Using Propensity Score Matching to Construct Comparable Samples in Evaluation Studies: A Case Study of a Road Transport Project[1]

Introduction

The Asian Development Bank has supported the Greater Mekong Sub-region (GMS) since its establishment in 1992 by Cambodia, China, Laos, Myanmar, Thailand and Vietnam in order to strengthen economic linkages among them. The countries adopted the GMS Strategic Framework 2002–12 in 2002. The framework outlines a comprehensive development agenda, underscoring the need to supplement infrastructure linkage with cooperation on policies and programmes to promote equitable and environmentally sustainable growth.

Improvement and rehabilitation of the transport system among GMS countries has been identified as one of nine priority cooperation areas.[2] The GMS Southern Coastal Corridor (GMS-SCC) project assists in the rehabilitation of National Road 33 in Cambodia and another section in southern Vietnam. This project is expected to have substantial impact on economic development, regional cooperation and better access to employment opportunities, markets and social services.

The GMS-SCC project has made significant progress. However, the mid-term review of the strategic framework reveals that the precise scale and scope of the project's contribution to poverty reduction and other socio-economic indicators may not be determined without comprehensive project baseline information to aid monitoring and evaluation. A naive estimate of intervention impacts would be to compare the relevant outcomes between participants (treatment group) and non-participants (comparison group). However, a difference in the outcomes between the two groups would not be reliable if any biases had not been eliminated (Ravallion 2001). Propensity score matching (PSM) is a statistical technique that allows investigators to account for biases in two groups and subsequently eliminate them.

This study aims at assessing the socio-economic effects of the GMS-SCC project in Cambodia, with the specific objectives: (1) design and construct a complete benchmark data set and description with relevant indicators; (2) assess and provide comprehensive baseline and evaluation data in the project areas (treatment group) and a few selected non-project areas (comparison group); and (3) construct comparable treatment and control groups using PSM.

Research Methodology

A household survey was used to construct baseline indicators. The household samples were classified into geographically similar treatment and control areas. The treatment area was selected within 5 km of both sides of the road, and the control area was selected from beyond 5 km of both sides of the road. Four hundred and fifty households were randomly selected for the treatment area and 360 for the control area. The household survey was used for a descriptive analysis of the current socio-economic and transport indicators in both treatment and control groups. Since this baseline data will be used for project monitoring and evaluation, PSM was used to determine similar characteristics of the two samples, reducing biases caused by sample selection error. Households that did not match a given set of characteristics were dropped from the samples, balancing the observed variables between the treatment and control groups.[3]

1 This article is prepared by Dr Theng Vuthy, programme coordinator, Mr Kem Sothorn, research associate and Mr Chhim Chhun and Ms Khiev Pirom, research assistants, at CDRI. It is a summary of the empirical findings of a CDRI project, Assessing the Socio-economic Effects of GMS-Southern Coastal Corridor, in 2009–10, funded by the Asian Development Bank.

2 The nine priority sectors are: transport, telecommunications, energy, environment, tourism, trade facilitation, investment, human resource development and agriculture.

3 Further details on the research method can be found in Theng and Kem (forthcoming) and Hing et al (forthcoming).

Empirical Key Findings

Descriptive Statistics

Table 1 summarises the differences between treatment and control groups, with their t-values. The t-values suggest that there are some significant differences between the groups in demographic and welfare characteristics. There appear to be significant differences in gender of household head, age, marital status, education level of household head and number of years settled in the area. There are also significant differences in income of household heads, small businesses of household head, wealth index and durable goods and house values. In addition, per capita income, per capita consumption and other agricultural expenditure and gross margin show significant differences between treatment and comparison groups (Table 1). Only a few demographic and livelihood indicators show no statistically significant differences. These results suggest that there are some biases in sampling selection, even though sampling was random.

Matching Households

To reduce biases and to construct similar characteristics of treatment and control groups for the future impact evaluation study, PSM was used (Ravallion 2001; Caliendo & Kopeinig 2008; World Bank 2002). Table 2 presents the estimates from the logit model used to match households in the GMS-SCC project area with the control area. The

Table 1: Description and Differences of Variables (sample mean)

Variable	Treatment	Control	Difference	T-value
Dependent variable (Asian Development Bank project)	1	0		
Control variables				
Household size	5.34	5.18	0.16	1.2
Number of women aged 15-45 years	1.36	1.33	0.03	0.4
Household labour power	4.15	3.96	0.19	1.6
Number of years in the village	4.53	3.39	1.14*	1.8
Gender of household head (dummy 1=male, 0=female)	0.76	0.84	-0.08***	-2.7
Age of household head	50.80	46.69	4.12***	4.3
Age of household head_ squared	2763.1	2358.2	404.9***	4.1
Marital status of household head (1=married, 0=otherwise)	0.76	0.82	-0.06**	-2.0
Educational level of household head (number of years)	3.64	4.20	-0.56**	-2.5
Household head's income from non-farming activities	100.96	67.50	33.46***	3.9
Household head's income from non-farming_ squared	27460.4	15691.6	11768.8**	2.2
Unemployment (dummy, 1: unemployed; 0: employed)	0.29	0.35	-0.05	-1.6
Civil servant as household head (dummy)	0.063	0.056	0.008	0.5
Farm worker as household head (dummy)	0.079	0.095	-0.016	-0.8
Small business of household head (dummy)	0.33	0.26	0.07**	2.3
Number of non-farming activities of household head	0.85	0.81	0.04	0.9
Index of all household assets	0.85	-1.04	1.89***	14.8
Total value of household assets	377.5	440.4	-62.9**	-2.0
Total value of household assets_squared	359788.1	343831.0	15957.1	0.2
Value of house	1014.4	424.7	589.7***	4.4
Value of house_squared	7409425.0	395630.6	7013794.0	1.6
Outcome variables (expected impact of road projects)				
Per capita daily income (riels)	3801.8	3218.9	582.9***	3.1
Per capita daily food consumption (riels)	2518.4	2372.0	146.4	1.6
Per capita daily non-food consumption (riels)	1770.2	1503.2	267.0**	2.0
Per capita daily total consumption (riels)	4288.6	3875.2	413.3**	2.2
Rice production cost per ha ('0000 riels)	180.4	148.7	31.8***	2.7
Rice gross margin per ha ('0000 riels)	52.9	71.1	-18.2*	-1.7
Non-rice gross margin per ha ('0000 riels)	90.84	153.5	-62.7	-1.4

Level of significance * at 10 percent; ** at 5 percent; *** at 1 percent.
Source: Household Survey, December 2009

Table 2: Propensity Score for Household Matching (logit estimation)

Control Variables	Coef.	Std. Err.	z
Household size	-0.0125	0.1243	-0.10
Number of women aged 15-45 years	0.0189	0.1294	0.15
Household labour power	-0.0496	0.1725	-0.29
Number of year in the village	0.0107	0.0129	0.82
Gender of household head (dummy 1=male, 0=female)	-0.3322	0.3406	-0.98
Age of household head	0.1027	0.0547	1.88
Age of household head_squared	-0.0007	0.0005	-1.35
Marital status of household head (1=married, 0=otherwise)	0.5167	0.3415	1.51
Education level of household head (number of year)	-0.0857	0.0332	-2.58**
Household head's income from non-farming activities	0.0024	0.0021	1.11
Household head's income from non-farming activities_ squared	0.0000	0.0000	-1.42
Unemployment (dummy, 1: unemployed; 0: employed)	0.0122	0.3557	0.03
Civil servant (dummy)	-0.5475	0.4511	-1.21
Farm worker (dummy)	-0.1324	0.3437	-0.39
Small business (dummy)	-0.0465	0.2379	-0.20
Number of non-farming activities of household head	0.2633	0.2506	1.05
Index of all household assets	0.8939	0.0880	10.15***
Total value of household assets	-0.0024	0.0006	-3.96***
Total value of household assets_squared	0.0000	0.0000	2.04**
Value of house	0.0003	0.0003	0.85
Value of house_squared	0.0000	0.0000	0.35
Constant	-2.0159	1.3431	-1.50
Number of observations	801	-	-
Pseudo-R^2	0.29	-	-

Level of significance * at 10 percent; ** at 5 percent; *** at 1 percent.

results were that only a few variables (household head education level and assets) are statistically significant, suggesting a similarity between matched households in treatment and control areas.

After the matching process, common on-support and off-support[4] households were identified; 98 households in the treatment area fell into off-support, and 344 individual households in the treatment area matched well with 359 households in the control area. The characteristics of households that could not find a suitable match, most of which are located in villages near the border and district town areas, may have more access to electricity, tap water or energy sources for cooking, which households in the control areas could not afford.

The results from the household matching in Table 3 generally suggest that the differences between the expected outcome indicators were significantly reduced, and more balance was obtained. Differences in per capita income, per capita non-food and total consumption, including both production expenditure and gross income from rice, were statistically non-significant after matching; per capita food consumption and non-rice gross margin differences were largely balanced after matching.

The matching has to be checked since we did not condition all variables, but rather used propensity scores, to see whether the matching procedure is able to balance the distribution of the relevant variables in control and treatment groups. Standardised bias, pseudo-R^2 and t-statistics before and after matching are suggested as important indications for checking matching quality (Caliendo & Kopeinig 2008). The median and mean absolute standardised biases

4 Common support indicates the individuals in the treatment area who find a suitable match, whereas off-support indicates the individuals in the treatment area who did not find a suitable match with control area.

Table 3: Comparison of Outcome Variables after Matching
(riels at December 2009 prices)

Outcome variables	Difference (Unmatched)		Difference (Matched)	
	Unmatched	T-stat	Matched	T-stat
Per capita income	582.87	3.08***	-86.53	-0.40
Per capita food consumption	146.37	1.65	64.52	0.63
Per capita non-food consumption	266.98	1.95**	43.22	0.25
Per capita total consumption	413.34	2.2**	107.74	0.48
Rice production cost per ha ('0000)	31.77	2.74***	15.94	1.22
Rice gross margin per ha ('0000)	-18.24	-1.75*	-7.10	-0.51
Non-rice gross margin per ha ('0000)	-62.67	-1.35	-39.01	-0.73

Level of significance * at 10 percent; ** at 5 percent; *** at 1 percent.

show substantial reduction, at 2.7 and 3.8 percent, respectively, after matching. Caliendo and Kopeinig (2008) suggest that median and mean of standardised bias below 3 and 5 percent after matching are sufficiently creditable. The pseudo-R^2 from the estimation of propensity score after matching was very low (0.03), compared to that before matching (0.29), and the P-value of the livelihood ratio test is non-significant after matching, indicating there are no systematic differences in the distribution of relevant variables between groups. These results suggest that a significant reduction in selection bias is obtained through household matching and that two similar study samples have been constructed for monitoring and evaluation. Therefore, the 98 households in the treatment area that fell into off-support were dropped, giving a total sample of 703 observations after matching.

Conclusions

This report presents baseline indicators of a potential longitudinal study area of the GMS-SCC project in Cambodia. It also provides indications of similarities and differences in the treatment and control areas before matching and significantly reduces the differences after matching. Before matching there were some statistically significant differences in household demographic and welfare indicators, indicating some biases in household sampling selection.

Matched households have similar characteristics in variables from demographics to socio-economic indicators and transport means indicators. Two sample groups with similar characteristics have been constructed by PSM, comprising 344 and 359 households in the treatment and control groups, respectively. Ninety-eight households have been dropped from the treatment sample because no matching partners could be found in the comparison group. This should save costs and time and estimate the project impact for the monitoring and evaluation.

References

Caliendo, M. & S. Kopeinig (2008), "Some Practical Guidance for the Implementation of Propensity Score Matching", *Journal of Economic Surveys,* Vol. 22, No. 1, pp. 31–72

Ravallion, M. (2001), "The Mystery of the Vanishing Benefits: An Introduction to Impact Evaluation", *World Bank Economic Review,* Vol. 15, pp. 115–140

Theng Vuthy & Kem Sothorn, (forthcoming), The Importance of PSM in Evaluation Studies: A Case Study of Physical Infrastructure in Cambodia, Working Paper (Phnom Penh: CDRI)

World Bank (2002), Socioeconomic Impact Assessment of Rural Roads: Methodology and Questionnaires

Hing Vutha, Saing Chan Hang & Khieng Sothy, (forthcoming). Using Propensity Score Matching to Generate Two Comparable Groups in the Baseline Survey: The Case of GMS Transmission Line Project. Annual Development Review 2010/2011(Phnom Penh: CDRI).