Social interactions and household fuel choice: evidence from rural China

Hang Fang,[1,2] Xuechen Jiang,[1] Qianheng Chen,[3] and Michael S. Delgado[4*]

[1]Department of Economics, Anhui Agricultural University, Hefei, Anhui, China, [2]Green Food Industry Chain in Philosophy and Social Science of Anhui Key Lab, Hefei, Anhui, China, [3]Department of Agricultural Economics, China Agricultural University, Beijing, China, [4]Department of Agricultural Economics, Purdue University, West Lafayette, IN, USA

[*]Corresponding author. E-mail: delgado2@purdue.edu

ONLINE APPENDIX

A.1 Derivation of the equilibrium

As noted, our theoretical model follows Brock and Durlauf (2001). We specify the random term $\varepsilon(Y_i)$ as $i.i.d.$ extreme-value, which leads the difference between the positive and negative choice on the random term to be logistically distributed. Thus the probability that a household chooses clean fuel is

$$Prob(Y_i) = \frac{exp(x_i Y_i + \lambda Y_i \bar{Y}_i^e)}{\sum_{Y\epsilon\{1,-1\}} exp(x_i Y_i + \lambda Y_i \bar{Y}_i^e)}, \tag{A1}$$

and the aggregate probability for all choices is

$$Prob(Y) = \frac{exp(\sum_{i=1}^{N}(x_i Y_i + \lambda Y_i \bar{Y}_i^e))}{\prod_{i=1}^{N}\sum_{Y\epsilon\{1,-1\}} exp(\sum_{i=1}^{N}(x_i Y_i + \lambda Y_i \bar{Y}_i^e))}. \tag{A2}$$

When $\lambda > 0$, household fuel choice is no longer independent and the social interactions effect is just the effect of the average village choice (captured by λ).[1] Furthermore, the expected value of a household's fuel choice given the expectation of its neighbors' choice is

$$E(Y_i) = \frac{exp\left(x_i + \lambda(N-1)^{-1}\sum_{i\neq j} Y_{ij}^e\right) - exp\left(-x_i - \lambda(N-1)^{-1}\sum_{i\neq j} Y_{ij}^e\right)}{exp\left(x_i + \lambda(N-1)^{-1}\sum_{i\neq j} Y_{ij}^e\right) + exp\left(-x_i - \lambda(N-1)^{-1}\sum_{i\neq j} Y_{ij}^e\right)}$$

$$= tanh\left(x_i + \lambda(N-1)^{-1}\sum_{i\neq j} Y_{ij}^e\right). \tag{A3}$$

Under rational expectations, for any households i and j from the same village, $Y_{ij}^e = E(Y_j)$, so that all households have the same overall average expectation about village fuel transition. The expected value is

$$E(Y_i) = tanh\left(x_i + \lambda(N-1)^{-1}\sum_{i\neq j} E(Y_j)\right). \tag{A4}$$

Since $tanh$ is a continuous function, there is at least one self-consistent equilibrium, $E(Y_i) = E(Y_j)$; there must be an expectation of average choice Y^* such that

$$E(Y^*) = tanh(x_i + \lambda Y^*). \tag{A5}$$

[1] The model is a common standard logit model if $\lambda = 0$ and the choice set $\{-1, 1\}$ is replaced with $\{0, 1\}$.

Yet, the solution Y^* need not be unique, and this is where the situation of multiple equilibria arises. This response function in equation (A5) provides a means of visually understanding the potential for multiple equilibria. The function is determined by two parameters: x_i, which reflects marginal private utility, and λ, which reflects marginal social utility. First, we observe the impact of a change in λ on the equilibrium while setting the marginal private utility to be fixed at $x_i = 0$. In figure A1, the horizontal axis is the expected value of the household's fuel choice, and the vertical axis is the expected value of the average village choice. The 45-degree line shows the set of points where the expected value of the household's choice is consistent with the average village choice, so that, the intersection of the 45-degree line and the response function is a self-consistent equilibrium point as defined by Y^*.

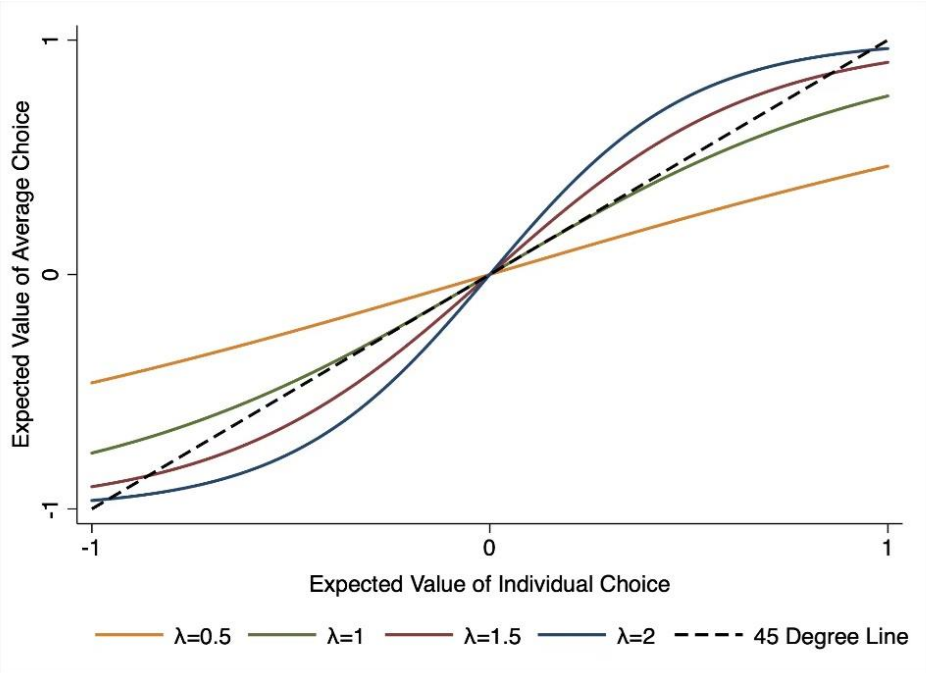

Figure A1. Equilibrium and social utility.

The four different curves in figure A1 correspond to the response function with the parameter λ set as 0.5, 1, 1.5, and 2, respectively. The shape of the response function is determined by the size of λ. A larger λ corresponds to a more curved function, whereas a smaller parameter value corresponds to less curvature. The greater the curvature of the response function, the more likely there are multiple intersections corresponding to multiple equilibria that arise when the social effect is relatively stronger.

Next, we fix $\lambda = 2$ and adjust marginal private utility, x, to observe the effect of changes in marginal private utility on the equilibrium. Figure A2 shows the response function with x equal to 0, 0.5, 1 and 1.5, respectively. As the marginal private utility increases, the probability of multiple equilibria gradually decreases. Based on these two comparative static analyses (figures A1 and A2), it is clear that multiple equilibria may emerge when the social interactions effect is relatively larger than the private incentive.

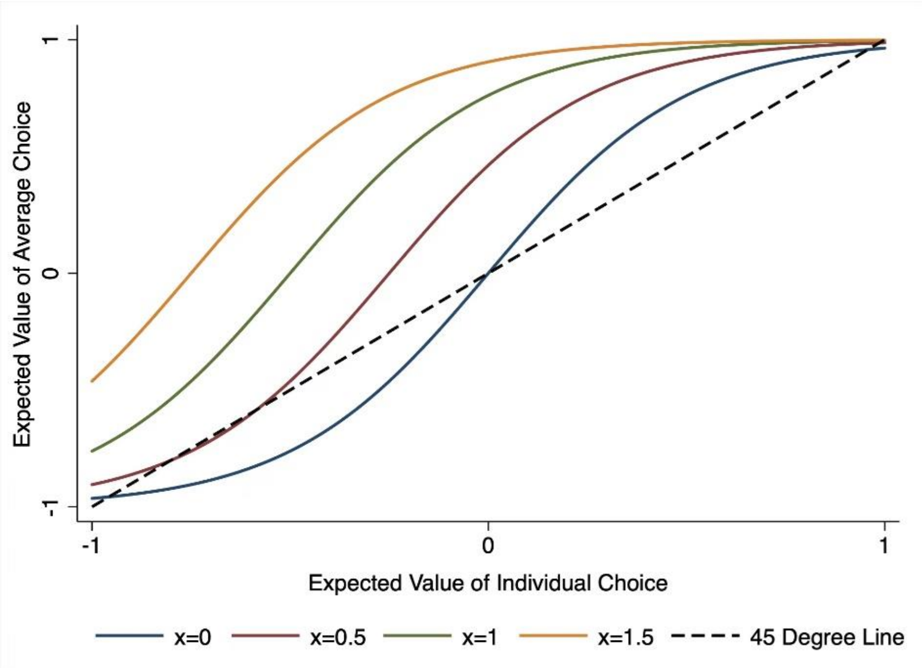

Figure A2. Equilibrium and private utility.

A.2 Robustness checks

A.2.1 Resizing group size

As a robustness check, we adjust the group (village) size as this sizing determines the peer reference group for the household. In a study of peer effects on academic achievement, Hsieh and Lee (2016) set the minimum group size to 11 so that each student has at least 10 peers. Compared with this setting, the minimum size of our groups (which is 6) is somewhat smaller. Following Hsieh and Lee (2016), we restrict the minimum group size to 11. These regression estimates are shown in column (1) of table A1, in which we see that the estimate of the effect of the average village adoption rate is still significant at the 1 per cent level, and is very close to that of the baseline model shown in table 2.

A.2.2 Household unobservable characteristics

In the above models, we use village-level fixed effects to account for the correlation effect. However, the village-level fixed effect can only deal with the unobserved factors between groups but not within the group (Lin, 2010). Characteristics like family members' habits and frugality can also affect fuel choice, and these household-level unobservable factors may also lead to biased estimates of the social interaction effects. To examine the potential impact of household-level unobservable characteristics, we use a panel fixed effects model to re-estimate the social interaction effects model.[2] The household fixed effects allow us to control for households' time-invariant unobserved characteristics, and this also creates a robustness check in terms of how we account for the correlated effect since the household fixed effect can simultaneously account for any unobserved village (or other group) effect (Bramoullé *et al*., 2020). Column (2) of table A1 reports the result: there remains a significant endogenous social effect for fuel choice after controlling for household-level fixed effects.

[2] We also try the random effects model, and a Hausman test indicates that the fixed effects model provides a better fit than the random effects model.

Table A1. The social interactions effect of fuel choice: robustness checks

	(1) Logit	(2) Panel FE	(3) IV-Probit
Fuel_peer	1.917	0.681	4.480
	(0.112)	(0.017)	(1.757)
Income	0.216	0.011	0.118
	(0.014)	(0.002)	(0.011)
Size	-0.063	-0.001	-0.031
	(0.010)	(0.002)	(0.006)
Elderly	-0.745	-0.024	-0.382
	(0.044)	(0.009)	(0.048)
Edu	0.523	0.018	0.281
	(0.066)	(0.015)	(0.039)
Agri	-0.063	-0.004	-0.034
	(0.004)	(0.001)	(0.003)
Business	0.911	0.030	0.479
	(0.057)	(0.010)	(0.058)
Income_peer	0.010	-0.004	-0.117
	(0.056)	(0.007)	(0.066)
Size_peer	-0.034	-0.001	0.026
	(0.039)	(0.005)	(0.033)
Elderly_peer	-0.044	-0.003	0.451
	(0.260)	(0.033)	(0.291)
Edu_peer	-0.339	-0.046	-0.090
	(0.332)	(0.038)	(0.195)
Agri_peer	0.009	0.003	0.013
	(0.023)	(0.002)	(0.013)
Business_peer	0.049	-0.006	-0.492
	(0.278)	(0.034)	(0.294)
Constant	-2.122	0.041	-2.472
	(0.660)	(0.071)	(0.758)
Time fixed effects	Yes	Yes	Yes
Village fixed effects	Yes	No	Yes
Household fixed effects	No	Yes	No
Observations	37577	38747	38278
Pseudo R^2 or R^2	0.354	0.495	

Notes: Robust standard errors in parentheses.

A.2.3 Placebo test

Although the village setting is reasonable in the social context of rural China (as discussed at the beginning of section 2), it is nevertheless important to test the reliability of the village reference group setting. We turn to a placebo test. We first randomly match each household with several other households from the database (irrespective of the village) in the same period to be the placebo reference group, with the size of the placebo reference group set equal to the average reference group size in the data (which is approximately 21 but varies slightly in different years). Second, we calculate the leave-i-out means based on those placebo reference groups and re-estimate the model following equation (2). Finally, we repeat the above steps 100 times to avoid any erroneous placebo results based on a single random placebo trial. If the original (i.e., true) village reference group is valid, then we expect the estimated endogenous social effect to be statistically insignificant across the 100 placebo trials. Instead, if we find broad significance of the social interaction effects in the placebo trials, we would be concerned about the validity of our primary regression model results. A summary of the results from the placebo trials are shown in figure A3. There are only five placebo trials for which the endogenous social effect is significant at the 5 per cent significance level, and in only eight placebo trials do we see a significant endogenous social effect at the 10 per cent significance level; together, we interpret these trials to indicate that the village reference group is valid because the majority of the placebo estimates are insignificant.

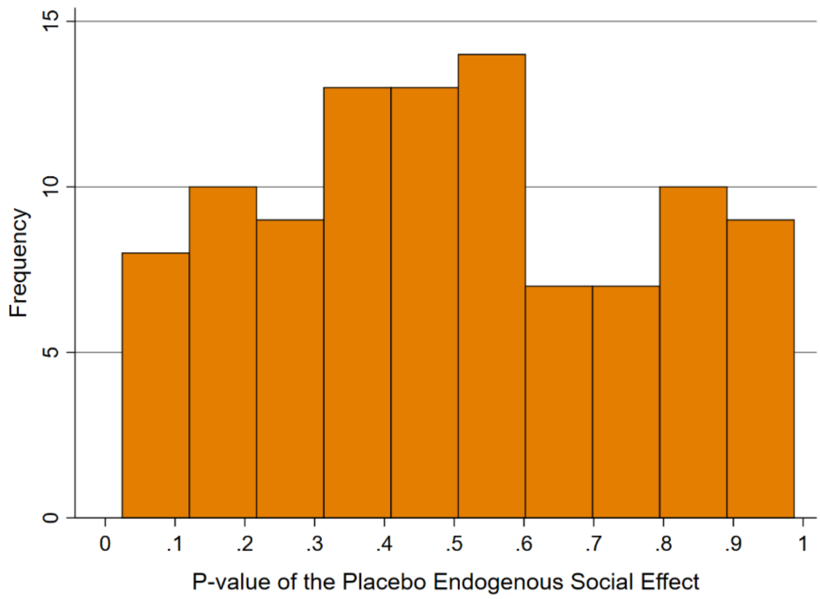

Figure A3. Summary of the results from the placebo tests.

A.2.4 Endogeneity concerns

Finally, despite the theoretical structure, a cautious reader might remain concerned about the possibility of reverse causality leading to an endogeneity problem. An increase in the village average adoption rate affects the individual's adoption decision, but one might still be concerned that the individual's decision might determine the village adoption level. Although we have noted that the impact of the individual's decision on the village adoption rate is negligible when the village size is large, for the sake of robustness, we use the IV probit model to address these concerns. Inspired by Duflo and Saez (2002) and Wen *et al*. (2021), we use the proportion of households that belong to the top 25th percentile of per capita income in each year from each village as the IV. The underlying assumption is that income is an important factor affecting fuel selection, but once household income, contextual effects, and village-level fixed effects are controlled for in the second-stage model, the IV is unlikely to directly affect household decision-making. Column (3) of table A1 shows the estimated results, and it is clear that we still find a significantly positive endogenous social effect.[3]

[3] In the first stage, the instrumental variable positively affects the village average adoption rate, which is consistent with our expectation. Via a Wald test, we fail to reject the null hypothesis that the correlation coefficient between the first and second stage residual terms is equal to zero, which means that the village average adoption rate is exogenous, and the basic model does not have an endogeneity problem.

References

Bramoullé Y, Djebbari H and Fortin B (2020) Peer effects in networks: a survey. *Annual Review of Economics* **12**, 603–629.

Brock WA and Durlauf SN (2001) Discrete choice with social interactions. *Review of Economic Studies* **68**, 235–260.

Duflo E and Saez E (2002) Participation and investment decisions in a retirement plan: the influence of colleagues' choices. *Journal of Public Economics* **85**, 121–148.

Hsieh C and Lee LF (2016) A social interactions model with endogenous friendship formation and selectivity. *Journal of Applied Econometrics* **31**, 301–319.

Lin X (2010) Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables. *Journal of Labor Economics* **28**(4), 825–860.

Wen H, Wang C and Nie P (2021) Acceleration of rural households' conversion to cleaner cooking fuels: the importance and mechanisms of peer effects. *Energy Policy*, **154**,112301.