Road to Specialization in Agricultural Production:

Tales of 18 Natural Villages in China

Yu QIN **yq38@cornell.edu**

The Dyson School of Applied Economics and Management

Cornell University, Ithaca, NY, 14850

Xiaobo Zhang **X.ZHANG@cgiar.org**

International Food Policy Research Institute

2033 K St. NW Washington, DC, 20006-1002

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association's 2011 AAEA & NAREA Joint Annual Meeting, Pittsburgh, Pennsylvania, July 24-26, 2011.

Abstract:

We use a panel survey at the household level primarily conducted in 18 natural villages over three waves in Guizhou province, China, to study how improved road access shapes farmers' cropping and input use. Our results show that access to roads significantly improves the level of specialization in household agricultural production as measured by the Herfindahl-Hirschman Index. In villages with better road access, farmers plant fewer crops and invest more intermediate input. Through specialization and increased use of intermediate input, road connections improve farmers' agricultural income and total income per capita. In addition, the cash income of agricultural product also increases, providing evidence of more trading of produced goods due to better connectivity. However, better access to rural roads does not seem to bring about significant changes in non-agricultural income.

"To get rich, build road first."

-old Chinese proverb

Introduction

Rural roads have been consistently documented as one of one the key instruments for overcoming spatial poverty traps and promoting rural nonfarm economy in developing countries (Calder 2009; Fan and Hazell 2001). However, because rural areas are often remote and hard to reach, it may be costly to build roads that improve access for these areas. Therefore, a rigorous impact assessment of roads would provide useful information for policy interventions. In developing countries, the rural poor often live in isolated areas. Being far away from the market, the poor are more likely to rely on self-sufficient farming to survive. Road connections reshape farmers' production choice sets, labor opportunity costs, and transportation costs to the outside world. Therefore, rural roads will likely affect agricultural production in at least two ways.

First, farmers may find more profitable opportunities in the nonfarm sector outside their villages after improved road access. Thus, they may change their approach toward agricultural production from one that is autarky-based and self-sufficient to a more market-oriented specialized production focus (Limao and Venables 2001; Renkow et al. 2003). Yang and Ng (1993) develop a theoretical model, showing that producers will choose to specialize in their production and simply purchase other goods from the market, provided that transaction costs are sufficiently small. Using a simulation approach, Omamo (1998) finds that as distance increases, small-scale farmers tend to move toward a more diversified cropping pattern and away from monocropping. However, the empirical findings are mixed. For example, in an empirical study, Stifel et al. (2003) show that in Madagascar the concentration level of agricultural production in the least remote areas is around 1.5 times that of the most remote areas. In other words, improved road access facilitates specialization in agricultural production. However, Gibson and Rozelle (2003) find that in Papua New Guinea, each extra hour to reach the nearest road induces a 2.6 percent reduction in the number of income earning activities.

Second, as improved road access reduces transportation costs, the price of intermediate inputs, such as fertilizer, is more likely to drop. Consequently, farmers may adopt more intermediate input use to increase their agricultural productivity. Gollin and Rogerson (2010) propose a theoretical model, suggesting that when transportation cost decreases, farmers will begin to use more intermediate inputs. The empirical findings on the impact of rural road improvements on input use, however, are inconclusive. In the empirical studies, Benziger (1996) finds that improved road access leads to increased fertilizer use in villages in Hebei, China. In the context of Madagascar, Stifel et al. (2003) show that farmers in more isolated regions use less fertilizer than those in places with relatively better road access. However, Dorosh et al. (2010) paint a more complicated story. Input use depends upon not only distance to roads but also the density of road networks. For example, in East Africa, reducing travel time significantly increases adoption of high-input/high-yield technology, while the impacts of road improvements are insignificant in West Africa, where the density of road networks is relatively higher at the beginning.

One difficult challenge of empirically evaluating the impact of improved road access on agricultural production is the factor of data limitations. Most of the empirical studies rely on cross-sectional data, making it hard to control for unobserved factors. To overcome this challenge, in this paper, we use a panel dataset at the household level primarily conducted in 18 natural villages over three waves in Guizhou province, China, to study how road access shapes farmers' cropping patterns and input use.

In the past several decades, the Chinese government has made huge investments to build a nationwide highway system. As the highway density increases, the marginal returns to future investments in highway increase as well. Fan and Chan-Kang (2005) argue that it may make more economic sense to shift such investment toward rural roads. However, rural roads carry less traffic, they are harder to maintain, and they are often more costly to build due to remoteness. Therefore, it is important to gather more evidence-based information about how rural roads affect agricultural patterns and rural livelihood. The information would be useful for policymakers to make better allocations of public investment. Moreover, as most of the rural poor reside in remote areas in China, rural road construction may play an important role in poverty alleviation.

While the literature on the impact of rural roads in China is limited, some discussion exists. Based on aggregated data at the provincial level, Fan and

Chan-Kang (2005) find a significant impact of road investment on both rural non-farm GDP and agricultural GDP, although the magnitude is larger relative to rural non-farm GDP than to agricultural GDP. However, China entered an era of labor shortage in 2003 (Zhang et al. 2011), suggesting that the marginal impact of roads on non-farm GDP may not be as large as it was in previous years. Still, it is widely known that it is hard to tackle the endogeneity issue using aggregate data. We try to examine the impact of rural road on agricultural production based on primary household data collected in remote rural villages in China over three waves, during which some of these villages built roads.

The data was collected for all the households in 18 natural villages within three of the administrative villages in Guizhou province China in the beginning of 2005, 2007 and 2010. The data includes detailed information on household agricultural production as well as other sources of household income. In addition to those items, the survey also collected public investment information at the natural village level. Several of the surveyed natural villages built roads during 2004-2009, thus enabling us to quantitatively evaluate the impact of road connection using the panel dataset.

Our data provides two advantages for studying the impact of access to road networks on isolated villages. First, as a non-recall panel data, our study provides relatively accurate and credible information with respect to household agricultural production. Second, the three waves of data enable us to conduct a difference-in-difference analysis, which helps to mitigate the bias of estimation due to the endogeneity of access to roads. To the best of our knowledge, this is the first paper to empirically document the causal impact of improved access to roads on agricultural specialization in China.

The main finding of our paper is that improved access to roads significantly increases the level of household agricultural specialization. The magnitude of this impact is around six percent as measured by the Herfindahl-Hirschman Index on a scale of zero to 1, which is around 13 percent relative to its mean level. In addition, there is also evidence suggesting that better connectivity for the villages lead to more intensive intermediate input use, an increase in agricultural income and total income per capita. However, the introduction of road access does not seem to improve farmers' non-agricultural income.

Our findings have important policy implications. As China's agricultural product prices has risen in recent years, it is of interest to determine whether improved road access motivates farmers to abandon their land, possibly leading to price inflation in the agricultural sector. However, our results show that roads actually facilitate the division of labor in the villages without hurting the output in the agricultural sector.

Our findings should be generalized with caution as our sample is generated from the mountainous rural areas in southwestern China, where small-scale farming is the dominant agricultural activity. As China is a large and spatially diverse country, the findings drawn from this sample may be not applicable to China as a whole.

Description of Data

As one of the poorest provinces in China, Guizhou has the least miles of road, due in part to its constraining mountainous geography. In these mountainous areas, households in the isolated villages live in a subsistence economy, whereas households in relatively level areas are much better connected to the market by roads. The large variation in respective levels of road access in our sample provides us with an opportunity to study the impact of road access on agricultural production in China.

The survey site is Puding County, which consists of 11 townships, 317 administrative villages and a total population of 448,000 people as of 2008.[1] The average household income in Puding County is around 5,800 *yuan* in the year of 2008, which is above the provincial median but below the provincial mean, suggesting that it is fairly representative of Guizhou province as a whole.

Three administrative villages that represent the broad range of economic development of Puding were chosen by survey enumerators, of which one is an author of this paper. A census-type survey of all 805 households in the 18 natural villages comprising the three administrative villages was administered in early 2005. The second wave of survey was then administered in early 2007 and included 833 households, among which several had resulted from new marriages. The third wave of survey was administered in early 2010, including 873 households.[2] The surveys collected detailed information relating to demographics, income, agricultural production and consumption. Information was collected for each household member, including members that were working outside the county at the time of the surveys.

[1] An "administrative village" is a bureaucratic entity comprised of several "natural" or "typical" villages.

[2] Valid observations are 782, 815 and 834 households in the three waves, respectively.

The three surveyed administrative villages exhibit variations in their levels of road access. In this paper, roads are defined as those that tractors can drive on during the rainy season. Table 1 presents the distribution of road access in the surveyed villages in 2004, 2006 and 2009. The information was collected from the village cadres' records. It can be seen from the table that Xinlin Village has the extent of road access in the three administrative villages. All the natural villages of Xinlin administrative village have been connected with roads since the first wave of the survey. Four natural villages in the Changtian administrative village—*Dongqingzhai, Yanshang, Baobaozhai and Pianpo*—constructed roads during our survey periods. However, until our most recent survey, some villages, such as *Yezhuwo* and *Yeniuzhuang*, still had yet to gain road access. In the third administrative village, the natural village of Xiadaba built a new road during 2004 and 2006, while the households in two of other natural villages were still living without road access until our most recent survey.

As we are interested in the impact of improved road access on agricultural specialization, we constructed an index similar to the Herfindahl-Hirschman index to measure varying degrees of specialization at the household level. The Herfindahl-Hirschman index is defined as the sum of the squares of the market shares of the firms in a certain industry and is used as a measure of the level of concentration in the industry. In the context of agricultural production, we use the sum of squares of the agricultural income share of the production activities to measure the level of specialization in household agricultural production.[3] Therefore, the specialization index is defined on [0, 1], where a greater the value reflects a higher level of specialization.

The households in the surveyed villages produced varieties of agricultural products as presented in Table 2. It can be seen from the table that corn income accounts for almost half of total household agricultural income in all the three surveyed years, ranging from 39 to 46 percent. As the second largest income source, oil seeds provide 16 to 21 percent of household agricultural income. Livestock represents the third most important source of agricultural income. However, only about 27-30 percent of households were engaged in livestock production, comparing to an approximately 90 percent participation rate in corn and rapeseed production.

It is worth noting that the categories of agricultural income decomposition are slightly different in the three waves of surveys due to questionnaire design inconsistency. For example, in the surveys of 2004 and 2009, there are nine sub-categories of agricultural income, while there are ten sub-categories in the 2006 survey. Additionally, there is no data available for vegetable income in year 2009. Therefore, we construct two alternative measures of the specialization indices as robustness checks. In the first alternative measurement, we predict the vegetable income in year 2009 based on the vegetable seed cost available in 2009 and the past relationship between vegetable seed cost and income in the previous surveys (variable

[3] For example, if the household produces maize and fruit, with an income of 2,000 *yuan* and 3,000 *yuan*, then the specialization index is calculated as $(2000/5000)^2+(3000/5000)^2=0.52$.

denoted as HHI [2]). In so doing, we obtain comparable household vegetable income for all the three waves. For the second alternative measurement, we categorize the non-overlapping subcategories and the rest of the income as “other.” After this adjustment, there are nine comparable subcategories of agricultural income across the three waves (variable denoted as HHI [3]).

Table 3 presents the summary of statistics for the key variables used in the analysis and shows that the average household income almost doubled during our survey periods. Average household income rises from 6,644.11 *yuan* in year 2004 to 11,994.57 *yuan* in year 2009. The rising income is mainly due to the increase in household non-agricultural income, which grew from 3,793.26 *yuan* in year 2004 to 6,540.62 *yuan* in year 2009. In comparison, the average household agricultural income has increased at a slower pace from 3,978.44 *yuan* to 5,453.96 *yuan* during the five years. Arable land is rather limited in our sample villages with an average land cultivated per person of 0.74 *mu*[4] in year 2009, lower than the national average of 1.4 *mu* per person in the year of 2006.

Lastly, the mean level of household agricultural specialization index is 0.46, 0.41 and 0.49 in the year 2004, 2006 and 2009, respectively. There are two possible reasons to explain the relatively low level of specialization in the year 2006. First, as the 2006 survey contains ten sub-categories of agricultural income while the other two waves contain only nine sub-categories, it is likely the measured level of specialization in the year 2006 will be lower. When we add the predicted vegetable

[4] 1 *mu*=0.066667 hectare.

income into the agricultural income in the year 2009 and re-categorize the agricultural income in the three waves into nine sub-categories, the mean of household agricultural specialization index in 2009 decreases from 0.49 to 0.47. Second, there were heavy droughts in 2005 and 2006 in Puding that affected almost all of the natural villages in our surveyed area. As a possible consequence, the share of corn income in the year 2006 is 39 percent lower than year 2004 (46 percent) and 2009 (42 percent), which is the main force driving down the specialization index in 2006.

When comparing the summary statistics by looking at the extent of road access, we find stark differences between the groups with and without road access. As shown in Table 4, the mean household income villages with road access is nearly two times as high in villages without road access. Both the agricultural and non-agricultural income in households with road access are higher than the households residing in disconnected villages. In terms of household agricultural specialization, the villages with roads are more specialized, though there is only a two percent difference in the specialization index between the two groups before regression adjustment. In general, households with road access tend to have smaller household size, larger area of cultivated land, higher education level and slightly lower number of active laborers.

Empirical Model

Our empirical question is: does road access have any impact on specialization and input use in agricultural production? In this paper, we adopt a difference-in-difference method to answer this question. The specification is as follows:

$$Y_{i,t} = \alpha_0 + \beta_1 \text{Connected}_{i,t} + Z_{i,t} + \varphi_{\text{village}} + \psi_{\text{year}} + \varepsilon_{i,t}$$

where $Y_{i,t}$ is a dependent variable for households i in time t; $\text{Connected}_{i,t}$ denotes whether the village individual i belongs to has road access in year t; $Z_{i,t}$ represents a series of control variables, including the acres of land cultivated by the household, the number of active labor force aging between 16 to 60 in the family, household size, the highest year of education in the household and whether there is village cadre in the household; φ_{village} stands for natural village fixed effects; ψ_{year} controls for year fixed effects; and $\varepsilon_{i,t}$ represents the error term.

Our coefficient of interest is β_1, which represents the impact of road access on the dependent variable. To answer our question, the main dependent variables in our estimation are (i) the household agricultural specialization index (HHI [1], HHI [2] and HHI [3] in the regression tables and the following discussions); (ii) input use measured by the natural log of monetary value of input use per *mu* of land, including seed cost, pesticide, fertilizer, animal traction and irrigation; (iii) the natural log of agricultural income, non-agricultural income and total income per capita in the household; and (iv) trading of agricultural products measured by agricultural cash income. If road access significantly leads to agricultural specialization, we should see β_1 to be positive and significant. Similarly, β_1 should be positive and significant as well if access to road significantly increases the level of input use and household income.

As the impact of roads may differ in the short run and in the long run, we allow for this difference by adding one more variable denoting a “new road” effect, which is as follows:

$$Y_{i,t} = \alpha_0 + \beta_1 Connected_{i,t} + \beta_2 Connectednew_{i,t} + Z_{i,t} + \varphi_{village} + \psi_{year} + \varepsilon_{i,t}$$

where $Connectednew_{i,t}$ denotes whether the village that individual i belongs to has road access built within two years of year t. Therefore, the short run impact of road is $\beta_1 + \beta_2$, while the long run impact of road is β_1.

As suggested by recent impact evaluation literature (Duflo and Pande, 2007), the non-random placement of the policy may bring endogeneity problems into the estimation. A typical solution to this problem is to carry out a Two-stage Least Square analysis by instrumenting the policy with a set of exogenous variables. However, as the variation of road access in the sample is at the natural village level and there are only 18 natural villages in the dataset, it is not appropriate to implement the first stage regression with such small observations.

However, we believe that the possible endogeneity or road placement will not cloud our findings for several reasons. First, if roads were placed in villages with higher growth potential, we should see that villages with roads before the first wave (year 2004) had higher growth rates than villages without road at that time. Moreover, we should see that the villages with roads before 2004 had higher growth rates than villages with roads after 2004, as higher priority with road construction indicates higher growth potential. However, the average net income growth rate for villages with roads during the three waves is 14.70 percent during 2004 to 2009, higher than the villages with roads before 2004 (12.99 percent) and villages without roads until 2009 (10.26 percent.) The above evidence shows that roads are not placed in villages with higher growth potential.

Second, as road construction requires relatively large investment, a significant portion of funding comes from the above-level government. For example, the funding from the above government on average accounts for 71 percent of the total investment for road construction in Puding County, according to a natural village survey conducted by the author in 285 natural villages in Puding, which includes all 18 villages in our dataset. As the funding from the above government depends on the liquidity of its finances, the timing of the road construction is more likely to be exogenous to the village itself.

Third, even if the road construction is endogenous to the village growth, our outcome variables, especially the specialization index and the level of input use in land, are not likely to directly affect road construction decisions. Therefore, the impact of endogeneity on the estimation to these outcomes is minimal.

Empirical results

The main result of road impact on agricultural specialization is shown in Table 5. Three specialization indices, namely HHI (1), HHI (2) and HHI (3), are used in the regressions as discussed in the above section. The first column of each regression presents the result for the most parsimonious specification, while the second and third columns allow for different impact of road in the short run and in the long run. As shown in the first and second rows of Table 5,the impact of roads on agricultural specialization is positive and significant, while the impact of “new road” is insignificant, suggesting that the short run and long run impact of roads do not differ much. Furthermore, regressions using three slightly alternative specialization indices yield consistent results in both magnitude and significance. On average, the introduction of roads will lead to a six percent increase in specialization measured by the HH indices, which amounts to around 13 percent of its mean level.

Table 6 presents the results for the impact of roads on input use. In the short run, the introduction of roads increases the input use per *mu yuan* by 53.7 percent (after translating the log form coefficient 0.43 into real growth rate), while in the long run the effect is not significant.

As road access facilitates agricultural specialization, shifting a village from an autarky world toward a more integrated economy, it is of interest to test whether more agricultural products are sold to the market after roads are constructed. As shown in Table 7, the coefficient for the short-run impact of roads on agricultural cash income is 0.38 (significant at 0.01). In other words, the cash income generated from agricultural activities increases by 46.2% after road construction.

Table 8 presents the findings for the impact of roads on household income, including agricultural income, non-agricultural income and total income per capita. There is a short-run impact of roads on agricultural income growth; that is, road access increases agricultural income by 29.7 percent (after translating the log form coefficient 0.26 into real growth rate), while the long-run impact is not significant. Moreover, the impact of roads seems to have no impact on the non-agricultural sector, as the effect of roads on non-agricultural income is neither significant in the short run nor the long run. In the short run, road construction boosts total income per capita by 15% (after translating the log form coefficient 0.14 into real growth rate).

Robustness Checks

In this section, we assume that the short-run and long-run effect of roads are the same, which enables us to implement a standard difference-in-difference analysis as follows:

$$Y_{i,t} = \alpha_0 + \beta_1 \mathrm{Road}_i + \alpha_0 + \beta_2 \mathrm{Road}_i * \mathrm{Beforeafter}_{i,t} + Z_{i,t} + \varphi_{village} + \psi_{year} + \varepsilon_{i,t}$$

where Road_i denotes whether the natural village individual i lives in has roads or not by the end of year 2009; $\mathrm{Beforeafter}_{i,t}$ denotes whether roads are constructed before time t or not. Therefore, β_2 is the coefficient of interest for the impact of roads.

Table 8 presents the results using this standard difference-in-difference specification. In order to save space, we only report the coefficient of interest, β_2, in this table. The impact of road on the level of agricultural specialization remains significantly positive at the 0.01 level with similar magnitude compared with the baseline estimation. The impact of roads on agricultural income is positive as well, but the significance decreases from the 0.05 level to the 0.1 level. The impact of roads on input use is not significant in this specification.

Conclusions

In this paper, through the use of primary census-type household surveys in remote villages in China, we examine the impact of road access on agricultural production, especially on the level of specialization and input use. We find that improved access to roads helps to facilitate agricultural specialization as well as to increase the level of input use, positively contributing positively in the end to the growth in agricultural income and total income per capita. However, road construction does not increase non-agricultural income.

A possible reason for the insignificant impact on the non-agricultural sector is that the era of rural surplus labor in China has been over since 2003 (Zhang et al. 2011). Since 2003, farmers have increasingly relied on remittances as the major nonfarm income. In this remote area, farmers' migration decisions may have little to do with local infrastructure conditions. However, road access may help facilitate market integration of the agricultural economy, thus enlarging the production scale of products with comparative advantage. For example, in *Xiadaba* natural village, the natural endowment is suitable for growing peaches. Before road construction, peaches were often damaged after being carried by hand to the nearest market, often over long distances. After building roads, farmers are able to sell their peaches at a collection point right in their natural village. As a result, peach production has significantly increased.

We add a note of caution that the finding of the positive impact of road investment on agricultural production do not necessarily mean that roads should be built to connect all remaining natural villages, as the marginal cost of building roads to the more remote communities may far outweigh the benefit. Thus, a cost-benefit analysis is needed when considering such rural road projects.

References

Benziger, V. 1996. "Urban Access and Rural Productivity Growth in Post-Mao China." *Economic Development and Cultural Change,* 44(3): 539–570.

Calderón, C. 2009. "Infrastructure and Growth in Africa." *World Bank Policy Research Working Paper Series No. 4914*.

Dorosh, P., H.G. Wang, L. You, and E. Schmidt. 2010. "Crop Production and Road Connectivity in Sub-Saharan Africa: A Spatial Analysis." *World Bank Policy Research Working Paper Series No.5385.*

Duflo, E. and R. Pande. 2007. "Dams." *Quarterly Journal of Economics*, 122(2): 601–646.

Fan, S. and C. Chan-Kang. 2005. "Road Development, Economic Growth, and Poverty Reduction in China." *IFPRI Research Report 138.* International Food Policy Research Institute, Washington D.C.

Fan, S. and P. Hazell. 2001. "Returns to Public Investments in the Less-favored Areas of India and China." *American Journal of Agricultural Economics*, 83(5): 1217–1222.

Gibson, J. and S. Rozelle. 2003. "Poverty and Access to Roads in Papua New Guinea." *Economic Development and Cultural Change*, 52(1): 159-85.

Gollin, D. and R. Rogerson. 2010. "Agriculture, Roads, and Economic Development in Uganda." *NBER Working paper* No.15863.

Limao, N. and AJ. Venables. 2001. "Infrastructure, Geographical Disadvantage, Transport Costs and Trade." *World Bank Economic Review,* 15(3): 451-479.

Omamo, S. W. 1998. “Farm-to-Market Transaction Costs and Specialization in Small-Scale Agriculture: Explorations with a Non-Separable Household Model.” *Journal of Development Studies,* 35(2):152-63.

Renkow, M., D.G. Hallstrom, and D. D. Karanja. 2004. “Rural Infrastructure, Transactions Costs and Market Participation in Kenya.” *Journal of Development Economics*, 73(1): 349– 367.

Stifel, D., B. Minten, and P. Dorosh. 2003. “Transaction Costs and Agricultural Productivity: Implications of Isolation for Rural Poverty in Madagascar.” Washington, D.C.: International Food Policy Research Institute, MSSD Discussion Paper No. 56.

Yang, X. and Y-k. Ng. 1993. “Specialization and Economic Organization, a New Classical Micro economic Framework.”Amsterdam, North-Holland.

Zhang, X., J. Yang, and S. Wang. 2011. “China Has Reached the Lewis Turning Point.” *China Economic Review,* forthcoming.

Table 1. Road Access in Three Surveyed Villages

	2004	2006	2009
Changtian Village			
Yezhuwo	0	0	0
Luojiawan	1	1	1
Yeniuzhuang	0	0	0
Dongqingzhai	0	1	1
Yanshang	0	0	1
Baobaozhai	0	0	1
Changtian	1	1	1
Pianpo	0	0	1
Naba	1	1	1
Xiadaba Village			
Naoyutang	0	0	0
Baobaozhai	1	1	1
Tianlinggang	0	0	0
Xiadaba	0	1	1
Xinlin Village			
Xiaojialin	1	1	1
Wangchengpo	1	1	1
Yejiaba	1	1	1
Liangshuijing	1	1	1

Source: Authors' Survey (2005, 2007, 2010).
Note: "1" denotes there is road access in the village in the specific year; "0" otherwise.

Table 2. Household Agricultural Income Shares in Different Categories

First Wave: Year 2004 (N=758)			
Income sources (%)	Mean	Std	Share>0
Corn	46.24	22.03	98.28
Paddy	11.87	17.68	46.04
Rapeseed	20.89	15.41	88.26
Vegetable	3.36	9.08	33.91
Fruit	2.20	7.02	20.71
Poultry	2.43	8.98	25.99
Livestock	11.93	21.21	29.95
Forestry	0.23	3.81	1.45
Fishing	0.84	6.46	2.11
Second Wave: Year 2006 (N=765)			
Income sources (%)	Share	Std	Share>0
Corn	38.70	21.48	94.90
Paddy	11.60	16.85	47.32
Rapeseed	19.28	14.34	92.81
Other grain	1.98	4.51	33.33
Vegetable	11.43	15.02	85.49
Fruit	3.38	10.72	23.27
Poultry	1.54	6.37	20.92
Livestock	10.14	19.06	27.32
Forestry	0.33	2.71	2.88
Fishing	1.62	10.43	3.14
Third Wave: Year 2009 (N=759)			
Income sources (%)	Share	Std	Share>0
Corn	42.04	24.34	94.99
Paddy	10.00	17.76	43.87
Rapeseed	16.00	12.40	93.15
Bean	4.00	7.62	82.35
Fruit	4.00	13.23	29.78
Poultry	2.00	10.60	37.94
Livestock	18.00	27.91	30.30
Forestry	0.00	5.00	5.27
Fishing	1.00	9.68	3.03

Source : Authors' survey (2005, 2007, 2010).

Table 3. Summary of Statistics

First Wave: Year 2004 (N=782)					
Variables	Obs	Mean	Std	Min	Max
Household income	782	6644.11	5156.50	0	50000
Agricultural income	782	3978.44	3861.81	0	37165
Non-agricultural income	782	2665.66	3472.30	0	50000
Household HH Index of agricultural production	758	0.46	0.16	0.19	1
Household size (migrants excluded)	782	3.69	1.55	0	8
Household land cultivated (*mu*)	781	3.66	2.76	0	20
Number of Labor (age 16-60)	782	2.53	1.42	0	7
Highest education in the household (year)	780	5.43	3.31	0	14
Village cadre in the household (dummy)	782	0.04	0.20	0	1
Second Wave: Year 2006 (N=815)					
Variables	Obs	Mean	Std	Min	Max
Household income	815	7618.57	10413.26	0	223080
Agricultural income	815	3825.31	3822.14	0	33148
Non-agricultural income	815	3793.26	9666.20	0	223000
Household HH Index of agricultural production	765	0.41	0.16	0.18	1
Household size (migrants excluded)	815	3.35	1.66	0	10
Household land cultivated (*mu*)	810	3.90	2.99	0	20.5
Number of Labor (age 16-60)	815	2.52	1.51	0	9
Highest education in the household (year)	811	6.12	3.47	0	18
Village cadre in the household (dummy)	811	0.02	0.13	0	1
Third Wave: Year 2009 (N=834)					
Variables	Obs	Mean	Std	Min	Max
Household income	834	11994.57	13933.61	0	191265
Agricultural income	834	5453.96	6613.53	0	67155
Non-agricultural income	834	6540.62	11868.99	0	182620
Household HH Index of agricultural production	759	0.49	0.18	0.22	1
Household size (migrants excluded)	834	3.85	1.83	0	12
Household land cultivated (acre)	806	3.10	2.92	0	32.5
Number of Labor (age 16-60)	834	2.47	1.48	0	7
Highest education in the household (year)	833	6.17	3.59	0	18
Village cadre in the household (dummy)	834	0.04	0.19	0	1

Source: Authors' survey (2005, 2007, 2010).

Table 4. Summary of Statistics by Road Access

Without Road Access (N=557)					
Variables	Obs	Mean	Std	Min	Max
Household income	557	5421.16	4388.54	0	38814.04
Agricultural income	557	3207.18	3216.94	0	38011.78
Non-agricultural income	557	2213.98	2590.37	0	20963.11
Household HH Index of agricultural production	535	0.44	0.16	0.18	1
Household size (migrants excluded)	557	3.79	1.81	0	12
Household land cultivated (*mu*)	551	3.23	2.75	0	20.5
Number of Labor (age 16-60)	557	2.58	1.50	0	9
Highest education in the household (year)	556	5.24	3.40	0	16
Village cadre in the household (dummy)	555	0.05	0.21	0	1
With Road Access (N=1874)					
Variables	Obs	Mean	Std	Min	Max
Household income	1874	9828.29	12066.59	0	221090.2
Agricultural income	1874	4791.11	5455.42	0	68806.35
Non-agricultural income	1874	5037.18	10510.64	0	221010.9
Household HH Index of agricultural production	1747	0.46	0.17	0.18	1
Household size (migrants excluded)	1874	3.59	1.66	0	12
Household land cultivated (*mu*)	1846	3.65	2.95	0	32.5
Number of Labor (age 16-60)	1874	2.48	1.46	0	8
Highest education in the household (year)	1868	6.12	3.47	0	18
Village cadre in the household (dummy)	1872	0.03	0.16	0	1

Source : Authors' survey (2005, 2007, 2010). All the prices are deflated to year 2003.

Table 5. Impact of Road Access on Agricultural Specialization

	Dependent Variable: HH index								
	HHI [1]			HHI [2]			HHI [3]		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Road	0.05***	0.06***	0.06***	0.05***	0.06***	0.06***	0.04***	0.05***	0.06***
	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)
Road built within last two years		-0.02	-0.02		-0.01	-0.02		-0.01	-0.02
		(0.02)	(0.02)		(0.02)	(0.02)		(0.02)	(0.02)
Long term impact = 'Road'; Short term impact = 'Road' + 'Road built within last two years'									
Land			-0.00***			-0.00**			-0.00***
			(0.00)			(0.00)			(0.00)
Number of Labor (age 16-60)			0.00			0.00			0.00
			(0.00)			(0.00)			(0.00)
Household size			-0.02***			-0.01***			-0.01***
			(0.00)			(0.00)			(0.00)
Highest education (year)			0.00			0.00			0.00
			(0.00)			(0.00)			(0.00)
Village cadre (dummy)			-0.04*			-0.03			-0.03
			(0.02)			(0.02)			(0.02)
Year fixed effect	YES	YES	YES	YES	YES	YES	YES	YES	YES
Natural village fixed effect	YES	YES	YES	YES	YES	YES	YES	YES	YES
R-squared	0.08	0.08	0.11	0.07	0.07	0.09	0.06	0.06	0.09
AIC	-1802.5	-1800.9	-1858	-1853.4	-1851.7	-1898	-1868.6	-1866.9	-1915.2
N	2124	2124	2124	2124	2124	2124	2124	2124	2124

Notes : 1.Households with no land are dropped from the regression for sample consistency for all regressions. In addition, seven observations with self consumption ratios larger than one are dropped, since they are possibly generated from recording errors. The main findings remain the same after including those dropped observations; 2. *significant at 0.10 level; ** significant at 0.05 level; *** significant at 0.01 level. 3. Robust standard errors are clustered at natural village * year level.

Table 6. Impact of Roads on Input Use

	Dependent Variable: Input Use Input Use (*yuan* per *mu* in log)		
	(1)	(2)	(3)
Road	0.18	-0.16	-0.19
	(0.13)	(0.20)	(0.20)
Road built within last two years		0.41**	0.43**
		(0.17)	(0.18)
Long term impact = 'Road'; Short term impact = 'Road' + 'Road built within last two years'			
Land			-0.05***
			(0.01)
Number of Labor (age 16-60)			0.01
			(0.01)
Household size			0.01
			(0.01)
Highest education (year)			0.01*
			(0.01)
Village cadre (dummy)			0.04
			(0.14)
Year fixed effect	YES	YES	YES
Natural village fixed effect	YES	YES	YES
R-squared	0.02	0.03	0.04
AIC	5965.8	5959.9	5921.3
N	2124	2124	2124

Notes : 1.Households with no land are dropped from the regression for sample consistency for all regressions. In addition, seven observations with self consumption ratios larger than one are dropped, since they are possibly generated from recording errors. Main findings remain the same after including those dropped observations; 2. *significant at 0.10 level; ** significant at 0.05 level; *** significant at 0.01 level. 3. Robust standard errors are clustered at natural village * year level.

Table 7. Impact of Roads on Agricultural Product Trading

	Dependent Variable: Agricultural income in cash Cash income (log)		
	(1)	(2)	(3)
Road	0.17	-0.08	-0.08
	(0.15)	(0.16)	(0.15)
Road built within last two years		0.30**	0.38***
		(0.14)	(0.13)
Long term impact = 'Road'; Short term impact = 'Road' + 'Road built within last two years'			
Land			0.07***
			(0.01)
Number of Labor (age 16-60)			0.03
			(0.02)
Household size			0.13***
			(0.02)
Highest education (year)			0.00
			(0.01)
Village cadre (dummy)			0.29
			(0.19)
Year fixed effect	YES	YES	YES
Natural village fixed effect	YES	YES	YES
R-squared	0.09	0.09	0.13
AIC	7613.6	7613.6	7517.7
N	2124	2124	2124

Notes : 1.Households with no land are dropped from the regression for sample consistency for all regressions. In addition, seven observations with self consumption ratios larger than one are dropped, since they are possibly generated from recording errors. Main findings remain the same after including those dropped observations; 2. *significant at 0.10 level; ** significant at 0.05 level; *** significant at 0.01 level. 3. Robust standard errors are clustered at natural village * year level.

Table 8. Impact of Roads on Household Income Composition

	Dependent Variable: Agricultural Income, Non-agricultural Income and Total Income								
	Agricultural income (log)			Non-agricultural income (log)			Total income per capita (log)		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Road	0.13	-0.02	-0.02	-0.09	0.32	0.25	0.07	-0.06	-0.07
	(0.10)	(0.13)	(0.14)	(0.32)	(0.33)	(0.31)	(0.08)	(0.10)	(0.09)
Road built within last two years		0.18*	0.26***		-0.49	-0.3		0.16**	0.14**
		(0.10)	(0.10)		(0.32)	(0.30)		(0.08)	(0.07)
Long term impact = 'Road'; Short term impact = 'Road' + 'Road built within last two years'									
Land			0.06***			0.01			0.05***
			(0.01)			(0.03)			(0.01)
Number of Labor (age 16-60)			0.04**			0.07			0.04***
			(0.02)			(0.05)			(0.01)
Household size			0.13***			0.24***			-0.14***
			(0.01)			(0.05)			(0.01)
Highest education (year)			0.00			0.04*			0.02**
			(0.01)			(0.02)			(0.01)
Village cadre (dummy)			0.25*			0.90**			0.38***
			(0.13)			(0.35)			(0.11)
Year fixed effect	YES	YES	YES	YES	YES	YES	YES	YES	YES
Natural village fixed effect	YES	YES	YES	YES	YES	YES	YES	YES	YES
R-squared	0.08	0.08	0.18	0.05	0.05	0.07	0.12	0.12	0.22
AIC	5692.5	5692.9	5452.8	10581.6	10582.3	10532.8	5136.3	5136.5	4895.7
N	2124	2124	2124	2124	2124	2124	2124	2124	2124

Notes : 1.Households with no land are dropped from the regression for sample consistency for all regressions. In addition, seven observations with self consumption ratios larger than one are dropped, since they are possibly generated from recording errors. Main findings remain the same after including those dropped observations; 2. *significant at 0.10 level; ** significant at 0.05 level; *** significant at 0.01 level. 3. Robust standard errors are clustered at natural village * year level.

Table 9. Robustness check using standard DID specification

	HHI [1]	HHI [2]	HHI [3]
Road*beforeafter	0.04***	0.04***	0.04***
	(0.01)	(0.01)	(0.01)
	Agricultural income	Non-ag income	Total income per capita
Road*beforeafter	0.19*	-0.01	0.05
	(0.10)	(0.30)	(0.07)
	Input use	Agricultural cash income	
Road*beforeafter	0.17	0.24	
	(0.13)	(0.14)	

Notes : 1.Households with no land are dropped from the regression for sample consistency for all regressions. In addition, seven observations with self consumption ratios larger than one are dropped, since they are possibly generated from recording errors. Main findings remain the same after including those dropped observations; 2. *significant at 0.10 level; ** significant at 0.05 level; *** significant at 0.01 level. 3. Robust standard errors are clustered at natural village * year level.