectively; and \( \rho_{\nu\eta} \) is the correlation coefficient between \( \nu_i \) and \( \eta_i \).

**Data-generating process**

There are five parameters in this data-generating process that are varied in the experiment. These parameters alter the character of the random data set. Coefficient \( \gamma_x \) determines the association between constant \( T_i \) and \( x_i \); \( m_\mu \) is the constant term in the outcome equation that determines the average probability of \( y_i \) being equal to 1; the treatment coefficient \( \beta \) reflects the influence of the treatment; the correlation coefficient \( \rho_{\nu\eta} \) determines the correlation between the error terms in the treatment and the outcome equation; and the correlation coefficient \( \rho_{z\eta} \) determines the power of the instrument. In the simulation exercises, \( \gamma_0 \) is first set at zero without loss of generality.

From the simulation exercises it is examined how each evaluation method performs in the case when the treatment \( T_i \) depends on \( x_i \), and in the case when the treatment \( T_i \) does not
depend on $x_i$, with $\gamma_s$ being 0 ($T_i^* = \nu_i$) and 0.5 ($T_i^* = 0.5x_i + \nu_i$), respectively\(^{12}\). Correlation coefficient $\rho_{z\eta}$ is set at 0.5 so that there is a valid and strong instrumental variable (for comparison, values 0.3 and 0.4 were also tried, which led to a similar qualitative conclusion). $\rho_{\nu\eta}$ is arbitrarily specified to be 0.2 (for comparison, values 0.1 and 0.3 were also tried, which led to a similar qualitative conclusion). In the main experiment, $\beta$ is varied between 0 and 2, while holding $m_o$ arbitrarily fixed at -1. In an alternative, $m_o$ is varied while holding the true $ATE$ arbitrarily fixed at 0.2 (for more details, see Bhattacharya et al., 2006). 5000 independent observations ($\nu_i,x_i,z_i,\eta_i$) are drawn from a multivariate normal distribution:

\[
\begin{pmatrix}
\nu_i \\
x_i \\
z_i \\
\eta_i
\end{pmatrix} \sim \text{MVN}
\begin{pmatrix}
1 & 0 & \rho_{z\nu} & \rho_{\nu\eta} \\
0 & 1 & 0 & 0 \\
\rho_{z\nu} & 0 & 1 & 0 \\
\rho_{\nu\eta} & 0 & 0 & 1
\end{pmatrix}
\]

In brief:

- Both the dependent variable and the treatment are binary variables;
- The treatment is correlated with the error term in the dependent variable (several values have been assigned: 0.1, 0.2, 0.3. I arbitrarily specify $\rho_{\nu\eta}$ to be 0.2);
- The instrumental variable is powerful (correlated strongly with the treatment, but not with the error term in the dependent variable. Correlation coefficient $\rho_{z\eta}$ is imposed to be 0.5 so that I have a valid and strong instrument.).

### 4.3 Simulation results

This section first compares the performance of the four methods when $T_i^* = \nu_i$ (the index function generating the treatment is also assumed to be independent of $x_i$ in the study of

\(^{12}\) Several non-zero numbers (0.25, 0.5, 0.75, 1) have been introduced for $\gamma_s$. The change of $\gamma_s$ does not alter the simulation comparison results.
Bhattacharya et al. (2006)). Figure 4.1 shows the bias in the $ATE$ estimate and the bias in its corresponding coefficient estimate – $\beta$ when the value of $\beta$ is varied (the bias from $\beta$ in the OLS is not presented as it is enormously large compared with the biases in other methods). Figure 4.2 shows the bias in the $ATE$ estimate and the bias in its corresponding coefficient estimate when the value of $m_o$ is varied.

The two-step probit (TSP) performs much worse than the other methods. The TSP estimator is noticeably biased for the estimate of $ATE$ and $\beta$, as the true $ATE$ approaches 0.5 (or $\beta$ approaches 2). Its bias in the $ATE$ or $\beta$ is also substantially different from zero as it tends to underestimate the $ATE$ or $\beta$ when the value of $m_o$ is varied. The BVP estimator produces unbiased estimates of the $ATE$ and $\beta$ for all the values tried for $\beta$ and $m_o$. This is not surprising for the large sample simulation exercise, since it is considered a consistent estimator. The 2SLS and the CFP approaches appear to have a good performance in the identification of the true $ATE$. Yet they are not unbiased and consistent estimators. The increasing of $\beta$ or $m_o$ ($m_o$ ranges from -3 to 0) tends to lead to a larger bias for both the 2SLS and the CFP approaches in the simulation exercises.

**Figure 4.1 Bias in the treatment estimate and bias in the coefficient of the treatment estimate**

Notes: 1. The coefficient from the 2SLS is not comparable to the coefficient from other models.

2. The X-axis refers to the value of $\beta$; the Y-axis refers to the value of bias.

3. “olsiv” refers to 2SLS; “ateiv” refers to two-stage probit; “cf” refers to control functions probit; “bvp” refers to bivariate probit.
In the previous setup, it was assumed that $T_i = \nu_i$, such that the choice of the treatment $T_i$ is independent of other observable covariates. This is an extreme case, and it is rare to see independent associations between the treatment $T_i$ and other observable covariates. To give a comprehensive illustration of the performance of the four estimators, a second Monte Carlo simulation is now conducted similar to the one just described, except that $T_i = 0.5x_2 + \nu_i$. 

Figure 4.3 shows the bias in the ATE estimate and the bias in its corresponding coefficient estimate when the value of $\beta$ is varied. Figure 4.4 shows the bias in the ATE estimate and the bias in its corresponding coefficient estimate when the value of $m_o$ is varied. The performance of the BVP is not affected by the change of model setup. The BVP produces unbiased estimates of the ATE and its corresponding $\beta$ for all values of $\beta$ and $m_o$ assigned in the simulation. The performance of the CFP estimator is significantly superior to that of the 2SLS estimator and the TSP estimator. Similar to the setup where $T_i = \nu_i$, the CFP provides an approximate estimate of the ATE that is comparable to the estimate from the BVP method.

Specifically, as shown in Figure 4.3, the TSP and the 2SLS estimators overestimate the ATE and its corresponding coefficient $\beta$ for all non-zero values of $\beta$. The bias increases...
dramatically and becomes noticeably large (up to 0.05) as the true ATE approaches 0.5 or coefficient $\beta$ approaches 2. Their performance in estimating the treatment changes substantially as $m_o$ changes. The TSP and the 2SLS estimator overestimate the ATE for $m_o$ between 0 and -2 and then rapidly decline with large negative bias as $m_o$ decreases.

It is clearly shown that the performance of the BVP is not affected by the change in the generation of the random data sets. The BVP produces unbiased and consistent estimates of the ATE for all values of $\beta$ and $m_o$ assigned. The CFP produces an approximate estimate of the ATE, which is very close to that obtained from the BVP. Similarly, it cannot produce an unbiased and consistent estimator. The absolute value of bias in the ATE or $\beta$ increases moderately as the absolute value of $\beta$ or $m_o$ increases.

**Figure 4.3 Bias in the treatment estimate and bias in the coefficient of the treatment estimate**

Notes: 1. The coefficient from the 2SLS is not comparable to the coefficient from other models.

2. The X-axis refers to the value of $\beta$; the Y-axis refers to the value of bias.

3. “olsiv” refers to 2SLS; “ateiv” refers to two-stage probit; “cf” refers to control functions probit; “bvp” refers to bivariate probit.
**Figure 4.4 The constant term and bias in the treatment estimate and bias in the coefficient of the treatment estimate**

Notes: 1. The coefficient from the 2SLS is not comparable to the coefficient from other models.
2. The X-axis refers to the value of $\beta$; the Y-axis refers to the value of bias.
3. “olsiv” refers to 2SLS; “ateiv” refers to two-stage probit; “cf” refers to control functions probit; “bvp” refers to bivariate probit.

**4.4 Conclusion**

This chapter has presented a simple review of the key assumptions and procedures of the single treatment framework with a binary outcome. The bivariate probit (BVP) and control functions probit (CFP) approaches are proposed to be valid in the tackling of treatment endogeneity. To support the argument, the Monte Carlo simulations are presented to compare the performance of the two-stage least square (2SLS), the two-stage probit (TSP), the BVP, and the CFP, which are similar to the simulation procedures applied by Bhattacharya et al. (2006).

The simulation exercises in this chapter provide identical findings to those of Bhattacharya et al. (2006) – the BVP estimator produces consistent estimates, while the 2SLS and TSP do not. Furthermore, the simulation exercises also confirm that the CFP approach produces a close estimate of the true $ATE$ that can be seen as a good approximation to that from the BVP. These two methods are major approaches in the studies in the later chapters to reduce endogeneity bias in the assessment of the education effect on the two dimensions of individual social capital.