Appendix A

This document is prepared for the paper titled "*Commercial cash crop production and households' economic welfare: Evidence from the pulse farmers in rural China*", which has been submitted to Journal of Integrative Agriculture (JIA).

The potential correlations among the explanatory variables

One of the concerns about the Heckman two-step model estimation and the endogenous treatment regression model was whether there is a strong collinearity issues among the explanatory variables. If there is a presence of strong collinearity between the explanatory variables and the main independent variable, then our estimations might become unstable.

To check the correlations among all the explanatory variables, we conducted the Pearson correlations test, and we presented the results in Table A1. We find there is rather minimum correlations among the listed explanatory variables. Perhaps the largest correlation observed is between households' pulse market sales price and the percentage of their farmland are irrigated.

Table A1: Correlation matrix of the explanatory variables

	Price (1)	Cost (2)	Pop (3)	Irri (4)	Age (5)	Edu (6)	Mach (7)	Loan (8)	Size (9)	Per (10)
Price	1									
Cost	-0.191	1								
Pop	-0.265	0.227	1							
Irri	-0.273	0.212	0.169	1						
Age	-0.134	0.071	-0.042	0.121	1					
Edu	-0.023	0.085	0.091	0.170	-0.137	1				
Mach	0.232	-0.049	-0.207	-0.174	-0.062	-0.002	1			
Loan	0.159	0.018	-0.086	-0.113	-0.131	-0.034	0.019	1		
Size	0.239	-0.110	-0.133	-0.079	-0.221	0.053	0.021	0.194	1	
Per	0.089	-0.020	-0.042	-0.159	-0.029	-0.013	0.226	0.100	0.091	1

Note:

a. Price=Previous households' pulse sales price; Cost=Production cost per mu; Pop=Family size, in person; Irri=% of irrigated land; Age=HH's age; Edu=HH's education, in years; Mach=Agricultural machinery service; Loan=Having a loan for pulse production; Size=Total farm size; Per= Perceived climate change.

b. The significant negative correlation indicates that when households have a larger share of farmland can be irrigated, the households sell their pulse in a lower market price. This in fact matches our expectation since pulse demands less water compared to other staple crops. When households have more farmlands can be irrigated, they have high probability to plant other crops instead of pulse.

c. None of the reported correlations have a value larger than 0.3, overall, we are less concern about the potential collinearity among all the explanatory variables.

Robustness check

In the manuscript we defined commercial pulse farming households and subsistence pulse farming households by the size of pulse farming size. If the pulse farming size larger than 1 mu, then we categorized the households as commercial pulse farming households, vice versa, the households are defined as subsistence pulse farming households.

To service as a robustness check, we further tighten the criteria of being a commercial pulse farming households or a subsistence pulse farming household, we defined the subsistence pulse farming households if their pulse farming size is less than 0.5 mu, which is extremely small in our study background; while if the household pulse farming size larger than 2 mu, then we defined the households as commercial pulse farming households. Although by this definition, we have excluded to some extent a sample of pulse farming households who were farming in between 0.5 ~ 2 mu, it provides us better and more clear differences between these two groups. Results obtained from this robustness check can be good references for the main results we have provided in the manuscript. In Table A2, we present the Heckman two-step estimation results of the determinants of being a commercial pulse farming household. The results are quite close to the results reported in Table 4. In Table A3 and A4 we further reported the results of the endogenous treatment regression (ETR) estimated effect on households' income and their expenditure. Again, we observe rather similar results as in Table 5 and Table 6 in the manuscript.

Table A2: Determinates of pulse commercial growing (Defining commercial farmers if size>2 mu; subsistence pulse farmers if size<0.5mu)

	Stage 2 Pulse farming size Coefficients (1)	**Stage 1** Growing purpose Coefficient (2)
1. Farmer's total farm size,	0.001***	0.762***
	(0.000)	(0.201)
2.Percentage of irrigated farmland, %	-0.302**	1.878*
	(0.118)	(1.000)
3.Family population	-0.150***	0.272
	(0.032)	(0.197)
4.HH's age, in years	-0.024***	0.025
	(0.005)	(0.032)
5.HH head's education, in years	-0.046**	0.250**
	(0.019)	(0.125)
6.Market sales price of the previous year	0.219***	0.177
	(0.029)	(0.170)
7.Production cost per mu	-0.761***	
	(0.176)	
8.Agricultural machinery service	0.183	
	(0.117)	
9.Having a loan for pulse production, 1=yes	0.032***	
	(0.006)	
10.Agricultural production subsidies		0.737***
		(0.228)
11.Perception of climate change, 1=yes		1.297**
		(0.608)
12. Lambda (coefficient of m_{it})	0.762***	
	(0.201)	
Constant	3.808***	-10.687***
	(0.425)	(3.712)
Observations	636	636

Note:

a. *** p<0.01, ** p<0.05, *p<0.1, standard errors are in presented in parentheses.

Data source: Author's survey.

Table A3: Effect of commercial pulse planting on farmer's incomes (Defining commercial pulse farmers if size>2 mu; subsistence pulse farmers if size<0.5mu)

	Model 1		Model 2		Model 3	
	Total income	Selection equation	Non-farm income	Selection equation	Farm income	Selection equation
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
1.HH's age, in years	-0.030***	-0.007	-0.015	-0.008	-0.031*	-0.007
	(0.005)	(0.009)	(0.027)	(0.008)	(0.017)	(0.009)
2.HH's education, in years	-0.022	0.039	0.146	0.024	0.041	0.042
	(0.022)	(0.035)	(0.111)	(0.035)	(0.068)	(0.036)
3.% of irrigated land	0.257*	-0.827***	-0.848	-0.731***	2.036***	-0.908***
	(0.136)	(0.235)	(0.712)	(0.241)	(0.426)	(0.245)
4.Pulse size, log form	0.335***		-0.684**		0.688***	
	(0.07)		(0.348)		(0.217)	
5.Total farm size, log form	0.002***		0.001		0.001*	
	(0.001)		(0.001)		(0.001)	
6.2019.year	0.240**		-0.007		0.962***	
	(0.098)		(0.486)		(0.306)	
7.Province dummies	-1.211***		-6.248***		-0.895	
	(0.19)		(0.955)		(0.593)	
8.1.nc	-0.936***		-6.408**		-1.155	
	(0.358)		(3.00)		(1.222)	
9.Perception of climate change		0.586***		0.487***		0.651***
		(0.175)		(0.177)		(0.175)
10.Family size, in persons		-0.161***		-0.173***		-0.164***
		(0.057)		(0.052)		(0.056)
athrho		0.153		0.618**		0.319**
		(0.103)		(0.279)		(0.132)
lnsigma		0.207***		1.834***		1.345***
		(0.028)		(0.039)		(0.03)
Constant	13.246***	2.430***	13.309***	2.591***	8.832***	2.427***
	(0.465)	(0.64)	(3.096)	(0.604)	(1.515)	(0.637)
Observations	639	639	640	640	636	636

Note:

a. *** p<0.01, ** p<0.05, *p<0.1, standard errors are in presented in parentheses.

Data source: Author's survey.

Table A4: Effect of commercial pulse planting on expenditure (Defining commercial pulse farmers if size>2 mu; subsistence pulse farmers if size<0.5mu)

	Model 1		Model 2	
	Total expenditure*	Selection equation	Food expenditure*	Selection equation
VARIABLES	(1)	(2)	(3)	(4)
1. HH's age	-0.03***	-0.01	-0.02***	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)
2. HH's education, in years	-0.01	0.04	0	0.04
	(0.02)	(0.04)	(0.02)	(0.04)
3. % of irrigated land	-0.04	-0.82***	0.40***	-0.83***
	(0.15)	(0.23)	(0.13)	(0.24)
4. Pulse size, log form	0.05			
	(0.08)			
5. Total farm size, log form	0.01**		0.01**	
	(0.01)		(0.01)	
6. 2019.year	0.24**		0.64***	
	(0.11)		(0.1)	
7. Province dummies	-0.76***		-0.69***	
	(0.21)		(0.19)	
8. 1.nc	-0.63		-0.95***	
	(0.4)		(0.33)	
9. Perceived climate change		0.58***		0.61***
		(0.17)		(0.18)
10. Family size, in person		-0.17***		-0.17***
		(0.06)		(0.06)
athrho		0.15		0.20**
		(0.1)		(0.09)
lnsigma		0.33***		0.19***
		(0.03)		(0.03)
Constant	12.49***	2.43***	10.13***	2.47***
	(0.52)	(0.64)	(0.44)	(0.64)
Observations	640	640	640	640

Note:

a. *** p<0.01, ** p<0.05, *p<0.1, standard errors are in presented in parentheses.

Data source: Author's survey.