◆◆ Scientific
◆◆ Research

Rural Labor Force Transfer Training Effects Evaluation by Matching Methods: Evidence from Yunnan Province of China

Jiachun Xie, Xingxu Li*

School of Statistics and Mathematics, Yunnan University of Finance and Economics, Kunming, China
Email: *xingxu1967@sina.com

ABSTRACT

Rural labor force transfer training is one of important poverty alleviation measures in China. This paper describes training participation situation and evaluates training effects by matching methods in the case of coastal autonomous minority nationality areas of Yunnan province by using 2053 rural household data. The result shows that the average training participation from 2006 to 2008 is 26.39 percent. In addition, ATE is 18.33 percent, TT is 18.99 percent, TUT is 18.09 percent. And, TUT < ATE < TT . It demonstrates that the rural labor force transfer training program is effective and well-directed in coastal autonomous minority nationality areas of Yunnan province of China.

Keywords: Coastal Autonomous Minority Nationality Areas; Treatment Effect; Training Effects Evaluation; Matching Evaluation

1. Introduction

Western regions are the main poverty areas in China, especially the autonomous minority nationality rural areas in the southwestern frontier regions. According to Poverty Monitoring Report of Rural China [1], there were 23.72 million people in poverty in western areas in 2009, which account for 65.9 percent of the country (35.97 million), with only the three of the 31 western provinces: Guizhou, Yunnan and Gansu, exceeding 3 million people in poverty respectively. Incidence of poverty in western regions was 8.3 percent in 2009, an increase of 4.5 percent from the national poverty incidence of 3.8 percent for the same year. In 2009, net income per capita for the western rural poverty population was CNY983, which accounted for 19.1 percent of the total rural population throughout the country.

The autonomous minority nationality rural region had 19.54 million people in poverty in 2009, which accounted for 54.3 percent of all rural people in poverty. The poverty incidence was 16.4 percent in these areas, and 12.6 percent higher than the country poverty inci-

dence. The autonomous minority nationality rural areas were more severely poor than any other areas of the country at the same time period. One of the minority nationality areas is Yunnan province.

Yunnan is the most southwestern province and shares a border of 4060 kilometers with Burma in the west, Laos in the south, and Vietnam in the southeast. It is noted for a very high level of ethnic diversity which has the second highest number of ethnic groups among the provinces in China. Because poverty is widespread in Yunnan and deep-rooted, and the geographical location is special, Yunnan becomes one of the most important poverty alleviation provinces, especially in its costal autonomous minority nationality areas.

A key strategy of poverty alleviation the Chinese government employs is to provide training to the rural largely unskilled working population. In March of 2004, "Suggestions on Accelerating Farmers' Income by the Party Central Committee and the State Council" was issued and recommended enhancement of the rural labor force through vocational skills training. Thereafter, a series of projects were initiated under the cooperation of the Ministry of Labor and Social Security, the Ministry

---

*Corresponding author.

of Agriculture, the Ministry of Finance, and the Ministry of Education, such as, the Sunshine Project, the Rural Labor Skills Training and Employment Project, and the Rain and Dew Project.

Since the micro survey data are scarce and hard to be obtained, there has been very little evaluation of the rural training programs domestically. Our aim of this study is to evaluate the effects of the rural labor force transfer training program by using Matching methods. The research targets are the Honghe Hani Nationality and Yi Nationality Autonomous Prefecture (Honghe Prefecture) and the Dehong Dai Nationality and Jingpo Nationality Autonomous Prefecture (Dehong Prefecture) of Yunnan province in southwest of China.

This paper is organized as: Section 2 is training participation and income description. A series of matching methods will be introduced in Section 3. Section 4 is empirical results. The last Section is concluding remarks.

2. Training Participation and Income Description

2.1. Data Sources

The data for our study were obtained from the China's Rural Household Survey (RHS) of Honghe Prefecture and Dehong Prefecture done by local State Statistical Bureau (SSB) offices in Yunnan province. The data appears to be of good quality and bunches of information about rural household income, consumption, production, accumulative and social behaviors. Two-stage sample was selected in each prefecture. The first stage involved the selection of 348 villages from 13 counties of Honghe Prefecture and 5 counties of Dehong Prefecture. In the second stage, it was involved the stochastic sampling of households from the selected villages. There were two main methods have been adopted for collecting data. One is the sampled households fill in a daily diary on expenditures and other relative information. Another one was visiting survey. Sampled households were visited on every month by an interviewer to check the diaries and collect data. Our sample data is a sequence of cross Sections since 2006 and is ongoing. The training information enrolled into this survey only from 2006 to 2008. The survey observations updated every year, tracked the households who appeared in this survey started from 2006 to 2008 continuously, there were 2280 households.

2.2. Training Participants and Their Income

There were 595 (26.10%), 612 (26.84%) and 598 (26.23%) households that had been trained during 2006 to 2008 respectively, the sample average training participation ration is 26.39 percent. In 2007, there were another new 111 households participated and another

new 41 households participated in 2008.

Income in this study is net income. In 2008 net income per capita of participants was CNY3841.77 while for non-participators was CNY2690.49. Households participating in training increased the net income almost CNY1000 than those who did not participate in training during these three years. The result shows that household who attended training benefited a lot.

Here the question is that income difference is a result, reasons for income difference could be training, or be other behaviors. It means that statistical description cannot tell whether there has causality between training and income difference. Therefore, it is necessary for us to use Matching methods which are special for cause-and-effect relationship research to estimate training effects in the case of coastal autonomous minority nationality areas of Yunnan province in China.

3. Methodology

The primary question for training programs effects evaluation is: what is the difference between participants' post-program income and the income that they would have received had they not participated in training [2]. In practice, it is quite difficult to answer this question straightforward. Suppose that there is a target population N has being studied. If we take T denotes the training status, $T = 1$ if a farm household participates in training, which is also say a household is treated. Here a household participates in training means any one family member participates in training. $T = 0$, denotes a household not participates in training, that is none family member participates. We are interested in income outcome Y and further denote Y_1 as the potential income of participants, Y_0 for non-participator. $Y_1 - Y_0$ is the treatment effect of the training. The difficulty for effects evaluation is, for a given household, that we observing either Y_1 or Y_0 at the same time, but not both, this also called missing data problem. In order to overcome this difficulty, we need to structure a counterfactual frame of causality which can be composed by untreated group members and shared with similar observable characteristics of those who are actually treated. Various methods have been employed to solve the above evaluation difficulty, for example, Instrument Variables (IVs) methods [3]; Marginal Treatment Effect (MTE)-based parametric estimations methods and MTE-based semi-parametric estimation methods [4-10]; a series of Matching methods [11-17].

The evaluation method of Matching has been used in many fields since it is easy to understand and easy to apply [3,15,17]. Over the previous literature, there are mainly three popular Matching methods, multivariate Matching based on Mahalanobis Distance (MD) [18-21],

Propensity-Score (PS) Matching [11], and Genetic Matching (GenMatch) [15,17,22]. In this research, we will employ these three popular Matching methods to evaluate training effects, and answer the following questions: a) what is the average treatment effect (ATE) for the target population; b) what is the treatment effect of the treated (TT); c) what is the treatment effect of the untreated (TUT); d) which Matching method can reduce bias mostly.

3.1. Parameters of Interests

There are three mean treatment effect parameters:

$$ATE = E(Y_1 - Y_0)$$

$$TT = E(Y_1 - Y_0 | T = 1)$$

$$TUT = E(Y_1 - Y_0 | T = 0)$$

The Average Treatment Effect (*ATE*) is defined for the whole population. ATE evaluates the average difference between a set of members in N that are randomly selected for treatment and another set of members that are randomly selected for control.

The Treatment effect of the Treated (*TT*) means to the average difference by treatment status for these people who are treated.

The Treatment effect of the Untreated (*TUT*) refers to the average difference by treatment status for these who are not treated.

3.2. Matching Based on Mahalanobis Distance Method

As is discussed before, the main difficulty of treatment effect evaluation is the missing data problem and we need to structure a counterfactual frame to overcome this problem. Matching is an excellent tool to structure a counterfactual by filling the missing data for each observation which is similar in terms of their observable characteristics and relies on the Conditional Independence Assumption (CIA) [23] which also has other names, "unconfoundness" or "ignorability" [11], and "exogeneity" [24]. If we let X be a vector of observed covariates, such as education level or whether living in a minority group village in this case, selection to participate in training is independent of potential outcomes, the CIA states:

$$(Y_1 - Y_0) \perp T | X$$

If CIA holds true, the above three parameters of interest ATE, TT and TUT can be expressed as:

$$ATE(X) = E(Y_1 - Y_0 | X)$$

$$TT(X) = E(Y_1 - Y_0 | X, T = 1)$$

$$TUT(X) = E(Y_1 - Y_0 | X, T = 0)$$

Till now, we can introduce multivariate matching which is based on Mahalanobis Distance to structure a counterfactual for each observation [25]. The Mahalanobis Distance between any two column vectors is:

$$MD(X_i, X_j) = \left\{ (X_i - X_j)^T S^{-1} (X_i - X_j) \right\}^{\frac{1}{2}}$$

where S is the sample covariance matrix of X. To estimate *TT* by matching with replacement, one matches each treated unit with the M closest control units, as defined by this distance measure, $MD(X_i, X_j)$. Under this method, the estimates will suffer bias when X consists of more than one continuous variable, which is equivalent to that the multivariate matching results in statistically efficient estimates of the treatment effect only when continuous variables are limited to one [13]. Additional continuous covariates will cause increasingly biased estimates.

A recommended alternative solution for more than one continuous covariate is known as the Propensity Score Matching method.

3.3. Propensity Score Matching Method

According to Rosenbaum and Rubin [11,26,27], the propensity score which means the probability of receiving treatment conditional on covariates X can reduce the dimensionality of the multivariate matching. Let $p(X_i)$ be the probability of a unit i being treated given X_i, which is a household participated in training in this case, $p(X_i)$ can be defined as:

$$p(X_i) = Pr(T = 1 | X_i) = E(T | X_i)$$

Given

$$0 < Pr(T = 1 | X_i) < 1$$

and

$$Pr(T_1, T_2 \cdots T_N | X_1, X_2 \cdots X_N)$$
$$= \prod_{i=1}^{N} p(X_i)^T (1 - p(X_i))^{(1-T)}$$

Rosenbaum and Rubin [11] proved that:

$$ATE(p)$$
$$= E\left\{ E(Y_1 | p(X_i), T = 1) - E(Y_0 | p(X_i), T = 0) \right\}$$

$$TT(p)$$
$$= E\left\{ E(Y_1 | p(X_i), T = 1) - E(Y_0 | p(X_i), T = 0) | T = 1 \right\}$$

The PS Matching differs the matching rule from the MD matching. PS Matching process involves matching as a function of propensity score, which is matching each treated member to the nearest control member on the unidimensional metric of the propensity score vector [28].

If a treatment observation matched with a control observation by matching on a correctly specified propensity score, that will asymptotically balance the observed covariates, and will asymptotically remove the bias conditional on such covariates [11,17]. By covariate balance it means that the treatment and control observations have the same joint distribution of observed covariates [17]. However, in practice, the correct $p(X_i)$ is unknown, so it must be estimated. Normally, $p(X_i)$ can be estimated by using *Probit* or *Logit* regression, here we choose *Logit* regression to estimate $p(X_i)$.

The MD Matching and PS Matching can be used alone, or in a way of combination. In this study, we set up different models, Model-1 for the MD Matching, Model-2 for the PS Matching and Model-3 combined the MD Matching with the PS Matching.

3.4. Genetic Matching Method

Genetic Matching (GenMatch) was proposed by Sekhon [17], Diamond and Sekhon [29], with a genetic search algorithm. Since the MD Matching is good at minimizing the distance between treatment and control observations but may fails optimal balance in a given dataset. Therefore, the MD Matching can be extended in a more generalizing method—GenMatch by including an additional weight matrix in the MD matrix:

$$D(X_i, X_j) = \left\{ (X_i - X_j)^T (S^{-1/2})^T W S^{-1/2} (X_i - X_j) \right\}^{1/2}$$

where W is a $k \times k$ positive definite weight matrix and $S^{1/2}$ is the Cholesky decomposition of S, *i.e.*, $S = S^{1/2}(S^{1/2})^T$ which is the variance-covariance matrix of X. All elements of W are zero except those down the main diagonal and k parameters must be chosen as the components of the main diagonal. It is easy to see that if each of those k parameters is equal to one, $D(\cdot)$ is the same as $MD(\cdot)$. Sekhon recommend that if one can estimate the propensity score correctly, it should be included as one of the covariates in GenMatch [28]. In this case, X in $D(\cdot)$ might be extended to X^*, which is a new matrix consisting of the propensity score $p(X_i)$ and X. It is not hard to understand that Gen-Match will be equivalent to PS Matching given a zero weight to covariate X_i. Thereby, both the PS Matching and the MD Matching are special cases of GenMatch.

In GenMatch, the genetic search algorithm automates the iterative process by checking and improving balance for each covariate or minimizing imbalance by minimizeing the largest observed covariate discrepancy. Thereby, the imbalance should be small after the optimal matching and it can be measured in a series of methods, such as the nonparametric Kolmogorov-Smirnov(KS)-test statistics and paired t-test and the smallest p-values from KS-tests

and t-test, which are need to be large.

In this study, we also set up Model-4 for the Gen-Match without propensity score. For each model, we report the covariate balance results to show the effectiveness of each matching method.

4. Empirical Results

4.1. Measures

Based on Mincer income model [30], not only human capital investment has been taken into consideration but also material resources capital investment, farm household characteristics and living village characteristics. We take households who participate in training as the treated group, who not participate in training as the control group. In our research, per capita annual net income Y of 2007 has been selected as our outcome variable and been taken logarithm. We did some data processing work before treatment effect evaluation. Since the income variable need to be taken logarithm, we save the households who with positive annual net income. Finally, totally 2053 households who with positive annual net income and participated in training only in 2006 but not in 2007 and 2008 have been selected in our research, in the consideration of the causal inference of training. In addition, there are 570 households in the treated group and 1483 households in the control group. The following **Table 1** is variables definition.

4.2. Statistical Analysis

In this research, R software of version 3.0.0 and Matching package are used for the analysis. Firstly, the sample statistical description and the mean differences are calculated with an unpaired Welch Two Sample t-test and are presented in **Table 2**. For the purpose of finding out the most optimal match for each treated observation, we set up a series models, Model-1 for the MD Matching, Model-2 for the PS Matching, Model-3 for the combination of MD and PS Matching and the last Model-4 for the GenMatch without PS. In each model we match with replacement and one-to-one match because allowing replacement reduces bias.

As is shown in **Table 2**, totally 27.6 percent modeled sample households participated in training in 2006. Per capita annual net income of treated group is significant higher than which of control group observations and increases from 2006 to 2007. For the highest education level, 56.5 percent of training participants attended junior high school but 47.1 percent for the control group and their difference is 9.4 percent which is significant. We can obtain abundant of information from **Table 2**, but the most important thing is, the pre-match mean differences are statistical significant between the treated group and

Table 1. Variables definition.

Variables	Definition
Y	Per capita annual net income in 2007 and take logarithm
Lin06	Per capita annual net income in 2006 and take logarithm
Edu	Highest education levels, Edu5 for college and above, Edu4 for special technical school or senior high school, Edu3 for junior high school, Edu2 for and Edu1 for illiteracy
T	Binary variable for participation in training, 1 indicate participation, 0 otherwise
Lnassets	Capital assets of farming per capita, take logarithm
Cadre	Binary variable, 1 indicate household with village official family members
Emlabor	Number of labor force that is migrating into other sectors or other areas
Burden	Calculated by (Residents- Labor)/Labor
Min	Binary variable, 1 indicate a minority group village, 0 otherwise
Geo	Topography indicator, Geo1 for plains, Geo2 for hills and Geo3 for mountains
Dis	Distance between village and county, Dis5 for distance more than 20 kilometers, Dis4 for distance between 10 and 20 kilometers, Dis3 for distance 10 kilometers, Dis2 for distance between 2 and 5 kilometers, Dis1 for less than 2 kilometers
Vitr06	Average ratio of the participants at a village level

Table 2. Sample statistical description and mean differences.

Variables	Total Sample		Treated Group		Control Group		
	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.	Mean Difference
	Obs = 2053		Obs = 570		Obs = 1483		
Y	7.590	0.737	7.942	0.714	7.455	0.701	0.487 ***
Lin06	7.444	0.741	7.758	0.695	7.324	0.723	0.434 ***
Edu1	0.044	0.206	0.009	0.093	0.058	0.234	−0.049 ***
Edu2	0.297	0.457	0.168	0.375	0.347	0.476	−0.178 ***
Edu3	0.497	0.500	0.565	0.496	0.471	0.499	0.094 ***
Edu4	0.151	0.358	0.235	0.424	0.119	0.324	0.116 ***
Edu5	0.011	0.103	0.023	0.149	0.006	0.078	0.017 ***
T	0.276	0.448	--	--	--	--	1.000 ***
Lnassets	6.185	2.131	6.977	1.993	5.880	2.105	1.097 ***
Cadre	0.105	0.306	0.163	0.37	0.082	0.275	0.081 ***
Emlabor	0.122	0.445	0.179	0.503	0.100	0.419	0.079 ***
Burden	0.626	0.603	0.529	0.532	0.663	0.624	−0.134 ***
Min	0.755	0.430	0.521	0.500	0.845	0.362	−0.324 ***
Geo1	0.151	0.358	0.314	0.465	0.088	0.283	0.226 ***
Geo2	0.123	0.328	0.247	0.432	0.075	0.263	0.173 ***
Geo3	0.727	0.446	0.439	0.497	0.837	0.369	−0.399 ***
Dis1	0.009	0.093	0.004	0.059	0.011	0.103	−0.007
Dis2	0.039	0.195	0.040	0.197	0.039	0.194	0.001
Dis3	0.068	0.251	0.154	0.362	0.034	0.182	0.120 ***
Dis4	0.150	0.357	0.228	0.420	0.120	0.325	0.108 ***
Dis5	0.734	0.442	0.574	0.495	0.796	0.403	−0.222 ***
Vitr06	0.278	0.407	0.695	0.319	0.038	0.099	0.657 ***

Note: Significant codes: '***' 0.001; '**' 0.01; '*' 0.05; '.' 0.1; ' ' 1.

the control group among most covariates except Dis1 and Dis 2.

4.3. Empirical Results

Before comparing each model, we need to estimate propensity score. As is mentioned before, we choose *Logit* model to estimate the propensity score and all coefficients estimation presented in **Table 3**. All the variables in **Table 3** are the determinants of the probability of participating in training.

The match balance results of each model which checks whether the results of match have achieved balance on a set of covariance are reported in **Tables 4** and **5**. For each covariance, kinds of tests and statistics are calculated, *i.e*. t-test, univariate and multivariate Kolmogorov-Smirnov (KS) tests and a variety of empirical-QQ plots statistics [28]. But only t-test p-value is reported before and after matching in our study, since for dummy variables, the KS tests results are not provided by the R software and they are the equivalent to the results from t-tests.

The balance results make clear that, each model succeed in reducing covariance differences between the treated and control group at different degrees. Take the Lin06 variable as an example, the balance of it has been

Table 3. Results of *logit* model.

Variables	Coefficient	Std. Error
Intercept	−9.89876***	1.21610
Lin06	0.53888***	0.08778
Edu2	0.70096	0.49156
Edu3	1.21421*	0.48353
Edu4	1.65044***	0.49761
Edu5	2.59881***	0.68397
Lnassets	0.24419***	0.03487
Cadre	0.26776	0.18417
Emlabor	0.10044	0.13294
Burden	−0.00799	0.10382
Min1	−0.85054***	0.15041
Geo2	−0.19314	0.20795
Geo3	−1.26714***	0.84592
Dis2	2.79108**	0.86386
Dis3	4.51532***	0.84592
Dis4	3.17671***	0.82175
Dis5	3.54656***	0.82854
Vitr06	0.26547.	0.14300

Note: Significant codes: '***' 0.001; '**' 0.01; '*' 0.05; '.' 0.1; ' ' 1.

made better by matching. The mean difference is 0.4341 with a p-value of 0.0000 which is statistic significant before matching. After matching, this difference decreases largely in each model and changes to not significant except in Model-1 with the t-test p-value of 0.006. Edu2 variable is balanced after matching in each model, which is what we expected; unfortunately this not happened to every variable. Dis2 variable has been made worse by MD Matching. Before matching, the mean difference is 0.0012 with a p-value of 0.8979, but after matching this mean difference changes to -0.0368 with a p-value of 0.0000 at 0.1 percent significant level. We can check all the balance statistics for every variable followed by the same logic. Then we summarized the comparison result in **Table 6**.

From **Table 6**, we can see that, there are totally 20 variables in each model. After matching, there are 14, 3, 13 and 8 variables still imbalanced in each model respectively. In all cases, Model-2 of PS Matching performs with a better balance result than the other three models, since only 3variables still imbalanced after matching, 1variable at the significant level of 5 percent and another 2 variables at the significant level of 1 percent. The results of Model-3 indicates that, combined PS Matching with MD Matching make improvement compared with Model-1, but not so remarkable in this case study. Gen-Match make greater progress than the MD Matching, only 2 variables are still significant imbalanced after matching at the level of 0.1 percent versus 8 variables in Model-1.

After matching, the three mean treatment parameters of ATE, TT and TUT and their standard errors are estimated. The results are summarized in **Table 7**.

As is shown in **Table 7**, all ATE, TT and TUT estimations are positive and statistical significant in four models. We observe that all three parameters estimation from MD Matching, MD and PS Matching and Genetic Matching are over 20 percent except PS Matching. For TT, the average income difference by treatment status for these farm households who are actually treated is 29.56 percent by MD Matching method, 24.92 percent and 27.35 percent for Model-3 Matching and Model-4 Matching respectively, but only 18.99 percent by the PS Matching method. In each model, the common result of TUT < ATE < TT indicates that the rural labor force transfer training program is well-directed.

From **Table 6**, we know that the PS Matching performs with an excellent balance result than the other three models. Under this method, TT is 18.99 percent. Compared with this result, the other three matching methods may overestimate the results.

Despite large differences among estimation results by different matching method, from **Table 7**, we conclude that, firstly, the positive estimation results demonstrate

Table 4. Description of the matching balance statistics for Model-1 and Model-2.

Coefficients		Model-1: MD Matching		Model-2:PS Matching	
		Before Match	After Match	Before Match	After Match
Lin06	Mean Difference	0.4341***	0.0507**	0.4341***	−0.0307
	t-test p-value	<2.22e-16	0.0059	<2.22e-16	0.4167
Edu2	Mean Difference	−0.1782***	−0.0070	−0.1782***	0.0286
	t-test p-value	<2.22e-16	0.3173	<2.22e-16	0.1644
Edu3	Mean Difference	0.0942***	−0.0737***	0.0942***	−0.0351
	t-test p-value	0.0001	2.86e-06	0.0001	0.2171
Edu4	Mean Difference	0.1164***	0.0561***	0.1164***	−0.0015
	t-test p-value	4.69e-09	0.0003	4.69e-09	0.9507
Edu5	Mean Difference	0.0167*	0.0211***	0.0167*	0.0067
	t-test p-value	0.0111	0.0005	0.0111	0.3934
Lnassets	Mean Difference	1.0969***	0.1469***	1.0969***	−0.1413
	t-test p-value	<2.22e-16	0.0005	<2.22e-16	0.1352
Cadre	Mean Difference	0.0809***	0.0035	0.0809***	−0.0328
	t-test p-value	2.48e-06	0.1571	2.48e-06	0.1327
Emlabor	Mean Difference	0.0785***	0.0246**	0.0785***	−0.0446
	t-test p-value	0.0010	0.0017	0.0010	0.1568
Burden	Mean Difference	−0.1344***	0.0442***	−0.1344***	0.0777*
	t-test p-value	1.23e-06	0.0002	1.23e-06	0.0120
Min1	Mean Difference	−0.3239***	−0.0035	−0.3239***	−0.0235
	t-test p-value	<2.22e-16	0.3173	<2.22e-16	0.3212
Geo2	Mean Difference	0.1725***	0.0333*	0.1725***	0.0172
	t-test p-value	<2.22e-16	0.0239	<2.22e-16	0.2474
Geo3	Mean Difference	−0.3989***	−0.0228*	−0.3989***	0.0191
	t-test p-value	<2.22e-16	0.0277	<2.22e-16	0.3192
Dis2	Mean Difference	0.0012	−0.0368***	0.0012	−0.0088
	t-test p-value	0.8979	2.33e-05	0.8979	0.4744
Dis3	Mean Difference	0.1200***	0.1158***	0.1200***	0.0566**
	t-test p-value	1.29e-13	<2.22e-16	1.29e-13	0.0011
Dis4	Mean Difference	0.1080***	−0.1105***	0.1080***	−0.0583*
	t-test p-value	4.10e-08	3.03e-12	4.10e-08	0.0232
Dis5	Mean Difference	−0.2220***	0.0351**	−0.2220***	0.0118
	t-test p-value	<2.22e-16	0.0020	<2.22e-16	0.6540
Vitr06	Mean Difference	0.657***	0.0189*	0.657***	−0.0189
	t-test p-value	<2.22e-16	0.0113	<2.22e-16	0.4432

Note: (1) e-16 means 10^{-16}; (2) Significant codes: '***' 0.001; '**' 0.01; '*' 0.05; '.' 0.1; ' ' 1.

Table 5. Description of the matching balance statistics for Model-3 and Model-4.

Coefficients		Model-3: MD and PS Matching		Model-4:Genetic Matching	
		Before Match	After Match	Before Match	After Match
Lin06	Mean Difference	0.4341***	0.0209	0.4341***	0.0070
	t-test p-value	<2.22e-16	0.2710	<2.22e-16	0.4403
Edu2	Mean Difference	−0.1782***	0.0000	−0.1782***	−0.0053
	t-test p-value	<2.22e-16	1.0000	<2.22e-16	0.1796
Edu3	Mean Difference	0.0942***	−0.0579***	0.0942***	−0.0211*
	t-test p-value	0.0001	0.0003	0.0001	0.0231
Edu4	Mean Difference	0.1164***	0.0333*	0.1164***	0.0123
	t-test p-value	4.69e-09	0.0345	4.69e-09	0.2086
Edu5	Mean Difference	0.0167*	0.0211***	0.0167*	0.0123*
	t-test p-value	0.0111	0.0005	0.0111	0.0345
Lnassets	Mean Difference	1.0969***	0.1435***	1.0969***	0.0431.
	t-test p-value	<2.22e-16	0.0010	<2.22e-16	0.0958
Cadre	Mean Difference	0.0809***	0.0053.	0.0809***	0.0105
	t-test p-value	2.48e-06	0.0830	2.48e-06	0.5988
Emlabor	Mean Difference	0.0785***	0.0176*	0.0785***	0.0070*
	t-test p-value	0.0010	0.0251	0.0010	0.0452
Burden	Mean Difference	−0.1344***	0.0539***	-0.1344***	0.0202
	t-test p-value	1.23e-06	2.74e-05	1.23e-06	0.5098
Min1	Mean Difference	−0.3239***	−0.0018	−0.3239***	−0.0421**
	t-test p-value	<2.22e-16	0.5638	<2.22e-16	0.0016
Geo2	Mean Difference	0.1725***	0.0561***	0.1725***	0.0176**
	t-test p-value	<2.22e-16	0.0006	<2.22e-16	0.0074
Geo3	Mean Difference	−0.3989***	−0.0088	−0.3989***	−0.0053
	t-test p-value	<2.22e-16	0.3173	<2.22e-16	0.1796
Dis2	Mean Difference	0.0012	−0.0105.	0.0012	−0.0105
	t-test p-value	0.8979	0.0575	0.8979	0.2888
Dis3	Mean Difference	0.1200***	0.0579***	0.1200***	0.1035***
	t-test p-value	1.29e-13	5.62e-09	1.29e-13	3.18e-10
Dis4	Mean Difference	0.1080***	−0.0842***	0.1080***	−0.1105***
	t-test p-value	4.10e-08	5.89e-09	4.10e-08	2.60e-07
Dis5	Mean Difference	−0.2220***	0.0404***	−0.2220***	0.0211
	t-test p-value	<2.22e-16	0.0006	<2.22e-16	0.261
Vitr06	Mean Difference	0.657***	0.0219**	0.657***	−0.0114
	t-test p-value	<2.22e-16	0.0041	<2.22e-16	0.5921

Note: (1) e-16 means 10^{-16}; (2) Significant codes: '***' 0.001; '**' 0.01; '*' 0.05; '.' 0.1; ' ' 1.

Table 6. Balance checking result.

		Model-1	Model-2	Model-3	Model-4
Total number of variables		20	20	20	20
Variables which significant before		18	18	18	18
Variables which still significant after		14	3	13	8
Significant level	0.1(.)	0	0	2	1
	0.05(*)	3	1	2	3
	0.01(**)	3	2	1	2
	0.001(***)	8	0	8	2

Table 7. ATE, TT and TUT estimation results.

	ATE	TT	TUT
Model-1: MD Matching	0.2533***	0.2956***	0.2375***
	(0.0489)	(0.0515)	(0.0585)
Model-2: PS Matching	0.1833**	0.1899**	0.1809*
	(0.0543)	(0.0540)	(0.0662)
Model-3: MD and PS Matching	0.2237***	0.2482***	0.2146**
	(0.0499)	(0.0522)	(0.0604)
Model-4: Genetic Matching	0.2415***	0.2735***	0.2281**
	(0.0499)	(0.0486)	(0.0626)

Note: (1)Numbers in parentheses are Abadie-Imbens (AI) standard error; (2) Significant codes: '***' 0.001; '**' 0.01; '*' 0.05; '.' 0.1; ' ' 1.

that the rural labor force training program is effective. Secondly, highest TT value illustrates that the rural labor force training program is well-directed.

5. Concluding Remarks

In this study, we use micro data obtained from China's Rural Household Survey (RHS) of Honghe Prefecture and Dehong Prefecture in Yunnan province, to estimate the Rural Labor Force Training Program effects in the autonomous minority nationality areas in southwestern frontier region of China.

We set up four models with different matching method, and in each model, three treatment effects specified as ATE, TT and TUT are estimated and the results answered the primary question which was proposed at the beginning of section 3 that, what is the difference between participants' post-program income and the income that they would have received if they haven't participated in training. The average TT value shows that participants' income will lose approximately 25.18 percent if they were not participating in training. Under each estimation method, people who participated in training gain the most and gain more than those who are randomly selected if they participated. TUT estimations are less than ATE and TT in each model, and clarify that if non-

participants are treated, their income will increase. Using our data set, positive ATE, TT and TUT estimations and the largest TT values, demonstrates that the rural labor force training program is effective and well-directed in Honghe Prefecture and Dehong Prefecture.

Our empirical study also provides a good example of using series matching methods. We compared four different matching methods for estimation rural labor force training program policy treatment effects on our sample data and showed the balance checking results with paired t-test p-value. In this empirical study, the Propensity Score Matching reduced imbalance mostly while this conclusion may be inconsistent with other literature. According to Sekhon, Genetic matching with a genetic search algorithm can directly optimize covariate balance even without the propensity score [28,29]. Sekhon estimated TT by Propensity Score Matching and Genetic Matching without propensity score methods using Lalonde experimental data. In his research, Genetic Matching is performed with better balance than Propensity Score Matching. This discrepancy conclusion may be caused by the diverse propensity score model and different constitution of co-variances. Different models give rise to different balance results and it is hard to obtain a correct propensity score model. So we suggest that it is better using a series of matching methods to estimate

treatment effects and combining their advantages, rather than only one matching method.

6. Acknowledgements

REFERENCES

[1]   Household Survey Department of National Bureau of Statistics, "Poverty Monitoring Report of Rural China," China Statistics Press, Beijing, 2010.

[2]   R. J. Lalonde, "The Promise of Public Sector-Sponsored Training Programs," *Journal of Economic Perspectives*, Vol. 9, No. 2, 1995, pp. 149-168. http://dx.doi.org/10.1257/jep.9.2.149

[3]   J. J. Heckman and N.-L. Salvador, "Using Matching, Instrumental Variables, and Control Functions to Estimate Economic Choice Models," *The Review of Economics and Statistics*, Vol. 86, No. 1, 2004, pp. 30-57. http://dx.doi.org/10.1162/003465304323023660

[4]   J. J. Heckman and V. Edward, "Local Instrumental Variable and Latent Variable Models for Identifying and Bounding Treatment Effects," *Proceedings of the National Academy of Sciences*, Vol. 96, No. 8, 1999, pp. 4730-4734. http://dx.doi.org/10.1073/pnas.96.8.4730

[5]   J. J. Heckman and V. Edward, "The Relationship between Treatment Parameters within a Latent Variable Framework," *Economics Letter*, Vol. 66, No. 1, 2000, pp. 33-39. http://dx.doi.org/10.1016/S0165-1765(99)00181-0

[6]   J. J. Heckman and V. Edward, "Structural Equation, Treatment Effects, and Economic Policy Evaluation," *Econometrica*, Vol. 73, No. 3, 2005, pp. 669-738. http://dx.doi.org/10.1111/j.1468-0262.2005.00594.x

[7]   J. J. Heckman, "Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture," *The Journal of Political Economy*, Vol. 109, No. 4, 2001, pp. 673-748. http://dx.doi.org/10.1086/322086

[8]   J. J. Heckman, U. Sergio and V. Edward, "Understanding Instrumental Variables in Models with Essential Heterogeneity," *The Review of Economics and Statistics*, Vol. 88, No. 3, 2006, pp. 389-432. http://dx.doi.org/10.1162/rest.88.3.389

[9]   J. J. Heckman, U. Sergio and V. Edward, "Estimation of treatment Effects under Essential Heterogeneity," Working Paper, University of Chicago and American Bar Foundation, Chicago, 2006.

[10] X. Zhou and Y. Xie, "Propensity-Score-Cased Method versus MTE-Based Methods in Causal Inference," Population Studies Center Research Report, University of Michigan, Ann Arbor, 2011. http://www.psc.isr.umich.edu/pubs/pdf/rr11-747.pdf

[11] P. R. Rosenbaum and D. B. Rubin, "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, Vol. 70, No. 1, 1983, pp. 41-55. http://dx.doi.org/10.1093/biomet/70.1.41

[12] H. D. Rajeev and W. Sadek, "Propensity Score Matching for Nonexperimental Causal Studies," *The Review of Economics and Statistic*s, Vol. 81, No. 1, 2001, pp. 151-161.

[13] A. Abadie and G. W. Imbens, "Large Sample Properties of Matching Estimators for Average Treatment Effects," Econometrica, Vol. 74, No. 1, 2006, pp. 235-267. http://dx.doi.org/10.1111/j.1468-0262.2006.00655.x

[14] J. S. Sekhon and R. M. Walter Jr., "Genetic Optimization Using Derivatives: Theory and Application to Nonlinear Models," *Political Analysis*, Vol. 7, No. 1, 1998, pp. 187-210. http://sekhon.berkeley.edu/genoud/genoud.pdf http://dx.doi.org/10.1093/pan/7.1.187

[15] J. S. Sekhon, "Alternative Balance Metrics for Bias Reduction in Matching Methods for Causal Inference," Working Paper, 2006. http://sekhon.berkeley.edu/papers/SekhonBalanceMetrics.pdf

[16] J. S. Sekhon, "Opiates for the Matches: Matching Methods for Causal Inference," *Annual Review of Political Science*, Vol. 12, 2009, pp. 487-508. http://dx.doi.org/10.1146/annurev.polisci.11.060606.135444

[17] A. Diamond and S. S. Jasjeet, "Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational in Studies," Working Paper, 2005. http://sekhon.berkeley.edu/papers/GenMatch.pdf

[18] W. G. Cochran and D. B. Rubin, "Controlling Bias in Observational Studies: A Review," *Sankhya*, *Series A*, Vol. 35, No. 4, 1973, pp. 417-446.

[19] D. B. Rubin, "Using Multivariate Sampling and Regression Adjustment to Control Bias in Observational Studies," *Journal of the American Statistical Associati*on, Vol. 74, No. 366, 1979, pp. 318-328.

[20] D. B. Rubin, "Bias Reduction Using Mahalanobis-Metric Matching," *Biometrics*, Vol. 36, No. 2, 1980, pp. 293-298. http://dx.doi.org/10.2307/2529981

[21] A. Abadie andG. W. Imbens, "Simple and Bias-Corrected Matching Estimators for Average Treatment Effects," Technical Working Paper, 2002. http://www.nber.org/papers/t0283.pdf

[22] J. S. Sekhon and G. Richard, "A Nonparametric Matching Method for Covariate Adjustment with Application to Economic Evaluation," Health Economics, 2011. http://sekhon.berkeley.edu/papers/GeneticMatching_SekhonGrieve.pdf

[23] M. Lechner, "Earnings and Employment Effects of Continuous Off-the-Job Training in East Germany after Unification," *Journal of Business and Economics Statistic*s, Vol. 17, No. 1, 1999, pp. 74-90.

[24] G. W. Imbens, "Nonparametric Estimation of Average treatment Effects under Exogeneity: A Review," *Review of Economics and Statistics*, Vol. 86, No. 1, 2004, pp. 4-29. http://dx.doi.org/10.1162/003465304323023651

[25] W. G. Cochran and D. B. Rubin, "Controlling Bias in

Observational Studies: A Review," *Sankhya*, *Series A*, Vol. 35, No. 4, 1973, pp. 417-446.

[26] P. R. Rosenbaum and D. B. Rubin, "Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity," *American Statistician*, Vol. 39, No. 1, 1985, pp. 33-38.

[27] P. R. Rosenbaum and D. B. Rubin, "The Bias Due to Incomplete Matching," *Biometrika*, Vol. 41, No. 1, 1986, pp. 103-116.

[28] J. S. Sekhon, "Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching Package for R," *Journal of Statistical Software*, Vol. 42, No. 7, 2011, pp. 1-52.

[29] J. S. Sekhon, "Matching: Multivariate and Propensity Score Matching with Balance Optimization," 2004. http://sekhon.berkeley.edu/matching/

[30] J. Mincer, "Schooling, Experience and Earnings," National Bureau of Economic Research, New York, 1974.