Estimating Impact of a Continuous Program under a Conditional Independence Assumption

Nguyen Viet, Cuong

Development Economics Group, Mansholt Graduate School, Wageningen University, Netherlands

01. January 2008

Estimating Impact of a Continuous Program under a Conditional Independence Assumption

Nguyen Viet Cuong[1]

Abstract

Traditional literature on impact evaluation often deals with binary programs. In reality, a program can provide different amounts of treatment for people, and it is regarded as a continuous variable. This paper discusses impact evaluation of a continuous program using matching methods and illustrates how to estimate impact of foreign remittances on per capita expenditure in Vietnam.

Keywords: Treatment effect, program impact, continuous program, matching, conditional independence.

JEL classification: C14; C15; H43; J41

[1] Development Economics Group, Mansholt Graduate School, Wageningen University, the Netherlands.
Address: Hollandsweweg 1, 6706 KN Wageningen, the Netherlands.
E-mail: c_nguyenviet@yahoo.com

1. Introduction

The main objective of impact evaluation is to assess the extent to which a program has changed outcomes of subjects.[2] The average impact of a program on a group of subjects is defined as the difference between their outcome in the presence of the program and their outcome in the absence of the program. However, for each subject, we are not able to observe these two potential outcomes at the same time. This missing data problem can be solved if an assumption of conditional independence of treatment and potential outcomes holds (Rubin, 1977). Under this assumption, the program impact can be estimated by OLS regression or matching methods. The idea of the matching method is to compare outcomes of participants and non-participants who have the same conditioning variables.

The literature of program impact evaluation often deals with impact of binary programs. In reality, a program can provide different amounts of treatment for people, and it can be regarded as continuous instead of binary. This paper discusses impact evaluation of a continuous program using matching methods and illustrates how to estimate impact of foreign remittances on per capita expenditure in Vietnam.

The paper is organized into five sections as follows. The second section presents the matching method in impact evaluation of a single binary program. The third section discusses the matching method in the case of a continuous program. Results of the empirical example are presented in the fourth section. Finally the fifth section concludes.

2. Impact Evaluation of a Binary Program

Two popular parameters in program impact evaluation are the Average Treatment Effect (ATE), and the Average Treatment Effect on the Treated (ATT), which are expressed as follows:[3]

$$ATE_{(X)} = E(\Delta \mid X) = E(Y_1 \mid X) - E(Y_0 \mid X), \qquad (1)$$

$$ATT_{(X)} = E(\Delta \mid X, D=1) = E(Y_1 \mid X, D=1) - E(Y_0 \mid X, D=1), \qquad (2)$$

[3] There are other parameters such as local average treatment effect, marginal treatment effect, or even effect of "non-treatment on non-treated" which measures what impact the program would have on the non-participants if they had participated in the program, etc.

where D is a binary variable of program participation, X are exogenous conditioning variables, Y_0 and Y_1 are potential outcomes without and with the program, respectively.[4]

The identification assumption for matching is the conditional independence assumption as follows:

Assumption 1: $Y_0, Y_1 \perp D | X$ (A.1)

The assumption states that conditioning on X, Y_0 and Y_1 are independent of D. Under this assumption, $ATE_{(X)}$ and $ATT_{(X)}$ can be identified as follows:

$$ATE_{(X)} = E(Y_1 \mid X) - E(Y_0 \mid X) = E(Y_1 \mid X, D=1) - E(Y_0 \mid X, D=0), \tag{3}$$

$$ATT_{(X)} = E(Y_1 \mid X, D=1) - E(Y_0 \mid X, D=1) = E(Y_1 \mid X, D=1) - E(Y_0 \mid X, D=0). \tag{4}$$

The matching method estimates these parameters by comparing the outcome of the participants and the outcome of a so-called comparison group who do not participate in the program but have the X variables identical to those of the participants.

A difficulty in the matching method is to find matched non-participants for the participants when there are many the X variables. In the literature of impact evaluation, three widely-used methods to match non-participants with participants are subclassification (see, e.g., Cochran and Chambers, 1965; Cochran, 1968), covariate matching (Rubin, 1979, 1980), propensity score matching (Rosenbaum and Rubin, 1983).

3 Impact Evaluation of a Continuous Program

In reality, a program can provide different amounts of treatment for people. Denote a continuous program by d whose value is equal to the treatment level that the program can provide for people. For those who do not take the program, d is equal to 0. We can be interested in impact of the program at a range (interval) of the treatment $[L_1, L_2]$:

$$ATE_{(X, d\in[L_1,L_2])} = E\left(Y_{d\in[L_1,L_2]} \middle| X\right) - E\left(Y_{d=0} \middle| X\right), \tag{5}$$

$$ATT_{(X, d\in[L_1,L_2])} = E\left(Y_{d\in[L_1,L_2]} \middle| X, d\in[L_1,L_2]\right) - E\left(Y_{d=0} \middle| X, d\in[L_1,L_2]\right). \tag{6}$$

[4] Y_0 and Y_1 can be vectors of outcomes, but for simplicity let's consider a single outcome of interest.

If d is real value and bounded, we can denote the range of the d variable by $[L_{min}, L_{max}]$. We can divide d into $n+1$ mutually exclusive and successive intervals $[L_{min}; L_1], (L_1; L_2], ..., (L_n; L_{max}]$.

Suppose that we can write the equations of the potential outcomes as follows:

$$d \notin [L_{min}, L_{max}] \quad \Rightarrow \quad Y_0 = \alpha_0 + X\beta_0 + \varepsilon_0,$$

$$d \in [L_{min}, L_1] \quad \Rightarrow \quad Y_1 = \alpha_1 + X\beta_1 + \varepsilon_1,$$

....

$$d \in (L_n, L_{max}] \quad \Rightarrow \quad Y_{n+1} = \alpha_{n+1} + X\beta_{n+1} + \varepsilon_{n+1}.$$

The observed outcome is then written in terms of the potential outcomes:

$$Y = Y_0 + (Y_1 - Y_0)I\{d \in [L_{min}, L_1]\} + (Y_2 - Y_0)I\{d \in (L_{min}, L_2]\} + ... + (Y_n - Y_0)I\{d \in (L_n, L_{max}]\} \tag{7}$$

Where I{ } is an indicator function that is equal to 1 if the value of { } is true, 0 otherwise.

Substitute the equations of the potential outcomes into (7), we get:

$$\begin{aligned} Y = {} & \alpha_0 + X\beta_0 + [(\alpha_1 - \alpha_0) + X(\beta_1 - \beta_0)]I\{D \in [L_{min}, L_1]\} + [(\alpha_2 - \alpha_0) + X(\beta_2 - \beta_0)]I\{D \in (L_1, L_2]\} \\ & + ... + [(\alpha_{n+1} - \alpha_0) + X(\beta_{n+1} - \beta_0)]I\{D \in (L_n, L_{max}]\} \\ & + [\varepsilon_0 + I\{D \in [L_{min}, L_1]\}(\varepsilon_1 - \varepsilon_0) + I\{D \in (L_1, L_2]\}(\varepsilon_2 - \varepsilon_0) + ... + I\{D \in (L_n, L_{max}]\}(\varepsilon_{n+1} - \varepsilon_0)] \end{aligned} \tag{8}$$

Similar to the case of a single binary program, impact of the program at $d = L \in (L_i, L_{i+1}]$ can be identified under the conditional independence assumption.

Assumption 2: Conditional independence between the potential outcomes and treatment levels:

$$Y_0, Y_i \perp I\{d \in (L_i, L_{i+1}]\} | X \qquad \forall i, j \tag{A.2}$$

Then, impact of the treatment at the interval $(L_i, L_{i+1}]$ can be estimated either by parametric methods, e.g., running an OLS regression on equation (8) with dummy variables d_i indicating $I\{d \in (L_i, L_{i+1}]\}$, or by non-parametric methods, e.g., the matching method by matching those who receive the treatment of level $(L_i, L_{i+1}]$ with those who do not receive the treatment (i.e. d = 0).

In reality, there can be a problem in estimating the program impact in the region of no data on the program treatment. To infer the program impact to the region of no treatment level, we can assume a functional form of the expected program impact conditional on X and d. Recall that $ATE_{(X)}$ and $ATT_{(X)}$ at a given treatment level $d = L \in (L_i, L_{i+1}]$ using a partition of n intervals are:

$$ATE_{(X, d \in (L_i, L_{i+1}], n)} = ATT_{(X, d \in (L_i, L_{i+1}], n)} = (\alpha_{i+1} - \alpha_0) + X(\beta_{i+1} - \beta_0). \tag{9}$$

The conditional parameters depend on X, d, and the number of partitioned intervals that is indicated by n. Thus we can write:

$$ATE_{(X,d,n)} = ATT_{(X,d,n)} = f_n(X,d), \tag{10}$$

where $f_n()$ are real functions defined on domain of X and d. $f_n(X,d)$ has an important property that $f_n(X, d=0) = 0$, i.e. the impact of the program equals to 0 when there is no treatment. Substitute $f_n(X,d)$ into equation (8), we have:[5]

$$Y = \alpha_0 + \beta_0 X + f_n(X,d) + \varepsilon \tag{11}$$

where:

$$\varepsilon = [\varepsilon_0 + I\{d \in [L_{min}, L_1]\}(\varepsilon_1 - \varepsilon_0) + I\{d \in (L_1, L_2]\}(\varepsilon_2 - \varepsilon_0) + ... + I\{d \in (L_n, L_{max}]\}(\varepsilon_{n+1} - \varepsilon_0)]$$

If assumption (A.2) holds for any interval of the treatment, ε has the traditional property that $E(\varepsilon | X, d) = 0$.

To define the average treatment effect at any arbitrarily small interval, let n go to infinity, and the interval, $[L_{min}, L_{max}]$, is divided into infinite sub-intervals as a result. Further, we can assume that functions $f_n(X,d)$ converge to a function $f(X,d)$ as n goes to infinity, i.e.:

$$f(X,d) = \lim_{n \to \infty} f_n(X,d). \tag{12}$$

Then, (11) becomes:

$$Y = \alpha_0 + X\beta_0 + f(X,d) + \varepsilon. \tag{13}$$

Once the functional form of $f(X,d)$ is specified, its parameters can be estimated by two ways. The first way is to run an OLS regression of Y on X and D using a function of Y, e.g., as in (13). The second way is to estimate program impact for all the participants by the matching method, and then to run an OLS regression of these impact estimates on X and D. The second method can be regarded as a semi-parametric method. Compared with the first approach, it is more robust in the sense it requires a functional form of $f(X,d)$, while the first approach requires functional forms of both Y and $f(X,d)$.

[5] Note that $I\{D \in [L_{min}, L_1]\} + I\{D \in (L_1, L_2]\} + ... + I\{D \in (L_n, L_{max}]\} = I\{D \in [L_{min}, L_{max}]\} = 1$.

We can also define the marginal treatment effect conditional on X and d as follows:

$$MTE_{(X,d)} = \frac{\partial ATE_{(X,d)}}{\partial d} = \frac{\partial f(X,d)}{\partial d}, \qquad (14)$$

which measures how the conditional average treatment effect changes as the treatment variable changes. In other words, $MTE_{(X,d)}$ measures the additional amount of outcome that one can gain when receiving an additional treatment level.

4. Empirical Example

Vietnam receives a large amount of foreign remittances annually. The total foreign remittances were nearly 4 billion USD in 2006. Foreign remittances are expected to increase household consumption. To measure impact of foreign remittances, we use Vietnam Household Living Standard Survey in 2004. The survey is conducted by General Statistical Office of Vietnam with technical support of World Bank. The number of households covered by the survey is 9188.

If we estimate the impact by OLS regression, we can assume that outcome has the following functional form:

$$Y = \alpha + \beta X + \gamma d + \theta d^2 + \varepsilon, \qquad (15)$$

where Y is per capita expenditure or log of per capita expenditure, X are household characteristics, d is amount of foreign remittance received by households. The X variables includes dummy regional variables (Vietnam is divided into geographic regions), urban/rural, age and occupation of household heads, and other variables of household composition and education.[6]

To estimate MTE or the impact distribution of the program by matching, we can implement two steps as follows. In the first step, we estimate impact for every household by a matching method. In this example, we perform matching based on the propensity score which is defined as the probability of receiving foreign remittances (regardless of remittance amount). The propensity scores can be estimated by logit or probit models. Then for each household who receives remittances, we calculate the difference in her observed outcome and the weighted average outcome of her matched households.

[6] Readers can contact the author for detailed description of variables and estimation results.

In the second step, we can use parametric regressions to estimate MTE. The main problem is to specify a functional form of $f(X,d)$. In this example, to compare this semi-parametric method with OLS regression we suppose that $f(X,d)$ has a form similar to (15), i.e.:

$$f(X,d) = \gamma d + \theta d^2, \tag{16}$$

Hence:

$$MTE = \gamma + 2\theta d \tag{17}$$

We can estimate γ and θ by running OLS regression of estimated impact on d and d^2 for remittance-receiving households.

Table 1 presents estimates of γ and θ using OLS regression and the so-called semi-parametric approach in which the propensity score matching is used to estimate program impact for remittance-receiving households in the first step. It shows that the two methods give rather close estimates. However, the semi-parametric method produces higher standard errors compared with the OLS regression.

Table 1: Impact of foreign remittances on per capita expenditure

Estimators	Explanatory variables	Dependent (outcome) variables			
		Log of per capita expenditure		Per capita expenditure	
		Mean	Std. Err.	Mean	Std. Err.
OLS regression	Remittances	0.000018	0.000001	0.138280	0.016050
	Remittances squared	-7.15E-11	6.32E-12	-4.54E-07	1.37E-07
Matching with 1 nearest neighbor in the first step	Remittances	0.000016	0.000003	0.126540	0.029646
	Remittances squared	-6.54E-11	3.02E-11	-4.46E-07	3.21E-07
Matching with 3 nearest neighbors in the first step	Remittances	0.000013	0.000002	0.117577	0.026430
	Remittances squared	-5.09E-11	2.25E-11	-3.90E-07	2.91E-07
Kernel matching with bandwidth of 0.05 in the first step	Remittances	0.000011	0.000002	0.129772	0.022332
	Remittances squared	-4.26E-11	1.61E-11	-4.55E-07	3.07E-07

Note: Both per capita expenditure and remittances are measured in thousand VND.

"E-x" means "multiplied by 10^{-x}", e.g. $-7.15\text{E}-11 = \frac{-7.15}{10^{11}}$

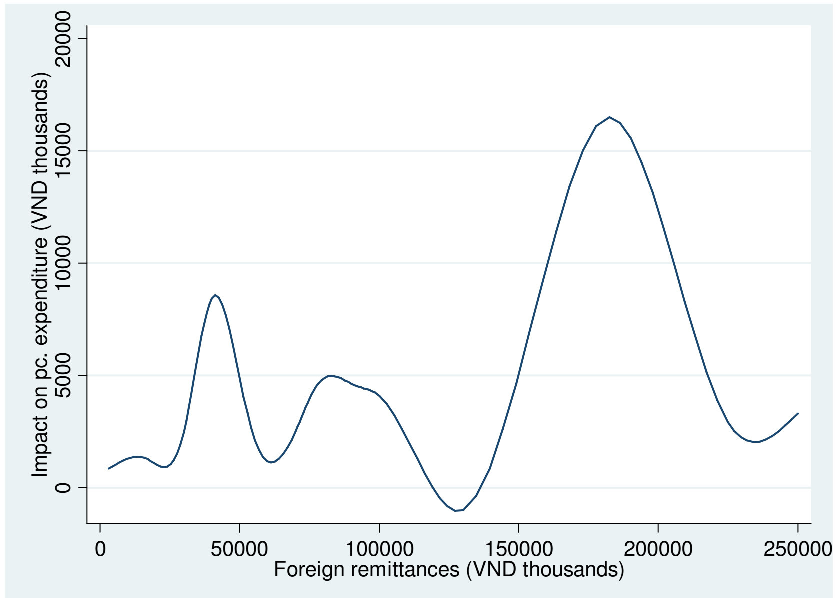
The number of observations is 9188. Ratio of households receiving foreign remittance is 6.1%.

Standard errors for matching estimates are calculated using bootstrap with 500 replications.

Source: Author's estimation

We can use nonparametric regression of estimated impact on the program variables. For example, Figure 1 graphs estimated impacts of remittance on per capita expenditure.

Figure 1: Median spline curve of estimated impact of foreign remittances on per capita expenditure



Source: Author's estimation

Finally, it should be noted that the matching method in the first step can allow for the estimation of the treatment effect at any interval of the treatment as long as the number of the treatment observations in that interval is large enough to get the reliable estimation.

5. Conclusions

In fact, a program can be continuous. People can receive different levels of treatment from a program. In this case, program impact can be identified under the conditional independence assumption and estimated by the matching methods. In addition to ATT and ATE, MTE can be defined, which measures the additional impact that people can gain when receiving an additional program level. These parameters can be estimated by a semi-parametric approach with two steps. In the first step, the treatment effect at different values of the conditioning and program variables is estimated by the matching method. In the second step, the distribution of program impact and MTE can be estimated non-parametrically or parametrically using the program impact estimates from the first step. Compared with OLS regression, it is more robust in the sense that it requires a functional form of program impact, while OLS regression requires functional forms of both outcome and program impact.

References

Cochran, W. G. and S. P. Chambers, 1965, The Planning of Observational Studies of Human Population, Journal of the Royal Statistical Society, Series A (General) 128, No. 2, 234-266.

Cochran, W. G., 1968, The effectiveness of Adjustment by Subclassification in Removing Bias in Observational Studies, Biometrics, 24, 295-313.

Rosenbaum, P. and Rubin R., 1983, The Central Role of the Propensity Score in Observational Studies for Causal Effects, Biometrika, 70 (1), 41-55.

Rubin, D., 1974, Estimating Causal Effects of Treatments in Randomized and Non-Randomized Studies, Journal of Educational Psychology, 66, 688-701.

Rubin, D., 1977, Assignment to a Treatment Group on the Basis of a Covariate, Journal of Educational Statistics, 2 (1), 1-26.

Rubin, D., 1979, Using Multivariate Sampling and Regression Adjustment to Control Bias in Observational Studies, Journal of the American Statistical Association, 74, 318–328.

Rubin, D., 1980, Bias Reduction Using Mahalanobis-Metric Matching, Biometrics, 36 (2), 293–298.