Impact Evaluation of Multiple Overlapping Programs using Difference-in-differences with Matching

Nguyen Viet, Cuong

Development Economics Group, Wageningen University, the Netherlands

1 March 2008

Impact Evaluation of Multiple Overlapping Programs using Difference-in-differences with Matching

Nguyen Viet Cuong[1]

Abstract

Difference-in-differences with matching is a popular method in impact evaluation. Traditional impact evaluation methods including difference-in-differences with matching often deal with impact measurement of a single binary program. Imbens (1999) and Lechner (2001) extend the matching method to the case of multiple mutually exclusive programs. Frölich (2002) discusses different impact evaluation methods in the similar context. In reality, one can participate in several programs simultaneously and the programs may be overlapping. This paper discusses the method of difference-in-differences with matching in a general context of multiple overlapping programs. The method is applied to measure impacts of formal and informal credit in Vietnam using panel data from two Vietnam Household Living Standard Surveys in 2002 and 2004.

Keyword: Treatment effect, impact evaluation, multiple programs, difference-in-differences, matching, propensity score.

JEL classification: C14; C21; H43; J41

---

[1] This paper was written during my Ph.D. study in the Development Economics Group of Wageningen University, the Netherlands under the supervision of Prof. David Bigman, Dr. Marrit Van Den Berg, and Prof. Vu Thieu. I am also very grateful to Dr. Daniel Westbrook in the Economics Department of Georgetown University, Washington DC. for his valuable comments and corrections. My e-mail: c_nguyenviet@yahoo.com ; Tel: (84) 904159258.

1. Introduction

Traditional literature on program impact evaluation often deals with a single binary program. Imbens (1999) and Lechner (2001) extend the matching method to the case of multiple mutually exclusive programs. Frölich (2002) discusses different impact evaluation methods in a context multiple mutually exclusive binary programs. However, in reality the programs can be overlapping. Some people can join several programs at the same time. For example, for evaluation of a micro-credit program that is provided by a bank, the participants and non-participants in the program can receive credit from other sources such as private lenders, relatives and other credit institutions.

difference-in-differences with matching is a popular method of program impact evaluation. Panel data become more available in both developed and developing countries. The method has two main advantages. Firstly, it allows for the selection of the program based on unobservable time-invariant variables. In this sense, it is more robust than evaluation methods which are based on conditional independence assumption such as matching using single cross-section data. Secondly, difference-in-differences with matching can be regarded a nonparametric method, which avoid the functional form assumptions invoked by parametric methods.

This paper discusses the difference-in-differences with matching method in a general context in which subjects can participate in several programs simultaneously. The method is illustrated by measuring impacts of formal and informal credit in Vietnam. The panel data used for the impact estimate are from two Vietnam Household Living Standard Surveys in 2002 and 2004.

The paper is organized into six sections as follows. The second section presents the problems and parameters of interest in impact evaluation. The third section discusses the method of difference-in-differences with matching in the case of a single binary program. The fourth section extends the method to the case of multiple overlapping programs. Next, the fifth section presents the application of the method in measuring impacts of formal and informal credit in Vietnam. Finally the sixth section concludes.

2. Evaluation of Program Impact: Problems and Parameter of Interest

The main objective of impact evaluation of a program is to assess the extent to which the program has changed outcomes of subjects. In other words, impact of the program on participants is measured by the change in welfare outcome that is attributed only to the program. In literature of impact evaluation, a broader term "treatment" instead of program/project is sometimes used to refer an intervention whose impact is evaluated.

To make the definition of impact evaluation more explicit, suppose that there is a program assigned to some people in population P. For simplicity, let's assume there is a single program, and denote D as a binary variable of participation in the program of a person, i.e. D equals 1 if she/he participates in the program, and D equals 0 otherwise. Let Y denote the observed value of an interested outcome. This variable can receive two potential values depending on the binary values of the participation variable, i.e. $Y = Y_1$ is the outcome in status of the program, and $Y = Y_0$ is the outcome in the status of no-program. Certainly, the potential outcomes are considered at a point of time after the program is implemented.

The impact of the program on the outcome of a person i is measured by the following difference:

$$\Delta_i = Y_{i1} - Y_{i0} \tag{2.1}$$

It is equal to the difference between the outcome of the person when she/he participates in the program and the potential outcome of that person when she/he does not participate in the program. The problem is that we cannot observe both terms in equation (2.1) for one person. For those who participated in the program, we can observe Y_1, but we cannot observe Y_0 – the outcome if they would had not participated in the program. Similarly, we can observe Y_0, but not Y_1 for those who did not participate in the program. In this context, outcomes that we cannot observe are called counterfactual.

It is wide consent that it is almost impossible to estimate program impact for each person (Heckman et. al., 1999). In fact, program impact can be estimated for a group of subjects. The most popular parameter of the program impact is Average Treatment Effect on the Treated (ATT) (Heckman et. al., 1999), which is the expected impact of the program on the actual participants:

$$ATT = E(Y_1 - Y_0 | D = 1) = E(Y_1 | D = 1) - E(Y_0 | D = 1) \tag{2.2}$$

More generally, we can allow these effects to vary across a vector of the observed variables X:

$$ATT_{(X)} = E(\Delta \mid X, D = 1) = E(Y_1 \mid X, D = 1) - E(Y_0 \mid X, D = 1) \tag{2.3}$$

Another parameter which is also widely mentioned in impact evaluation is Average Treatment Effect (ATE). This parameter measures impact of the program on randomly selection people. ATE is defined as:[2]

$$ATT = E(Y_1) - E(Y_0) \tag{2.4}$$

In this paper, we focus on ATT since this is the most popular parameter in impact evaluation. The estimation strategy of ATE is very similar. In the following discussion we will focus more on identification of the conditional parameters, since once the conditional parameters are identified, the unconditional parameters are also identified:

$$ATT = \int_{X \mid d_1 = 1} ATT_{(X)} \mathrm{dF}(\mathrm{X} \mid d_1 = 1) \tag{2.5}$$

Estimation of $\text{ATT}_{(X)}$ is not straightforward, since $E(Y_0 \mid X, D = 1)$ are not observed and cannot be estimated directly. $E(Y_0 \mid X, D = 1)$ is called counterfactual which is the expected outcome of non-participants if they would had participated in the program.

3. Impact Estimation of a Program using Difference-in-Difference with Matching

[2] There are other parameters such as local average treatment effect, marginal treatment effect, or even average 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.

3.1. Matching Method

There is a large amount of literature on matching methods of impact evaluation. Important contributions in this area can be found in studies such as Rubin (1977, 1979, 1980), Rosenbaum and Rubin (1983, 1985a), and Heckman, et al. (1997). The matching method can be used to estimate ATT under the conditional independence assumption. Formally, this assumption is written as:[3]

Assumption 3.1: $Y_0 \perp D \mid X$ (A.3.1)

Actually, we just need a weaker form of (A.3.1) in order to identify the program impact parameter.

Assumption 3.2: $E(Y_0 \mid X, D) = E(Y_0 \mid X)$ (A.3.2)

This is called the conditional mean independence assumption. It is weaker than (A.4.1) in sense that (A.3.1) implies (A.3.2) but the reverse is not correct.

The basic idea of the matching method is to find a control group (also called comparison group) that has the same (or at least similar) distribution of *X* as the treatment group. By doing so, we have controlled for the differences in *X* between the participants and non-participants. The potential outcomes of the control and treatment group are now independent of the program selection. The difference in outcome of the control group and the treatment group then can be attributed to the program impact.

However for the matching method to be implemented, we must find a control group that is similar to the treatment group but does not participate in the program. This similarity assumption is called common support. If we denote $p(X)$ as the probability of participating in the program for each subject, i.e. $p(X) = P(D=1 \mid X)$, the assumption can be stated formally as follows:

Assumption 3.3: $0 < p(X) < 1$ (A.3.3)

Proposition 3.1: Under assumptions (A.3.2) and (A.3.3), $ATT_{(X)}$ and ATT are identified by the matching method.

Proof: the proof is straightforward using the conditional independence assumption.

$$ATT_{(X)} = E(Y_1 \mid X, D=1) - E(Y_0 \mid X, D=1) = E(Y_1 \mid X, D=1) - E(Y_0 \mid X, D=0) \ . \quad (3.1)$$

Both terms in (3.1) can be observed. In addition, assumption (A.3.3) ensures that there are some participants and non-participants whose values of *X* are the similar so that we are able to use sample information to estimate (3.1). ATT is also identified as in (2.5).■

The difficulty in the matching method is to how find matched non-participants for the participants when there are many variables *X*. A popular solution is proposed by Rosenbaum and Rubin (1983) who show that if the potential outcomes are independent of the program assignment given by the variables *X*, then they are also independent of the program assignment given the balance score.[4]

---

[3] If we want to estimate both ATT and ATE, we need the conditional independence assumption for both Y_0 and Y_1, i.e., $Y_0, Y_1 \perp D \mid X$.

[4] Other matching methods are subclassification (Cochran and Chambers, 1965) and (Cochran, 1968), and covariate matching (Rubin, 1979, 1980).

Proposition 3.2 (Rosenbaum and Rubin, 1983)**:**

$Y_0, Y_1 \perp D | X \Rightarrow (Y_0, Y_1) \perp D | b(X)$

where $b(X)$ is any function such that $p(X) = f[b(X)]$ and $p(X) = Pr(D=1|X) = E(D|X)$.

A natural choice of the balance score is the propensity score, i.e., the probability of being assigned to the program. Using this proposition, $ATT_{(X)}$ is rewritten as:

$$ATT_{(X)} = E(Y_1 | p(X), D=1) - E(Y_0 | p(X), D=0) . \quad (3.2)$$

Thus non-participants are matched with the participants based on the propensity score. Once the comparison is constructed, the parameters of program impact can be estimated by comparing the outcome of the comparison and treatment groups.

The matching method which relies on assumption (A.3.1) or (A.3.2) will lead to biased estimation of the program impacts if the program selection is based on not only observed but also unobserved variables. For example, people can participate in a micro-credit program because they have higher motivation for high income or better business and production skills. If these variables are not observed and controlled, the matching method will produce biased estimators of the program impacts.

3.2. Difference-in-Difference with Matching

When panel data on the participants and non-participants in a program before and after the program implementation are available, we can estimate the program impacts using the method of difference-in-differences with matching. This method allows the program selection to be based on unobserved variables. However it requires these unobserved variables time-invariant.

Let's denote Y_{0B}, X_B as the outcome and conditioning variables before the program. After the program, the potential outcomes are denoted as Y_{0A}, Y_{1A} corresponding to the states of no-program and program, and the conditioning variables are denoted as X_B. The identification assumptions of the difference-in-differences with matching method are as follows.

Assumption 3.4: Conditional on *X*, the difference in the expectation of outcomes between the participants and non-participants are unchanged before and after the program, i.e.:

$$E(Y_{0B} | X_B, X_A, D=1) - E(Y_{0B} | X_B, X_A, D=0) = E(Y_{0A} | X_B, X_A, D=1) - E(Y_{0A} | X_B, X_A, D=0) \quad \text{(A.3.4)}$$

Assumption 3.5: $0 < P(D=1|X) = P(D=1|X_B, X_A) < 1$ (A.3.5)

Assumption (A.3.5) is the common support assumption which means that there are non-participants who have variables X_B and X_A similar to those of the participants in the program.

Proposition 3.3: Under assumptions (A.3.4) and (A.3.5), $\mathrm{ATT}_{(X)}$ and ATT are identified by the difference-in-differences with matching method.

<u>*Proof:*</u>

Recall the parameter $\mathrm{ATT}_{(X)}$ is equal to:

$$ATT_{(X)} = ATT_{(X_B, X_A)} = E(Y_{1A} \mid X_B, X_A, D=1) - E(Y_{0A} \mid X_B, X_A, D=1) \quad (3.3)$$

Insert equation in (A.3.4) into (3.3) to obtain:

$$\begin{aligned} ATT_{(X_B, X_A)} = {} & E(Y_{1A} \mid X, D=1) - E(Y_{0A} \mid X, D=1) - \left[E(Y_{0A} \mid X, D=0) - E(Y_{0B} \mid X, D=0)\right] \\ & + \left[E(Y_{0A} \mid X, D=1) - E(Y_{0B} \mid X, D=1)\right] \\ = {} & \left[E(Y_{1A} \mid X, D=1) - E(Y_{0A} \mid X, D=0)\right] - \left[E(Y_{0B} \mid X, D=1) - E(Y_{0B} \mid X, D=0)\right] \end{aligned}$$

The unconditional parameter is also identified by (2.5).■

According to the method, the non-participants are matched with the participants based on their variables X before and after the program. The matched non-participants will form a comparison groups.

Note that the term $\left[E(Y_{0A} \mid X_B, X_A, D=1) - E(Y_{0A} \mid X_B, X_A, D=0)\right]$ in (A.3.4) is set equal to zero if we want to identify the program impacts using single cross-section data. This bias arises when the conditional expectation of outcome of non-participants is used to predict the conditional expectation of outcome of participants if they had not participated in the program. Matching method using single cross-section data assumes this bias equals zero once conditional on X. Thus the panel data matching method is more robust than the matching method in sense that it allows this bias to differ from zero. It, however, requires that this bias be time-invariant.

4. Difference-in-Difference with Matching in Multiple Overlapping Programs

4.1. The Case of Two Overlapping Programs

For illustration of the ideas, this section discusses impact evaluation of two programs. In the next section, the method will be extended to the case of multiple programs.

Suppose that there are two programs that are assigned to some people in population P. Denote D as a vector variable of program participation of a person. D has two binary variable elements: d_1 and d_2, i.e.:

$$D = \begin{pmatrix} d_1 \\ d_2 \end{pmatrix}$$

where $d_1 = 1$ if the person receives the program 1, and $d_1 = 0$ otherwise; similarly $d_2 = 1$ if the person receives the program 1, and $d_2 = 0$ otherwise. As a result, the set of the potential treatment have 4 values:

$$\Omega_D = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}; \begin{pmatrix} 1 \\ 0 \end{pmatrix}; \begin{pmatrix} 0 \\ 1 \end{pmatrix}; \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\}$$

Further let Y_B denote the value of an interested outcome before the program implementation. After the program, the potential outcome set is $\Omega_{Y^P} = \{Y_{11};Y_{10};Y_{01};Y_{00}\}$[5], corresponding to the values of the participation variable.

Suppose we are interested in program impact of the program d_1 measured by ATT. The identification of the program d_2 is the same. Denote Y_1 as the potential outcome of a person when she/he participates in the program d_1 (d_1 = 1), and Y_0 as the potential outcome when she/he does not participate in the program (d_1 = 0). ATT for the program d_1 is defined as:

$$ATT1_{(X)} = E(Y_1 - Y_0 | X_B, X_A, d_1 = 1) = E(Y_1 - Y_0 | X, d_1 = 1) \qquad (4.1)^{6}$$

To express this parameter in terms of the four potential outcomes, we rearrange (4.1):

$$\begin{aligned}
ATT1_{(X)} &= E(Y_1 \mid X, d_1 = 1) - E(Y_0 \mid X, d_1 = 1) \\
&= \left[E(Y_{11} \mid X, d_1 = 1, d_2 = 1)\Pr(d_2 = 1 \mid X, d_1 = 1) + E(Y_{10} \mid X, d_1 = 1, d_2 = 0)\Pr(d_2 = 0 \mid X, d_1 = 1)\right] \\
&\quad - \left[E(Y_{01} \mid X, d_1 = 1, d_2 = 1)\Pr(d_2 = 1 \mid X, d_1 = 1) + E(Y_{00} \mid X, d_1 = 1, d_2 = 0)\Pr(d_2 = 0 \mid X, d_1 = 1)\right] \\
&= \left[E(Y_{11} \mid X, d_1 = 1, d_2 = 1) - E(Y_{01} \mid X, d_1 = 1, d_2 = 1)\right]\Pr(d_2 = 1 \mid X, d_1 = 1) \\
&\quad + \left[E(Y_{10} \mid X, d_1 = 1, d_2 = 0) - E(Y_{00} \mid X, d_1 = 1, d_2 = 0)\right]\Pr(d_2 = 0 \mid X, d_1 = 1)
\end{aligned} \qquad (4.2)$$

The above formula allows for the overlap between the program d_1 and the program d_2. If the two programs are mutually exclusive, then the term $\Pr(d_2 = 1 \mid X, d_1 = 1)$ will be equal to zero, and the term $\Pr(d_2 = 0 \mid X, d_1 = 1)$ is equal to 1. In this case the implementation of the matching method is similar to the case of a single binary program, taking into account that the comparison group should exclude those who participate in the program d_2.

Similar to the case of a single program, to identify ATT using the matching method we require that the difference in the expectation of potential outcomes between the participants and non-participants are the same before and after the program given the variables X and d_2:

Assumption 4.1:

$$\begin{aligned}
&E(Y_{0B} \mid X, d_1 = 1, d_2 = 1) - E(Y_{0B} \mid X, d_1 = 0, d_2 = 1) = E(Y_{01} \mid X, d_1 = 1, d_2 = 1) - E(Y_{01} \mid X, d_1 = 0, d_2 = 1) \\
&E(Y_{0B} \mid X, d_1 = 1, d_2 = 0) - E(Y_{0B} \mid X, d_1 = 0, d_2 = 0) = E(Y_{00} \mid X, d_1 = 1, d_2 = 0) - E(Y_{00} \mid X, d_1 = 0, d_2 = 0)
\end{aligned} \qquad (A.4.1)$$

To estimate the program impact by matching, it is required that there be remaining people who do not participate in the program d_1 but have similar distribution of the variables X given the treatment variable d_2. This is the common support assumption, and is stated formally as follows:

Assumption 4.2:

$$\begin{aligned}
&0 < P(d_1 = 1 \mid X, d_2 = 0) < 1 \\
&0 < P(d_1 = 1 \mid X, d_2 = 1) < 1
\end{aligned} \qquad (A.4.2)$$

Proposition 4.1: Under assumptions (A.4.1) and (A.4.2), the conditional and unconditional parameters $ATT_{(X)}$ and ATT for the program d_1 are identified.

Proof:

---

[5] For simplicity, the subscript "A" is dropped.

[6] For simplicity in denotation, we denote variables $\{X_B, X_A\}$ as X.

Similar to the proof of the proposition 3.3, substitute two equations in (A.4.1) to (4.2) to identify ATT1$_{(X)}$:

$$ATT1_{(X)} = E(Y_1 \mid X_B, X_A, d_1 = 1) - E(Y_0 \mid X_B, X_A, d_1 = 1)$$

$$= \left\{ \begin{array}{l} \left[E(Y_{11} \mid X, d_1 = 1, d_2 = 1) - E(Y_{01} \mid X, d_1 = 0, d_2 = 1) \right] \\ - \left[E(Y_{0B} \mid X, d_1 = 1, d_2 = 1) - E(Y_{0B} \mid X, d_1 = 0, d_2 = 1) \right] \end{array} \right\} \Pr(d_2 = 1 \mid X, d_1 = 1) \tag{4.3}$$

$$+ \left\{ \begin{array}{l} \left[E(Y_{10} \mid X, d_1 = 1, d_2 = 0) - E(Y_{10} \mid X, d_1 = 1, d_2 = 0) \right] \\ - \left[E(Y_{0B} \mid X, d_1 = 1, d_2 = 0) - E(Y_{0B} \mid X, d_1 = 1, d_2 = 0) \right] \end{array} \right\} \Pr(d_2 = 0 \mid X, d_1 = 1)$$

The unconditional parameter, ATT is identified because of (2.5).■

To estimate the program impacts, the non-participants in the program d_1 will be matched to the participants in the program d_1 based on the closeness of the distance between the variables to construct the comparison group. The matching is performed for people who have the same program variable d_2, i.e. the participants and matched non-participants have the same participation statuses in the program d_2.

For a participant i, denote n_{ic} as the number of non-participants j who are matched with this participant, and $w(i,j)$ is weight is attached to the outcome of each non-participant. These weights are defined non-negative and sum up to 1, i.e.:

$$\sum_{j=1}^{n_{ic}} w(i,j) = 1 \tag{4.4}$$

Weight can be equal weights, e.g. as in n-nearest neighbor matching or different weights e.g. kernel matching and local linear regression matching.

For those who do not participate in the program d_2 (i.e. $d_2 = 0$), the difference in outcome between the participants and matched non-participants is given by:

$$A\hat{T}T1_{(X=x, d_2=0)} = \frac{1}{n_{x1}} \sum_{d_2=0, X_i=x} \left\{ \left[Y_{i1A} - \sum_{j=1}^{n_{ic}} w(i,j) Y_{j0A} \right] - \left[Y_{i1B} - \sum_{j=1}^{n_{ic}} w(i,j) Y_{j0B} \right] \right\} \tag{4.5}$$

Where:

- n_{X1} is the number of those who have $d_1 = 1; d_2 = 0; X = x$.
- Y_{i1A} and Y_{j0A} are the observed outcomes of participant i and non-participant j with $X = x$ after the program.
- Y_{i1B} and Y_{j0B} are the observed outcomes of participant i and non-participant j with $X = x$ before the program.

Similarly, we have the estimator $A\hat{T}T_{(X=x, d_2=1)}$, and the estimator $A\hat{T}T_{(X=x)}$ is:

$$A\hat{T}T1_{(X=x)} = \frac{1}{n_{x1} + n_{x2}} \left\{ n_{x1} A\hat{T}T1_{(X=x, d_2=0)} + n_{x2} A\hat{T}T1_{(X=x, d_2=1)} \right\} \tag{4.6}$$

where:

- n_{X2} is the number of those who have $d_1 = 1; d_2 = 1; X = x$

The estimators of unconditional parameter are:

$$A\hat{T}T1 = \frac{1}{\sum I\{d_1 = 1; x \in S_X\}} \sum_{x \in S_X} A\hat{T}T1_{(X=x)} \qquad (4.7)$$

Where *I{}* is an indicator function that is equal to 1 if the value of {} is true, 0 otherwise, S_X is sample space of the *X* variables.

4.2. The Case of Multiple Overlapping Programs

Now suppose that there are m programs that are assigned to subjects in population *P*. Denote participation in the programs by a vector variable *D*:

$$D = (d_1, d_2, ..., d_m) \ . \qquad (4.8)$$

where d_k is a variable that equals 1 if she participates in program *k*, and 0 otherwise. Subjects who do not participate in any program will have the value of the vector *D* equal to $D = (0,0,...,0)$. In contrast, subjects who participate in all the programs will have the value of the vector D equal to $D = (1,1,...,1)$. The set of the potential treatments has 2^m values:

$$\Omega_D = \left\{ \begin{pmatrix} 0 \\ 0 \\ . \\ . \\ 0 \end{pmatrix}; \begin{pmatrix} 1 \\ 0 \\ . \\ . \\ 0 \end{pmatrix}; ...; \begin{pmatrix} 1 \\ 1 \\ . \\ . \\ 1 \end{pmatrix} \right\}. \qquad (4.9)$$

Before the program implementation, the outcome variable is observed and denoted by Y_{0B}. After program implementation, corresponding to each value of the vector variable D, there is a potential outcome, denoted by $Y^P_{(D)}$.[7] Thus for each subject, there are 2^m potential outcomes. However we are able to observe only one outcome of those, depending on the realization of the vector variable D.

In general, one can estimate impact of a treatment state $D = D_g$ relative to a treatment state $D = D_h$:

$$ATTgh_{(X)} = E(Y^P_{D=D_g} \mid X, D = D_g) - E(Y^P_{D=D_h} \mid X, D = D_g). \qquad (4.10)$$

However, explanation of (4.10) is complicated and less practical. For simplicity, we focus on impact of a particular program, e.g., program d_k. The impact parameter of program *k* is defined:

$$ATTk_{(X)} = E(Y^P_{d_k=1} \mid X, d_k = 1) - E(Y^P_{d_k=0} \mid X, d_k = 1). \qquad (4.11)$$

It should be noted that in (4.10) and (4.11) the *X* variable denotes variables $\{X_B, X_A\}$ for simplicity in formulas. The identification assumptions for the difference-in-differences with matching method in the case of multiple overlapping programs are extended as follows:

Assumption 4.3:

[7] For simplicity, the subscript "A" is dropped.

$$E(Y_{0B} \mid X, \overline{D}, d_k = 1) - E(Y_{0B} \mid X, \overline{D}, d_k = 0) = E(Y_{\overline{D}, d_k = 0} \mid X, \overline{D}, d_k = 1) - E(Y_{\overline{D}, d_k = 0} \mid X, \overline{D}, d_k = 0),$$

where $\overline{D} = D \setminus d_k$ i.e. $\overline{D}$ does not include d_k. (A.4.3)

The matching method requires the assumption on common:

Assumption 4.4: $0 < P(d_k = 1 \mid X, \overline{D}) < 1$. (A.4.4)

$P(d_k = 1 \mid X, \overline{D})$ is the conditional probability of participating in program d_k given the variables X and other program variables. It is required that there be still subjects who do not participate in program d_k but have the same variables X and participation statuses of the other programs (not include program d_k) as those of the participants of program d_k.

Proposition 4.2: Under the assumptions (A.4.3) and (A.4.4), $ATTk_{(X)}$ and $ATTk$ for program d_k are identified by the difference-in-differences with matching method.

Proof:

Similar to (4.3), the $ATTk_{(X)}$ is written as follows:

$$\begin{aligned} ATTk_{(X)} &= E(Y_{d_k=1} \mid X, d_k = 1) - E(Y_{d_k=0} \mid X, d_k = 1) \\ &= \sum_{D_g \in \Omega_D} \left\{ \left[E(Y_{\overline{D}, d_k=1} \mid X, \overline{D}, d_k = 1) - E(Y_{\overline{D}, d_k=0} \mid X, \overline{D}, d_k = 1) \right] \Pr(\overline{D} \mid X, d_k = 1) \right\}, \end{aligned} \quad (4.12)$$

There are unobserved terms in (4.12) i.e., $E(Y_{\overline{D}, d_k = 0} \mid X, \overline{D}, d_k = 1)$. However, under assumptions (A.4.3) and (A.4.4), we have:

$$\begin{aligned} ATTk_{(X)} &= E(Y_{d_k=1} \mid X, d_k = 1) - E(Y_{d_k=0} \mid X, d_k = 1) \\ &= \sum_{\overline{D} \in \Omega_D} \left\{ \begin{pmatrix} \left[E(Y_{\overline{D}, d_k=1} \mid X, \overline{D}, d_k = 1) - E(Y_{\overline{D}, d_k=0} \mid X, \overline{D}, d_k = 0) \right] \\ - \left[E(Y_{0B} \mid X, \overline{D}, d_k = 1) - E(Y_{0B} \mid X, \overline{D}, d_k = 0) \right] \end{pmatrix} \Pr(\overline{D} \mid X, d_k = 1) \right\}, \end{aligned}$$

This parameter is identified since all the terms are observed. The unconditional parameters are also identified by formulas (2.5).■

To estimate the program impacts, the participants of program d_k will be matched to the non-participants based on the closeness of the distance in the variables X before and after the program implementation. In addition, the matching is performed for people who have the same program statuses D (except program d_k). The estimator of the $ATTk_{(X)}$ has a similar form as in the case of two programs, i.e., formula (4.6), in which the sample mean outcomes of the participants are estimators of $E(Y_{\overline{D}, d_k = 1} \mid X, \overline{D}, d_k = 0)$, and the sample mean outcomes of the matched non-participants are estimators of $E(Y_{\overline{D}, d_k = 0} \mid X, \overline{D}, d_k = 1)$ (before and after the program).

4.3. Matching Using the Propensity Scores

To perform the matching using propensity scores, Proposition 3.2 is extended to the case of multiple overlapping programs as follows:

Proposition 3.5: $Y^P_{(D)} \perp D \middle| X \Rightarrow Y^P_{(D)} \perp D \,|\, b(X)$,

where $b(X)$ is any function such that $P(D \mid X) = f[b(X)]$.

Since we focus on impact of a program of interest, e.g., program d_k, we will state the proposition in a different way which emphasizes a program of interest.

Proposition 3.6: $Y^P_{(D)} \perp d_k \middle| X, \overline{D} \Rightarrow Y^P_{(D)} \perp d_k \middle| b_k(X, \overline{D})$,

where:

$\overline{D} = D \backslash d_k$ i.e. $\overline{D}$ does not include d_k,

$P(d_k = 1 \mid X, \overline{D}) = E(d_k \mid X, \overline{D})$,

$b_k(X, \overline{D})$ is any function such that $P(d_k = 1 \mid X, \overline{D}) = f[b_k(X, \overline{D})]$.

Proof:

It is equivalent to show that (Dawid, 1979):

$$P\left[d_k = 1 \mid Y^P_{(D)}, b_k(X, \overline{D})\right] = P\left[d_k = 1 \mid b_k(X, \overline{D})\right]. \tag{4.13}$$

The following manipulations using law of iterated expectation:

$$\begin{aligned} P[d_k = 1 \mid Y^P_{(D)}, b_k(X, \overline{D})] &= E[d_k \mid Y^P_{(D)}, b_k(X, \overline{D})] \\ &= E\left\{E[d_k \mid Y^P_{(D)}, X, \overline{D}, b_k(X, \overline{D})] \middle| Y^P_{(D)}, b_k(X, \overline{D})\right\} \\ &= E\left\{E[d_k \mid Y^P_{(D)}, X, \overline{D}] \middle| b_k(X, \overline{D})\right\} \\ &= E\left\{E[d_k \mid X, \overline{D}] \middle| b_k(X, \overline{D})\right\} \\ &= E\left\{P(d_k = 1 \mid X, \overline{D}) \middle| b_k(X, \overline{D})\right\} \\ &= P[d_k = 1 \mid b_k(X, \overline{D})]. \end{aligned} \tag{4.14}$$

The propensity score is usually selected as the balancing score. The above two propositions 3.5 and 3.6 suggest two ways to estimate the propensity score. The first is to estimate propensity score for the treatment variable D, i.e. $P(D \mid X)$ by a multinomial model. The second is to estimate the propensity score for the program d_k conditional on the variables X and $\overline{D}$. If we are interested in a particular program, it is more convenient and easy to estimate the probability of participation in the program given the variables X and $\overline{D}$.

5. Impact of Formal and Informal Credit in Vietnam

This section illustrates the impact estimation of the borrowing from formal and informal credit in Vietnam. Some households can borrow from both the formal and informal credit sources. Thus the borrowing from the formal and informal credit sources can be regarded as two overlapping programs, and the method of difference-in-differences with matching can be applied to measure impact of the borrowing.

5.1. Data Source

The study relies on data from the two Vietnam Household Living Standard Surveys (VHLSS) to analyze the poverty targeting and impact of the formal and informal credit in Vietnam. The surveys were conducted by the General Statistical Office of Vietnam (GSO) with technical support from World Bank in the years 2002 and 2004. Information on household characteristics is collected using detailed household questionnaires. The collected information includes basic demography, employment and labor force participation, education, health, income, expenditure, housing, fixed assets and durable goods, the participation of households in poverty alleviation programs, and especially information on credit that households had borrowed during the past 12 months before the year 2004.

The 2002 and 2004 VHLSSs sampled 29530 and 9188, respectively. The samples are representative for the whole country and 8 geographic regions. It is very interesting that these samples of VHLSS 2002 and 2004 construct a panel data set of 4008 households, which is representative for the whole country, and regions of large population.

5.2. Formal and informal credit in Vietnam

It is often argued that micro-credit is an important tool for smoothing consumption and promoting production, especially for the poor households (e.g. Zeller, et. al. 1997; Conning and Udry, 2005). In Vietnam there are alternative sources of credit that a household can borrow from. Among the formal credit institution, the Bank for Agriculture and Rural Development (BARD) is the largest lender. The Vietnam Bank for Social Policies (VBSP) is a State bank which is targeted at the poor households. For the informal credit sources, friends and relatives are important lenders for the households in Vietnam.

Table 1: Households borrowing from formal and informal credit sources

		Households not borrowing from informal credit sources	Households borrowing from informal credit sources	Total by column
Households not borrowing from formal credit sources	%	60.3	14.6	74.9
	No. obs.	2416	585	3001
Households borrowing from formal credit sources	%	20.7	4.4	25.1
	No. obs.	831	176	1007
Total by row	%	81.0	19.0	100
	No. obs.	3247	761	4008

Note: the percentages are estimated using the sampling weights of the 2004 VHLSS.
Source: Estimation from the 2004 VHLSS.

Table 1 shows that 25% and 19% of households borrow from the formal and informal credit sources, respectively. About 4.4% of households can have access and borrow from both the formal and informal credit sources.

It should be noted that panel data from the 2002 and 2004 VHLSSs are used in the difference-in-differences with matching method. Data from the 2002 VHLSS are considered as baseline data. Thus only loans which were obtained by households between 2002 and 2004 are included in Table 1.

5.3. Impact Estimation of Formal and Informal Credit

The first step in the measuring impact is to predict the probability of receiving credit between the year 2002 and 2004 for all households in the sample. Since the dependent variable is binary, a logit regression is often used. The main problem in the estimation is how to select explanatory variables. All variables that are exogenous to the credit borrowing and expected to affect the credit borrowing as well as outcomes should be included in the model. Variables pre-program are clearly unaffected by the program implementation. Conditioning variables used include: the regional variables; household demography such as household size, percentage of the elderly and children; education and main job of the head; ratio of working people in 2002; saving, foreign and domestic remittances and household asset in 2002. Variables of the number of sick days and sick persons in 2002 and 2004 are also included, since a part of credit is used for healthcare treatment.

It should be noted that the usage of the predicted propensity score is mainly aimed to overcome the multidimensionality problem of matching by covariates. The quality of a constructed comparison group should be assessed by testing whether the distribution of characteristics covariates is similar between the comparison and treatment groups given the predicted propensity score. In this research two types of test are performed to examine the similarity of covariates between the matched non-participants and participants. The first is simply the test for the mean equality of covariates between the treatment and comparison groups. The second is the test for the mean equality of covariates within strata of the predicted propensity score.[8] If there exists a covariate not balanced in many strata, e.g. three strata, the comparison group should be reconstructed by modifying the logit model of propensity score.[9] Figures 1 and 2 in Appendix graph the propensity score for the recipients and non-recipients of formal and informal credit.

Tables 2 and 3 present the estimation of the program impact measured by the parameter ATT. Three matching schemes are used, namely the 1 nearest-neighbor, 3 nearest-neighbors, and kernel matching with bandwidth equal to 0.01. The standard errors are estimated using nonparametric bootstrap with 500 replications.

Table 2: Impact of Formal Credit

Matching schemes	Outcome variables (thousand VND)			
	Expenditure per capita	Household fixed assets	Household durables	Income per capita
1 nearest neighbor	224.7	9833.9**	574.2*	297.1

[8] The method of testing the equality in mean of covariates within stratum is proposed by Dehejia and Wahba (2002). They perform the test for all the participants and non-participants after estimating the propensity score. In this research the test is applied for the treatment and comparison groups after they are matched. Since what we need is the similarity of covariates between the treatment and comparison groups.

[9] The logit regression results are not presented in this paper, but they can be provided on request.

matching	(165.9)	(4902.6)	(300.6)	(574.8)
3 nearest neighbors	189.0	6100.2**	393.8*	350.1
matching	(143.6)	(3094.4)	(254.3)	(458.9)
Kernel matching with	129.5	5968.7**	430.9*	340.3
bandwidth of 0.01	(113.5)	(3047.8)	(229.4)	(357.1)

Standard errors are calculated using bootstrap with 500 replications (in parentheses).
* significant at 10%; ** significant at 5%; *** significant at 1%
Source: Estimation from the 2002-2004 VHLSSs

Table 2 shows that the formal credit has positive impact on fixed assets and durable assets of the borrowing households. The results are similar between the three matching scheme. However, the impacts on expenditure and income per capita are not statistically significant. This can be because the period 2002-2004 is quite short, and the effect of credit on income and expenditure is not clear.

Table 3 presents impact of informal credit on household welfare. It shows that informal credit does not have any statistically significant estimate of ATT on all the household outcomes.

Table 3: Impact of Informal Credit

Matching schemes	Outcome variables (thousand VND)			
	Expenditure per capita	Household fixed assets	Household durables	Income per capita
1 nearest neighbor	208.5	7533.1	389.2	388.2
matching	(180.8)	(4979.9)	(415.6)	(456.1)
3 nearest neighbors	157.5	4449.7	29.8	309.3
matching	(156.1)	(4260.8)	(404.6)	(366.8)
Kernel matching with	116.8	3250.5	40.7	335.6
bandwidth of 0.01	(144.1)	(3999.5)	(389.7)	(308.2)

Standard errors are calculated using bootstrap with 500 replications (in parentheses).
* significant at 10%; ** significant at 5%; *** significant at 1%
Source: Estimation from the 2002-2004 VHLSSs

6. Conclusion

Traditional literature on program impact evaluation often deals with a single binary program. In reality, some people can join several programs at the same time. This paper discusses the difference-in-differences with matching method in a general context in which people may participate in several programs simultaneously. The parameter of interest is the Average Treatment Effect on the Treated (ATT). It is shown that impact of a program can be measured as a weighted average of impacts of the program on groups with various program statuses, which are estimated by the difference-in-differences with matching method.

The method is illustrated by measuring impacts of formal and informal credit in Vietnam. The panel data used for the impact estimate are from two Vietnam Household Living Standard Surveys in 2002 and 2004. It is shown that formal credit has positive and statistically significant impact estimate on the fixed and durables assets of the borrowing households. However, there is no impact of formal credit found on income and consumption expenditure of the borrowing households. For the informal credit, the impact estimates are not statistically significant for all the four outcomes of interest.

References

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

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)*, Vol. 128, No. 2. (1965), pp. 234-266.

Conning, J. and Christopher U. (2005). Rural Financial Markets in Developing Countries, Economic Growth Center, Yale University, Center Discussion Paper No. 914.

Dawid, A. P. (1979). Conditional Independence in Statistical Theory. *J. R. Statist. Soc.*, 41, No. 1: 1-31.

Dehejia, R. H. and Wahba S. (1998). Propensity Score Matching Methods for Non Experimental Causal Studies. NBER Working Paper 6829, Cambridge, Mass.

Frölich, M. (2002). Program Evaluation with Multiple Treatments. Discussion Paper 2002-17, Department of Economics, University of St. Gallen.

Heckman, J., Lalonde, R., and Smith, J., (1999). The Economics and Econometrics of Active Labor Market Programs. *Handbook of Labor Economics, Volume 3*, Ashenfelter, A. and D. Card, eds., Amsterdam: Elsevier Science.

Heckman, J., Ichimura H. and Todd P. (1997). Matching as an Econometric Evaluation Estimators: Evidence from Evaluating a Job Training Programme. *Review of Economic Studies*, 64 (4), 605-654.

Imbens, G. (1999). The Role of the Propensity Score in Estimating Dose-Response Functions NBER Technical Working Paper 237.

Lechner, M. (2001). Identification And Estimation of Causal Effects of Multiple Treatments under the Conditional Independence Assumption (In Lechner, M. and Pfeiffer, F. (Eds.), *Econometric Evaluation of Labour Market Policies.* Heidelberg: Physica-Verlag.)

Quandt, R. (1972). Methods for Estimating Switching Regressions. *Journal of the American Statistical Association,* 67(338):306-310.

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.

Zeller, M., A. Diagne, and C. Mataya (1997). Market Access by Smallholder Farmers in Malawi: Implications for Technology Adoption, Agricultural Productivity, And Crop Income. *Agricultural Economics*, 19 (1 – 2): 219 – 229.

Appendix

Figure 1: Density of propensity score for recipients and non-recipients of formal credit

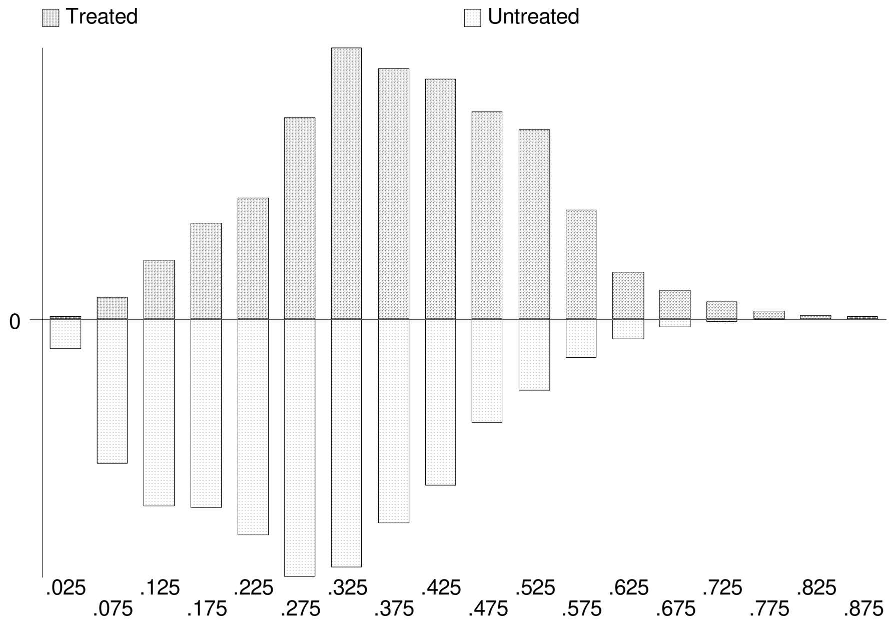

Figure 2: Density of propensity score for recipients and non-recipients of formal credit