Propensity Score Matching (PSM)

Third method: To analyze the potential effects of growing rubber or oil palm on household food security, several evaluation challenges need to be addressed (Kissoly et al. 2017) since that farmers may have different probability of being selected as participants (i.e. self-selection). This is addressed by estimating the average treatment effect of participating in rubber or oil palm farming on food security. The PSM approach generates a control group and then addresses the bias due to selection-on-observables, overt bias (Awotide et al. 2015). The PSM is used in observational studies to adjust for differences in pre-treatment variables, and to draw inferences about the effects of binary treatments or participants (Rosenbaum and Robin 1983). The basic idea is to match each participant with an otherwise identical nonparticipant (the comparator) –based on observed pretreatment characteristics –and then to measure the measure the average difference in the outcome variable between the participants and the comparison group (Haughton and Haughton, 2011).

The PSM was implemented: **(a)** probit models were first used to estimate the propensity scores for the two different treatments for analyzing the determinants of the participatory. For a particular treatment, the probit specified as follows:

$$Y = \beta_0 + \beta_1 X + \epsilon \qquad (1)$$

Where Y is a binary variable representing a household's participatory in growing rubber or oil palm, (two different treatments and a pooled of these two were also tested) and X is a set of covariates relevant in the choice of participation in growing industrial crops. β stands for a vector of coefficients to be estimated while ε captures a vector of random unobserved factors affecting the choice of the participatory in growing industrial crops. The PSM generates a variable known as the propensity score, which is the probability or tendency that a grower would adopt any industrial crop (rubber or oil palm) conditional on the grower's observable characteristics. The propensity score (P(x)) is written as:

$$P(x) = Pr(T = 1/X = x) \qquad (2)$$

The generated propensity score can usually be used to create matched samples, uniform subgroups, and weight for balancing characteristics between farmers and a variable for controlling or adjusting the data (Guo and Fraser 2009). The proficiency of the PSM to control for the differences in observable covariates that might influence the decision of the participatory in growing industrial crops is based on the conditional independence assumption. This states that conditional on observables characteristic of industrial crops' growers (X), food security outcomes are independent in the participatory in rubber or oil palm specified as: (TY1, Y0⊥T|X). Another vital assumption is the common support or overlap condition: $0 < P(T = 1| X) < 1$. According to Heckman et al. (1999), this condition ensures that the treatment observations (participants) have comparison observations (control groups) "nearby" in the propensity score distribution. Only in areas of common support can inferences be made about causality. It is also very important to conduct a balancing test, that is, to ascertain if: $P^*(X/T = 1) = P^*(X/T = 0)$. However, it is worthy of note that the estimation of the propensity score is a necessary but not sufficient condition to calculate the parameters of interest such as the Average Treatment Effect (ATE), Average Treatment Effect on the Treated (ATT), and Average Treatment Effect on the Untreated. There is a need to search for the appropriate counterfactuals that match each participant depending on its propensity score. **(b)** The effect of participation on food security outcomes was estimated using the three most commonly adopted matching methods (nearest neighborhood matching (NNM), kernel based matching (KBM) and caliper matching (CBM)) in the literature (Awotide et al. 2015, Caliendo and Kopeinig, 2008). All matching estimators contrast the food security outcome of a treated individual with outcomes of comparison group members. The Average Treatment Effect on the Treated (ATT), which is the most important parameter to us in this study, is then estimated by averaging within-match differences in the food security outcome variables between participants and the controls (Awotide et al. 2015).

Endogenous Treatment Effect Regression (ETER)

The endogenous treatment effect regression accounts for the other bias related to unobservable characteristics of the growers, which cannot be controlled using the PSM technique alone (Awotide et al. 2015). In addition, the endogenous treatment effect model supports to maintain the assumption related to the stable unit treatment value assumption (SUTVA) (Rubin, 1978), that there is neither interference between units nor different versions of the treatment. Equations (3) and (4) are the two potential outcomes equations in the two possible states (participants and controls) of the industrial crops adoption.

$$Y1 = \gamma_1 X_1 + \vartheta_1 \qquad (3)$$

$$Y0 = \gamma_0 X_0 + \vartheta_0 \qquad (4)$$

$$T^* = \varphi_T Z_T + \vartheta_T \qquad (5)$$

Where T^* is latent variable generating T(Z) as shown in equation (6). If T(Z) stands for the observed adoption decision, where T(Z) = 1 if the producer is a participant in growing rubber or oil palm and T(Z) = 0 if the farmer belongs to the control group.

$$T(Z) = 1[T^*(Z) \geq 0] = 1\left[X_\varphi + \vartheta_D \geq 0\right] \qquad (6)$$

The counterfactual choice variables are also defined. For any z which is a potential realization of Z, we define the variable $T(z) = 1[z\varphi \geq \vartheta_T]$, which shows whether or not the individual industrial crop grower would have adopted any rubber or oil palm farming had the value of Z been externally set to z, holding constant the observed ϑ_T. This requires an exclusive restriction and denoted by Z_n some element of Z, which is not in X, it is possible for us to manipulate an individual industrial crop grower's probability of participating in rubber or oil palm farming without tampering with the potential dietary diversity, food insecurity related experiential measures and consumption behaviors. Finally, we assume that (ϑT, ϑ1, ϑ0) is independent of X and Z. If the observed outcome is:

$$Y = TY1 + (1 - T)Y0 \qquad (7)$$

The Average Treatment Effect (ATE), defines the gain or impact of participation in growing rubber or oil palm on considered food security indicators and this can be expressed as follows:

$$Y = Y1 - Y0 \qquad (8)$$

Therefore the ATE conditional on X = x can be expressed as shown in equation (9). The Average Treatment Effect on the Treated (ATT), which is the gain or impact in considered food security for those growers that actually participate in growing rubber or oil palm, can be expressed as written in equation (10).

$$ATE(x) = E(\Delta | X = x) = x_1(\gamma_1 - \gamma_0) \qquad (9)$$

$$ATT = (x, z, T(z) = 1) = E(\Delta | X = x, Z = z, T(z) = 1)$$

$$= x_1(\gamma_1 - \gamma_0) + E\left(\vartheta_1 - \frac{\vartheta_0}{\vartheta_T} \geq -z_1\varphi\right) \qquad (10)$$

We use *"etregress"* to estimate the parameters of the endogenous treatment regression model. *etregress* estimates an average treatment effect (ATE) and the other parameters of a linear regression model augmented with an endogenous binary-treatment variable. Estimation is by either full maximum likelihood or a two-step consistent estimator. When there are no interactions between the treatment variable and the outcome covariates, *etregress* directly estimates the ATE and the ATT. The ATT is the same as the ATE in this case because the treatment indicator variable has not interacted with any of the outcome covariates (STATAcorp 2013).

Awotide, B. A, Alene, A. D., Abdoulaye, T., & Manyong, V. M. (2015). Impact of agricultural technology adoption on asset ownership: the case of improved cassava varieties in Nigeria. Food Security 7(6): 1239-1258.

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. Journal of Economic Surveys, 22(1): 31-72.

Guo, S., & Fraser, W. M. (2009). Propensity score analysis: Statistical methods and applications. Thousand Oaks: Sage Publications.

Heckman, J., Lalonde, R., & Smith, J. (1999). The economics and econometrics of active labour market programs. In O. Ashenfelter & D. Card (Eds.) Handbook of Labor Economics, Vol. III, Amsterdam: Elsevier.

Kissoly, L., Fabe, A. & Grote, U. (2017). The integration of smallholders in agricultural value chain activities and food security: evidence from rural Tanzania. Food sec. Doi 10.1007/s12571-016-0642-2

Haughton, D. and Haughton, J. (2011). Living standards analytics: Development through the lens of household survey data. Springer Science & Business Media. Springer New York Dordrecht Heidelberg London. 245-267.

Rosenbaum, P. R., & Rubin D. B. (1983). The central role of the propensity score in observational studies for causal effects. Biometrika 70:41-55.

Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. The Annals of statistics, 6, 34–58.

STATAcorp (2013). Endogenous treatment effects. http://www.stata.com/stata13/endogenous-treatment-effects/. Accessed 9 Aug 2018.

Table S1. Mean score and distribution across food security thresholds for each study group.

Indicators	Thresholds		Rubber Growers	Oil palm Growers	Subsistence Farmers
FCS	Food Secure	Acceptable	55.3 (46.6 %)	62.7 (29.7 %)	57.9 (59.7 %)
	Food Insecure	Borderline	35.2 (15.1 %)	38.0 (24.4 %)	34.9 (23.9 %)
		Poor	14.1 (38.3 %)	14.2 (45.9 %)	13.9 (16.4 %)
HFIAS	Food Secure	Secure	1.0 (1.4 %)	-	1.0 (4.5 %)
		Mildly food insecure	1.5 (2.9 %)	-	2.7 (4.5 %)
	Food Insecure	Moderately food insecure	4.5 (5.7 %)	6.3 (8.6 %)	6.6 (14.9 %)
		Severely food insecure	12.3 (90.0 %)	8.4 (91.4 %)	13.4 (76.1 %)
HHS	Food Secure	Little to no hunger	0.4 (31.6 %)	0.6 (32.4 %)	0.3 (34.3 %)
	Food Insecure	Moderate hunger	2.7 (50.0 %)	2.3 (45.9 %)	2.8 (49.3 %)
		Severe hunger	4.5 (18.4 %)	4.4 (21.6 %)	4.5 (16.4 %)
CSI	Food Secure	Secure	0.4 (18.4 %)	-	0.3 (17.9 %)
		Mildly food insecure	6.9 (25.0 %)	6.9 (37.8 %)	5.8 (20.9 %)
	Food Insecure	Moderately food insecure	23.9 (23.7 %)	29.6 (18.9 %)	26.1 (23.9 %)
		Severely food insecure	70.1 (32.9 %)	77.7 (43.2 %)	67.7 (37.3 %)
rCSI	Food Secure	Secure	0.5 (40.8 %)	0.2 (45.9 %)	0.4 (41.8 %)
		Mildly food insecure	6.6 (25.0 %)	5.4 (13.6 %)	6.4 (11.9 %)
	Food Insecure	Moderately food insecure	14.7 (17.1 %)	14.3 (18.9 %)	12.8 (28.4 %)
		Severely food insecure	25.8 (17.1 %)	24.5 (21.6 %)	32.2 (17.9 %)

Note: This table includes only food security indicators that can categorize respondents according to a classification system of food security (Maxwell et al 2014).

Table S2. Mean food security scores for HHS and rCSI and differences between study groups.

	Rubber growers (Group 1)	Oil palm growers (Group 2)	Subsistence farmers (Group 3)	Group 1 vs. Group 2	Group 2 vs. Group 3	Group 1 vs. Group 2
	Mean score (Standard deviation)			(Mean difference) p-value		
HHS	2.30 (1.55)	2.19 (1.49)	2.19 (1.63)	(0.11) 0.683	(-0.005) 0.988	(0.11) 0.712
rCSI	8.78 (9.83)	8.81 (10.47)	10.34 (12.33)	(-1.57) 0.400	(-1.53) 0.524	(-0.03) 0.986

Table S3. Determinants of participation in industrial crop production.

	Rubber Growers (1) and Subsistence Farmers (0)	Oil palm growers (1) and Subsistence Farmers (0)
	Coefficient (SE)	Coefficient (SE)
Origin of respondent	0.26 (0.42)	0.28 (0.75)
Household size	0.07** (0.03)	0.06 (0.04)
Dependency ratio	-0.01 (0.010)	-0.008 (0.013)
Off-farm income	0.002 (0.002)	0.002 (0.002)
Gender of household head	0.34 (0.32)	0.22 (0.55)
Cultivated land	0.01 (0.02)	-0.01 (0.06)
Livestock	0.006 (0.09)	-0.59 (0.47)

Income	0.0002 (0.00014)	0.0004** (0.0001)
Farming experience	0.56*** (0.16)	0.67*** (0.17)
Constant	-1.60 (0.47)	-2.36 (0.80)
Number of observations	143	104
LR chi2(9)	94.16***	98.82***
Log likelihood	-51.75	-18.29
Pseudo R2	0.48	0.73

Note: *** = p<0.01; ** = p<0.05.

Table S4. Impacts of participation in industrial crop production on HFIAS and HHS.

	Rubber growers vs. Subsistence farmers				**Oil palm growers vs. Subsistence farmers**			
Parameters	**Treated**	**Untreated**	**Difference (S.E)**	**T-stat**	**Treated**	**Untreated**	**Difference (S.E)**	**T-stat**
			HHS					
			NNM					
Unmatched	2.30	2.19	0.11 (0.27)	0.41	2.19	2.19	0.00 (0.32)	-0.01
ATT	2.30	4.04	-1.74 (1.24)	-1.4*	2.19	4.38	-2.19 (1.15)	-1.91**
			CBM					
Unmatched	2.30	2.19	0.11 (0.27)	0.41	2.19	2.19	0.00 (0.32)	-0.01
ATT	2.30	2.05	0.25 (0.65)	0.39	1.40	2.80	-1.40 (0.94)	-1.5*
			KBM					
Unmatched	2.30	2.19	0.11 (0.27)	0.41	2.19	2.19	0.00 (0.32)	-0.01
ATT	2.38	2.63	-0.26 (0.45)	-0.56	2.26	4.62	-2.36 (1.41)	-1.68**
			rCSI					
			NNM					
Unmatched	8.78	10.34	-1.57 (1.86)	-0.84	8.81	10.34	-1.53 (2.40)	-0.64
ATT	8.78	16.14	-7.37 (5.37)	-1.37*	8.81	16.32	-7.51 (12.85)	-0.58
			CBM					
Unmatched	8.78	10.34	-1.57 (1.86)	-0.84	8.81	10.34	-1.53 (2.40)	-0.64
ATT	7.05	8.80	-1.75 (3.12)	-0.56	8.40	11.60	-3.20 (10.81)	-0.3
			KBM					
Unmatched	8.78	10.34	-1.57 (1.86)	-0.84	8.81	10.34	-1.53 (2.40)	-0.64
ATT	6.84	9.68	-2.83 (3.13)	-0.9	9.29	18.30	-9.01 (10.71)	-0.84

Note: **=p<0.05; *=p<0.1; NNM = Nearest neighbor matching; CBM = Caliper-based matching; KBM = Kernel-based matching.

Table S5. Test of balancing for confounders.

Variable		Rubber growers and Subsistence farmers				Oil palm growers and Subsistence farmers			
		Mean		%bias	p>t	Mean		%bias	p>t
		Treated	Untreated			Treated	Untreated		
Native	U	0.89	0.84	17.2	0.304	0.89	0.84	16.2	0.441
	M	0.88	0.67	58.4	0.057	0.87	0.06	235.5	0.000
Household size	U	8.20	6.34	35.7	0.037	10.84	6.34	67.4	0.000
	M	8.97	7.58	26.7	0.354	11.23	2.47	131.4	0.000
Dependency ratio	U	7.80	16.68	-21.0	0.202	12.01	16.68	-10.9	0.629
	M	5.92	17.69	-27.8	0.104	14.34	84.57	-163.8	0.000
Non-farm income	U	64.97	28.06	29.5	0.088	75.11	28.06	31.5	0.080
	M	52.97	52.03	0.8	0.964	76.03	1.14	50.2	0.059
Gender	U	0.28	0.30	-4.9	0.772	0.27	0.30	-6.2	0.764
	M	0.34	0.35	-1.5	0.956	0.16	0.02	30.7	0.058
Cultivated land	U	7.17	5.07	21.3	0.213	7.51	5.07	32.9	0.105
	M	6.07	5.01	10.7	0.544	8.35	4.70	49.3	0.019
Livestock	U	1.16	0.65	12.0	0.488	0.28	0.65	-29.1	0.189
	M	0.73	0.57	3.8	0.684	0.33	0.07	20.5	0.159
Income	U	754.30	421.58	29.8	0.077	1406.10	421.58	75.7	0.000
	M	725.63	720.30	0.5	0.988	1340.60	774.88	43.5	0.255
Farming experience	U	7.82	0.10	129.1	0.000	7.81	0.10	167.6	0.000
	M	0.59	1.17	-9.7	0.258	9.03	5.87	68.8	0.013

Notes: U = Unmatched; M = Matched.

Table S6. Linear Regression with Endogenous Treatment Effects for participation in rubber production (HHS and rCSI metrics)

Outcome equation	**HHS**	**rCSI**
Coefficient (Standard Error)		
Household size	0.02 (0.03)	-0.36** (0.17)
Dependency ratio	0.003 (0.003)	-0.02 (0.02)
Ratio of land given to SOGUIPAH to current land	4E-05 (6E-05)	-0.001* (0.0004)
Home garden area	0.04* (0.02)	-0.02 (0.16)
Natural palm grove area	-0.06 (0.05)	0.67* (0.35)
Livestock	0.02 (0.03)	0.47** (0.20)
Monthly food expenditure	-0.002 (0.002)	0.007 (0.01)
Annual food expenditure	0.0002 (0.0001)	0.0005 (0.001)
Amount borrowed for food	0.001 (0.001)	0.0013 (0.008)
Origin of respondent	-0.87** (0.42)	4.04 (2.86)
Gender of household head	0.05 (0.30)	-1.08 (2.11)
Participation in industrial crop production (1= Rubber, 0=Subsistence)	-0.58 (0.87)	-10.20** (4.80)
Constant	3.01*** (0.53)	12.93*** (3.35)
Treatment equation		
Off-farm income	0.002* (0.001)	0.002 (0.001)
MIHFP	0.07 (0.05)	0.07 (0.05)

Income	0.0002** (1E-04)	0.0003*** (1E-04)
Origin of respondent	0.50 (0.32)	0.52 (0.32)
Gender of household head	0.004 (0.24)	-0.0006 (0.24)
Constant	-0.75** (0.35)	-0.78** (0.34)
/athrho	0.42 (0.36)	0.44 (0.29)
/lnsigma	0.45*** (0.10)	2.38*** (0.09)
rho	0.40 (0.31)	0.41 (0.24)
sigma	1.56 (0.16)	10.85 (0.97)
lambda	0.62 (0.53)	4.46 (2.89)
Number of observations	143	143
Wald chi2(12)	16.17	22.28**
Log likelihood	-352.97	-629.10

Notes: *** p <0.01; ** p < 0.05; * p < 0.1.

Table S7. Linear Regression with Endogenous Treatment Effects for participation in oil palm production (HHS and rCSI metrics).

Outcome equation	**HHS**	**rCSI**
Coefficient (Standard Error)		
Household size	-0.02 (0.03)	-0.50*** (0.19)
Dependency ratio	0.004 (0.003)	-0.005 (0.02)
Ratio of land given to SOGUIPAH to current land	-9E-06 (6E-05)	-0.0005 (0.0004)
Home garden area	0.04 (0.03)	0.37** (0.18)
Natural palm grove area	0.03 (0.09)	0.83 (0.60)
Livestock	0.03 (0.11)	0.94 (0.81)
Monthly food expenditure	-0.0002 (0.002)	0.009 (0.015)
Annual food expenditure	-8E-05 (0.0002)	-0.0004 (0.001)
Amount borrowed for food	0.001 (0.001)	0.003 (0.009)
Origin of respondent	-0.35 (0.46)	2.37 (3.29)
Gender of household head	-0.05 (0.36)	0.21 (2.55)
Participation in industrial crop production (1= oil palm, 0=subsistence)	-0.39 (0.66)	-5.37 (3.99)
Constant	2.54*** (0.53)	11.03*** (3.67)
Treatment equation		
Off-farm income	0.003 (0.002)	0.002 (0.002)
MIHFP	0.21*** (0.06)	0.19*** (0.06)
Income	0.0004*** (0.0001)	0.0004*** (0.0001)
Origin of respondent	0.90** (0.45)	0.88** (0.44)
Gender of household head	-0.03 (0.32)	-0.013 (0.31)
Constant	-2.13*** (0.53)	-2.05*** (0.53)
/athrho	0.25 (0.31)	0.36 (0.24)
/lnsigma	0.42*** (0.08)	2.39*** (0.08)
rho	0.24 (0.29)	0.35 (0.21)
sigma	1.52 (0.12)	10.92 (0.86)
lambda	0.37 (0.45)	3.77 (2.48)
Number of observations	104	104
Wald chi2(12)	8.55	22.14**
Log likelihood	-244.36	-447.31

Notes: *** p <0.01; ** p < 0.05; * p < 0.1.

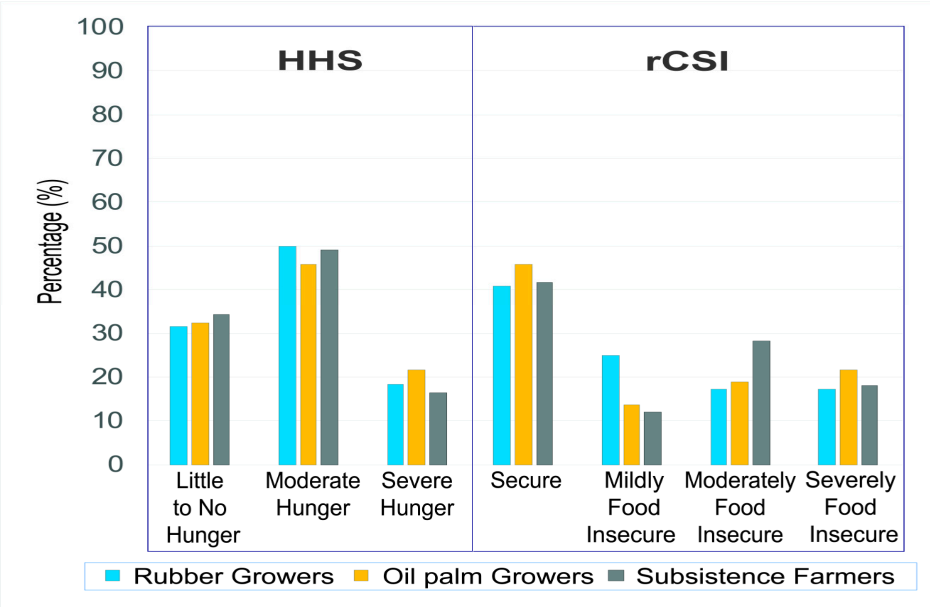

Figure S1. Distribution of respondents across food security thresholds for the HHS and rCSI metrics.

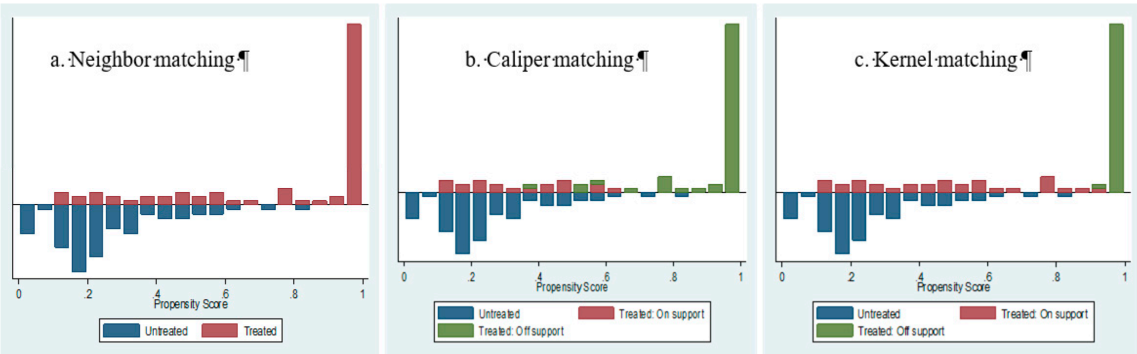

Figure S2. Propensity score distribution and common support for propensity score estimation for rubber growers and subsistence farmers.

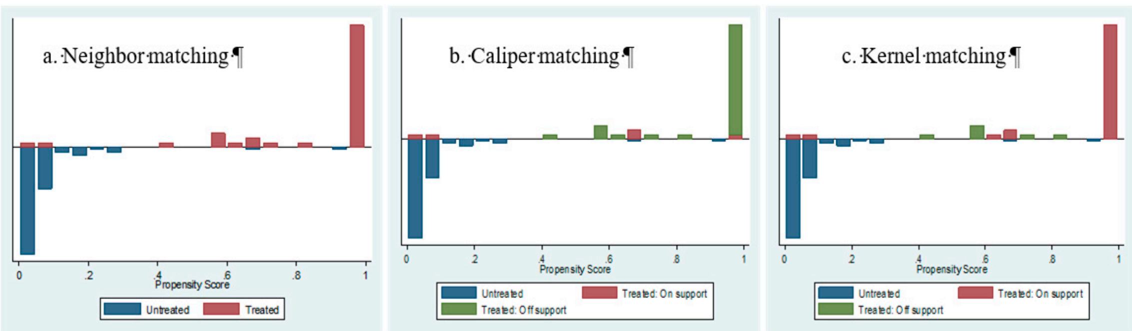

Figure S3. Propensity score distribution and common support for propensity score estimation of oil palm growers and subsistence farmers.