Education spillovers in farm productivity: empirical evidence in rural India

Véronique Gille*

Centre d'Économie de la Sorbonne, University of Paris 1 Panthéon Sorbonne

September 2010

Abstract

Empirical evidence of education spillovers in developing countries and rural contexts is scarce and focuses on specific channels. This paper provides evidence of such spillovers in rural India, by evaluating the overall impact of education of neighbors on farm productivity. Spatial econometric tools are used to take into account social distance between neighbors. We use data from the India Human Development Survey (IHDS) of 2005. Our main results show that education spillovers are substantial: one additional year in the mean level of education of neighbors increases households' farm productivity by 3%.These findings are robust to changes in specification. They show the importance for policy makers of taking into account education spillovers and policies' complementarity when facing political trade-offs.

Keywords: Education externalities, Rural India, Farm productivity

JEL Classification Numbers: D1, I2, O1, R3

*Address: 106-112 bvd de l'Hôpital, Maison des Sciences Économiques, 75647 Paris cedex 13 FRANCE, telephone: +33631210602, e-mail:*veroniquegille@gmail.com*

1 Introduction

In the last twenty years, India has massively invested in education. Its efforts and accomplishments in terms of education have hugely increased with the launching of the Sarva Shiksha Abhiyan program in 2002, which aimed at providing primary education to all 6-14 years old children by 2010. Even if this goal has not yet been reached, the number of children out of school has been reduced from 25 millions in 2003 to 8.1 million in 2009[1].

This process is considered of major importance for India because as mentioned by the World Bank "Education is one of the most powerful instruments for reducing poverty and inequality. Education is equally key to enhance India's competitiveness in the global economy. Therefore, ensuring access to quality education for all, in particular for the poor and rural population, is central to the economic and social development of India.".

The impact of education on productivity, growth and more generally on development in India has been widely asserted by researchers [Psacharopoulos and Patrinos, 2002]. Nevertheless, education does not induce growth only because it improves individual productivity. Education is a key issue for development also because it has positive externalities, notably in terms of learning spillovers. Increasing education has a higher social return than its private return.

Even if education externalities are not a new idea in the literature [Marshall, 1890, Lucas, 1988], their empirical assessment and measurement is relatively recent [Moretti, 2003, 2004, Mas and Moretti, 2006, e.g.,]. One reason is the identification problem: human capital is not distributed randomly across places. Consequently, research on this subject is scarce.

The existing literature mainly focuses on education externalities in cities or in firms in developed countries. In developing countries these works have very little implications. Though, having a clear idea of education spillovers in a rural context should help resolving trade-off in development policies.

[1]Source: World Bank

In the case of India, the question of education spillovers in villages is of particular concern, in part because 72% of the population is rural[2], and in part because the agricultural policy is at a turning point. India's past agricultural policies have been largely based on input subsidies and on price control. After the important increase in its agricultural production occurred in the 1970's during the Green Revolution, its growth has progressively slowed down to reach 2.5% per year in the last ten years. Moreover, this policy is very costely and has important effects on environment, in particular because of subsidized fertilizers. Criticized by researchers [Gulati and Sharma, 1995, e.g.] and international institutions, India has redefined its agricultural policy for the period 2007/08 - 2011/12 toward a better shared growth. Being relatively free from international pressure due to little constraining agreements with WTO[OCDE, 2009], India has the opportunity to use differently the money currently spent in subsidies. Determining the true impact of diverse policies, and in our case of education policies, is consequently important for policy makers.

Our goal in this paper is to assert the existence of education externalities in rural India, by estimating the overall impact of neighbors' education level on household farm productivity. We use spatial econometrics tools to evaluate the spillover effect while taking into account social interactions in Indian villages. To my knowledge, this paper is one of the first to consider this issue in a rural context in a developing country[3].

Cross-sectional data from the India Human development survey are used. This survey was conducted in 2005 in 41,554 households in 1503 villages and 971 urban neighborhoods. All States and Union Territories of India were covered with the exception of the islands of Andaman and Nicobar and Lakshadweep.

Our paper is organized as follows. Section 2 is a brief literature review on education externalities. Section 3 describes the empirical specification, the econometric issues and the

[2]Source: Census of India, 2001

[3]One exception is the article by Conley et al. [2003] which considers education spillovers in Malaysia.

data used. Section 4 presents the results. Section 5 tests the robustness of the results and section 6 concludes.

2 Literature Review

Education externalities have been mainly invoked for urban contexts in developed countries (2.1). But learning spillovers do exist in agriculture (2.2).

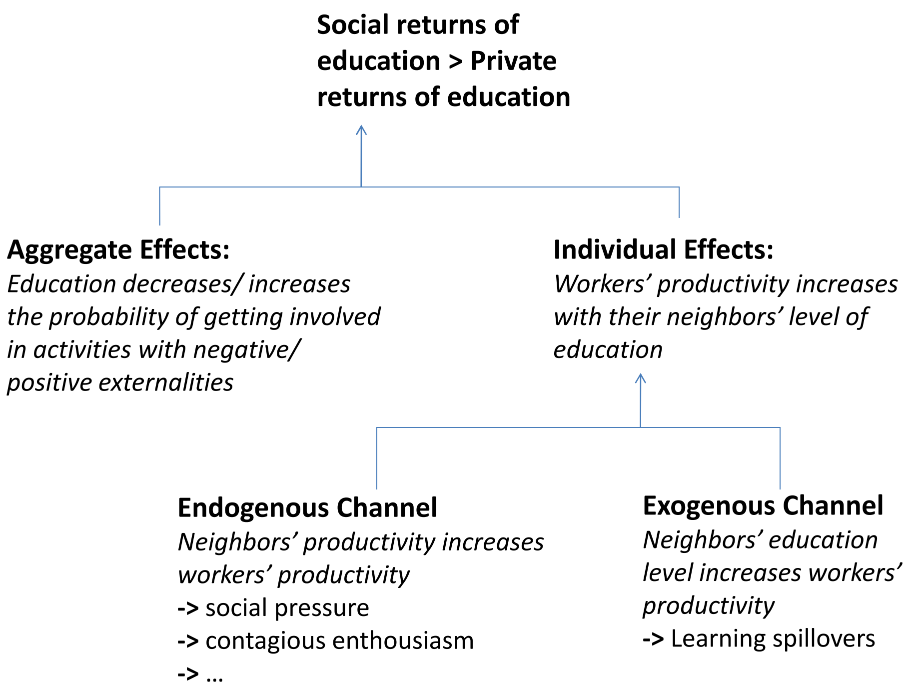
2.1 Education externalities

Theoretically, there are several ways which make human capital's social return higher than its private return. The different channels throw which this phenomenon occurs can be classified into two types. Here on we will use indifferently the terms human capital and education.

The first channel is a broad one: education has a higher social return at a community level such as a city or a State because education reduces the probability of getting involved in activities which produce negative externalities. It also increases the probability of engaging in activities with positive externalities. Moretti [2003] for example underlines that education can "reduce criminal participation and improve voters political behavior". These changes of behavior thanks to education will either reduce the negative externalities generated by criminal activities, either lead to better policy decisions which in turn affect positively the community.

The second channel is what is called a spillover effect. It occurs at the individual level. As Kremer [1993] assumes in its O-ring theory, human capital of a worker may have a marginal return which grows with the human capital of other workers. This effect can occur because the productivity of each people can affect directly the productivity of others. This has been called by Manski [1993] the "endogenous channel". Mas and Moretti [2006] underline two specific reasons why it could happen: there can be a social pressure between workers who suffer from a disutility if other workers notice that there are less productive. There can also be contagious enthusiasm: a more productive worker has a leading effect on his coworkers. The spillover effect can also happen because there is an "exogenous channel" [Manski, 1993]: workers' productivity increases with their neighbors' education. This can take place through transfers of knowledge from skilled workers to unskilled workers as Martins [2004] underlines or from a worker to another worker, through exchange of ideas, imitation or learning [Acemoglu and Angrist, 1999]. Figure 1 summarizes these different channels.

Figure 1: Education Externalities channels



The existence of educational spillover has been empirically tested at different level of aggregation, following the theoretical literature.

The first empirical strategy is to evaluate the impact of human capital at a city or regional level. This empirical work captures then aggregate as well as individual effects of education. The methodology is to compare individual wages or plants' output in cities or regions with different average level of education. Acemoglu and Angrist [1999], for example, use an instrumental variable strategy to evaluate the impact of higher average schooling in US States on individual wages. They find mitigated results. Using a different methodology, Moretti [2004], finds that in the US, plants in cities where the share of college graduates grows faster, have a faster increase in productivity than plants in cities where the share of college graduates grows slower.

Nevertheless, most interactions between workers take place in a more reduced environment than cities or regions [Martins, 2004]. That's why the second empirical strategy is to consider inside plants distribution of human capital and its impact on productivity. Consequently, it excludes the broader channels to focus on the individual channel. Mas and Moretti [2006] consider the impact of high productivity workers on other workers' productivity of a grocery chain in the United States. They find that there is a productivity spillover going from highly productive workers to low productivity workers. Further exploring the channels through which this spillover occur, they find that it is due to mutual monitoring. Similarly, Martins [2004] tests the spillover between educated and less educated workers on a set of Portuguese firms. He finds that the social return at a firm-level is higher than the private return, thanks to a transfer of skills from more educated workers to less educated workers. Navon [2009], using data from Israeli manufacturing plants, tests the existence of spillovers between workers with different type of knowledge. He finds that employing diverse workers according to the academic discipline where they get their diploma increases firms' productivity.

If this empirical work mostly agrees on the existence of human capital spillovers, it only focuses on externalities in firms in an urban context. Does this work could be extended to other environments? Does Kremer's O-ring theory also apply to rural contexts for example?

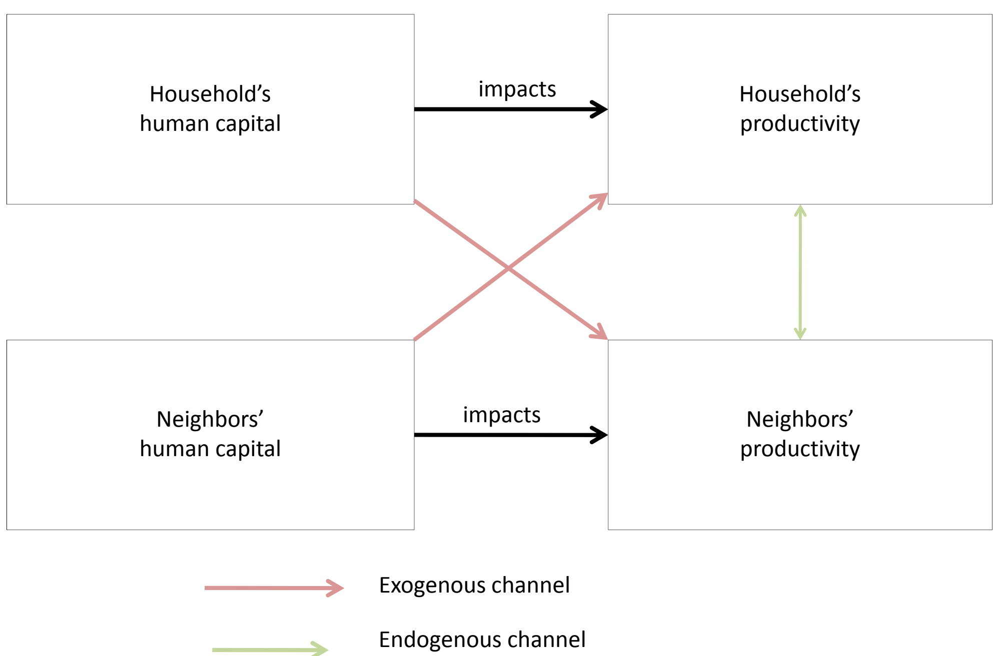
2.2 Education externalities in a rural context

The impact of education on agricultural productivity has been documented in the literature. But the issue of education spillovers in this context is still to explore. Only specific channels through which spillovers occur have been tested.

The channels through which education has an impact on agricultural productivity are well-known. Rosenzweig (1995) argue that education improves information learning. On one hand it gives better access to sources of information, such as newspapers or manuals. On the other hand education helps understanding new information. Consequently education has positive returns where there are learning opportunities. In particular, education facilitates the adoption of new technologies. Foster and Rosenzweig (1996), using datasets from the Green Revolution in India, show that the returns from schooling in rural areas increase in periods of fast technological change. Another channel we can think of is a more indirect one: by improving health, education can improve productivity.

The channels through which neighbors' level of education could improve one's farm productivity are the same as the channels described in figure 1. There can be as well endogenous channels as well as exogenous channels. That is to say the level of education of the neighborhood can have an indirect effect through its impact on neighborhood productivity. Or it can have a direct effect through learning spillovers. Figure 2 makes these channels more explicit.

Figure 2: Endogenous and Exogenous channels



Only one specific channel for education spillovers in a rural context has been massively documented: the adoption of new technologies in agriculture [Foster and Rosenzweig, 1995, Besley and Case, 1994, Munshi, 2004, e.g]. This process is massively influenced by neighbor's behavior which depends in part on their education level and on learning spillovers.

Studying specific channels is essential, but is less relevant for education policies. For this reason, the goal in this paper is to assert and evaluate the overall impact of neighbors' human capital on households' farm productivity.

3 Empirical methodology and issues

3.1 Empirical specification

The influence of neighbors' human capital on farm productivity is estimated in a standard linear framework, using spatial econometrics terminology. The equation estimated is

$$lnY_i = \alpha + \beta W_i E + \gamma X_i + u_i \tag{1}$$

Where Y_i is the farm productivity of the household i, X_i is a set of control household variables which influence the productivity and u_i is a household error term which control for other determinants of the household productivity.

The level of education of neighbors is taken into account by using spatial econometric tools: W_i is the ith row of a matrix W which allocates to each household its neighbors. More precisely, each element w_{ij} of W is defined as follows: $w_{ij} = \theta$ with $\theta > 0$ if i and j are neighbors and $w_{ij} = 0$ otherwise. Whether i and j are neighbors depends on the definition of the neighborhood we use. $w_{i,i} = 0$ for all i to exclude the household level of education from the calculation of neighbors' educational level. E is a column vector whose elements represents the level of education of each household in the database. As each row W_i is also normalized such that $W_i = 1$ for all i, $W_i E$ is a weighted average of neighbor's education level. The different matrix used are explained below.

The control variables are divided into three categories: there is one Human Capital variable, one Physical Capital variable and three Input variables.

One major issue when estimating this equation is the endogeneity problem. Unobserved variables which impact farm productivity can be correlated to neighbors' education.

The first kind of omitted variables we can think of are the geological variables such as the quality of land (dryness, fertility, etc) or the climate variables. It includes permanent and temporary characteristics. These unobserved variables may be correlated to education of neighbors through a fiscal channel: richer places (because of good land fertility for example) have more public money to invest in education thanks to higher tax receipts, improving the level of education of the whole village. These unobserved variables generate an upward bias in the estimation of the impact of the neighbor's education.

The second kind of omitted variables is also problematic. It concerns infrastructure variables such as roads or public transports. This kind of infrastructure increases educational level through the offer side because people have better access to schools, or through the demand side: roads create job opportunities which increase incentives of getting educated. At the same time it may also increase farm productivity, through better access to new technologies or to input markets for example. These unobserved variables (at least in our database) generate again a positive bias: it increases the coefficient of neighbors' education.

One way to control for these biases would have been by adding village dummies: it would have absorbed variables which affect at the same time household productivity and neighbors' behavior. Nevertheless this econometric solution is not possible for two reasons. The first reason is that in our sample there are not enough households per village. Even if the survey has been conducted on at least thirty households per village, not every habitants of each village cultivate land. So in some villages there are less than ten households in the sample. The second reason is that the education of neighbors' variable has by construction - which will be explicated latter- a small within-village variance. So its coefficient cannot be efficiently estimated with village dummies [Plumper and Troeger, 2007]. Furthermore an estimation with village dummies do not solve the problem of temporary unobserved characteristics.

In this context, the solutions chosen are as follows: to control for the unobserved "geological-kind" variables, we add regional fixed-effects. The region is a statistical entity which has been constructed by the National Sample Survey Organization (NSSO)[4] to conduct its surveys. Regions are composed of several districts of the same State and they are homogeneous in their "agro-climatic conditions and socio-economic features" (Murthi, Srinivasan and Subramanian 1999). They must then well take into account the first kind of omitted variable. We also estimate farm productivity in prices calculated for each crop at the district level. As prices vary temporally depending on weather conditions, and geographically depending on climate conditions, it also partially integrate as well temporary as constant unobserved variables[5].

The second kind of unobservable variables are taken into account by adding village-level variables which have an impact on farm productivity. The first village characteristic which we are interested in is the isolation of the village. As explained earlier, people from an isolated village have less incentives to get educated and at the same time they don't have access to new technologies. The distance of villages to the nearest city is used to measure the isolation of each village. We also add a variable which proxies credit access in the village: it is a dummy variable which is equal to 1 if there is a bank branch office or a credit cooperative in the village. Finally, we add the proportion of household having electricity in the villages which also captures the isolation of the village.

Consequently, the extended model is:

[4]The NSSO is an organization which depends of the Ministry of Statistics and Program implementation. It conducts various socio economic surveys.

[5]Specifically, concerning the education of neighbors variable, as it takes into account only adults (which means more than 15 years old people), there is a lag between the time at which the level of education is determined and the year for which the level of productivity is calculated. It is consequently quite unlikely that a temporary shock during the year of the survey could have an effect on the level of education of the concerned people which was determined years before.

$$lnY_{ivr} = \alpha + \beta W_i E + \gamma X_i + \delta Z_v + c_r + u_i \tag{2}$$

Where Z_v is the set of village-level variables and c_r is the regional fixed effect.

Under the hypothesis that migration has not as a consequence to sort people according to factors influencing their farm productivity, this model can be consistently estimated with OLS estimates [GALLO, 2000]. In India, there is a really low rate of migration. According to the Indian Census of 2001, 72.4% of the rural population was born in their place of residence. The migrants are mainly women for reason of marriage. If we only consider men, 89.1% where born in there current place fo residence. For this reason, many authors choose to ignore the migration problem in rural India [Banerjee et al., 2007, Foster and Rosenzweig, 1995, Anderson, 2005]. We follow this work in our study. This hypothesis is discussed in part 5.

3.2 Weighting matrix construction

One issue in estimating neighborhood impact is the neighborhood definition. Goux and Maurin [2007]underline that "distant neighbors have less influence than close ones" and that using a too broad definition of neighbors can lead to an underestimate of the influence of close neighbors. To control for this bias we test several definitions of neighborhood. These definitions are based on geographic contiguity, caste group membership and occupation.

The first matrix we use to define the neighborhood is a simple contiguity matrix: according to the literature on neighborhood effect in agriculture in India [Foster and Rosenzweig, 1995, Munshi, 2004], the social unit where interactions occur is the village. So we define neighbors as people from the same village. Furthermore we assume that each neighbor has the same influence on farm productivity. In other words, in a first time, $w_{i,j} = 1$ if i and j are from the same village, $w_{i,j} = 0$ otherwise. The consequence of this construction is that each neighbor's observation has equal weights: $w_{i,j} = w_{i,k}$.

The other definitions of neighborhood we use are still based on village membership, but each neighbor takes different weight depending on its social or economic distance with the household. The idea is that people can be influenced differently by their neighbors, either because they are not from the same social group and consequently they communicate less, either because they don't do the same work, so they can't share their experience.

In a first time, we allow people to have different weights on account of their respective social groups. Due to a lack of details in our data, we cannot define very precisely social groups. Nevertheless, the specific social organization of India in castes helps us understanding social interactions occurring in villages.

Srinivas [1962][6] defines a caste as an "hereditary, endogamous group which is usually localised. It has a traditional association with an occupation, and a particular position in the local hierarchy of castes. Relations between castes are governed, among other things by the concepts of pollution and purity, and generally maximum commensality i.e. interdining occurs within the caste". Castes are traditionally specialized in a specific occupation which make the different groups interdependent from each other. So they interact between each others. Nevertheless, castes at the bottom of the traditional hierarchy are considered as impure and are ostracized by people from other castes in the village. Consequently they are less likely to interact with people from other castes[7].

These groups have been targeted by the government for specific policies, so surveys refer to them as Scheduled Tribes(ST) or Scheduled Castes(SC). Our data specifies for each household if the head of the household is or not a ST or a SC. So we are able to classify households in three social groups: ST, SC and others.

The main problem is to determine which weight we use for people who are from the same

[6]Indian sociologist

[7]for more information on castes, please refer to Delige [2004].

social group and for people who are from a different social group. To avoid arbitrary decision we first evaluate the real impact of neighbor's education by running the same regression, but the influence of people from the same social group are considered separately from people of other social groups, that is

$$lnY_{ivr} = \alpha + \beta_1 W_{1i}E + \beta_2 W_{2i}E + \gamma X_i + \delta Z_v + c_r + u_i \tag{3}$$

Where W_{1i} is the ith row of the matrix W_1 whose elements w_{1ij} are equal to one if i and j are from the same village and from the same social group, 0 otherwise and W_{2i} is the ith row of the matrix W_2 with $w_{2ij} = 1$ if i and j are from the same village but not from the same social group, 0 otherwise. The rows of W_1 and W_2 are normalized. The set of control variables is the same.

The coefficients β_1 and β_2 give the respective impact of people from the same social group and from another social group. They are used to compute the weight matrix. In equation (1), $w_{i,j}$ is still equal to 1 if i and j are from the same village and from the same social group, but if i and j are from the same village but not from the same social group, $w_{i,j} = \beta_2/\beta_1$. People who are not from the same village get a weight of 0. Again $w_{i,i} = 0$ for all i and the rows of the matrix are normalized.

Finally, the third matrix we use is a matrix based on neighbors' occupation. Again neighbors are people from the same village, but they get different weight depending on their occupation. The same methodology is used to calculate the weights: it is most likely that knowledge spillovers are more important between neighbors who have the same job. The hypothesis here is then that the level of education of people whose main activity is farming will have more influence on farm productivity than people who have different works.

The equation estimated is the same as previously, but $w_{1,i,j} = 1$ if i and j are from the same village and the main occupation of j is agriculture, 0 otherwise and $w_{2,i,j} = 1$ if i and j are from the same village but j's main activity is not agriculture, 0 otherwise.

3.3 Data

3.3.1 India Human Development Survey

The database used to evaluate human capital spillovers comes from the India Human Development Survey. This survey was jointly conducted in 2005 by researchers from the University of Maryland and the National Council of Applied Economic Research, New Delhi. It took place in all States and Union Territories (UT) of India, with the exception of the Islands of Andaman and Nicobar and Lakshadweep. These places have not been surveyed because of their small population. Across these 33 States and UT, 41,554 households in 1503 villages and 971 urban neighborhoods were interviewed.

The goal is to evaluate the impact of neighbors' human capital in agriculture, so we only keep rural households. Since the survey's drafting, 19 villages of the sample have been classified as urban zones by the 2001 National Census of India. We take the census definition to define rural households. This leads to 26,734 households.

The education of neighbors variable is calculated with the whole set of rural household. That is to say we don't only consider the impact of neighbors who also cultivate land but the impact of all neighbors in the village. Out of these 26,734 rural households, 43 have missing values for the educational level. So the education of neighbors is calculated with 26,691 households.

For the equation estimation, we only consider rural households who cultivate land that is to say 14,298 households. Out of these households, 3236 have missing values for one of the variable[8]. Consequently the estimation is made on 11,062 households.

3.3.2 Variables building

The data indicate for each household the crops grown and the total production and amount cultivated for each crop. 68 types of crops are referenced. People were also asked if they sold

[8] In particular, due to logistical constraints the interviewers were only able to complete 1454 village questionnaires, resulting in 49 villages being omitted. According to the survey managers, there were no consistent pattern to these omissions.

any of their production and how much they sold it. The total production for each household is calculated in rupees by using the mean price of each crop in the district[9]. The productivity is the total production in rupees divided by the total amount of land cultivated.

The education variable is based on years completed at school. It varies from 0 to 15 which is equivalent to a bachelor degree. The household level of education is the mean level of the most educated female adult (more than 21 years old) and the most educated male adult. The level of education of neighbors is calculated using the same estimation for each household and different types of matrix of contiguity.

For the control variables, the Human capital variable is the number of hours worked per year per household on the field in logs. The Physical Capital variable is a dummy variable which is equal to one if the household owns agricultural equipment, 0 otherwise. Agricultural equipments includes tractors, threshers and bio-gas plants. Input variables include the value of seeds used per acre cultivated, if the land is irrigated or not and the proportion of fertilizer used per acre. Finally, the village-level variables are the distance to a town and the percentage of households having electricity in the village which proxy the isolation of the village and the credit access which is a dummy variable equal to 1 if the village has bank offices or credit cooperatives and 0 otherwise.

4 Results

4.1 Weights estimation

The first step is the estimation of the weights for the weighting matrix. The results are given in table 2.

Column (1) differentiates between the education of people from the same social group and the education of people from another social group. The results confirm the choice of

[9]for the mean price calculation, answers which are outside the price boxplot are considered as outliers and eliminated (i.e. prices which are not in the interval $[\bar{x} - (Q_3 - Q_1); \bar{x} + (Q_3 - Q_1)]$, where Q_1 and Q_3 are respectively the first and third quartile.

Table 1: DESCRIPTIVE STATISTICS

.

	mean	sd	min	max
Productivity	868.761	3930.93	1.027875	181666.7
HH Education	5.070286	4.03148	0	15
Neighbors' Education	4.447781	2.059923	0	15
Days worked	353.1867	280.2019	2	3480
Agr. equipment	.0942867	.2922404	0	1
Irrig. land	.6110107	.4875429	0	1
Fertilizer p.a.	697.781	1679.833	0	75000
Nb of seasons	1.714337	.6007037	1	3
Distance to town	14.03526	10.8453	1	85
% of HH with elect	63.6987	34.76666	0	100
Credit access	.4706201	.4991586	0	1
Observations	11062			

caste as an indicator of social interactions: farm productivity is much more influenced by the educational level of neighbors from the same caste, than from the education of people from another caste. The spillover effect of people from the same social group is of 3% whereas the one from people in the same village but not from the same social group is of 0.5% and is not statistically significant.

Column (2) shows the results when we differentiate people on account of their occupation. Not surprisingly, the level of education of neighbors whose main occupation is agriculture has more impact on farm productivity than the level of education of people whose main occupation is not agriculture. The spillover effect of neighbors who have the same occupation is around 1.7% and it is significant at a 0.1% level whereas the spillover effect from other neighbors is around 0.8% and is only significant at a 5% level.

Table 2: WEIGHTS CALCULATION

	(1) Productivity	(2) Productivity
HH Education	0.0119***	0.0112***
	(0.00225)	(0.00212)
Same Caste group Neigh's educ level	0.0298***	
	(0.00405)	
Diff Caste group Neigh's educ level	0.00536	
	(0.00382)	
Same Occupation Neigh's Educ level		0.0170***
		(0.00422)
Diff Occupation Neigh's Educ level		0.00756*
		(0.00349)
Control Variables	YES	YES
Regional Fixed-Effects	YES	YES
Observations	9215	10348
r2	0.387	0.398

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

These coefficient are then used to compute the weight matrix as explained earlier.

4.2 Main results

The results are shown in table 4. The coefficients are very stable across the different specifications. The results for the estimation which gives equal weights to each neighbor in the village are given in column (1). Column (2) gives lower weights to people from a different social group. Column (3) gives the results for the estimation with an occcupation-weighted matrix.

The results show a strong spillover effect of education: the variable "education of neighbors" is highly significant at a 0.1% level for each estimate. Moreover, the measured magnitude of the spillover is the same for the three definitions of neighborhood: on additionnal

Table 3: LEVEL OF SPILLOVER EFFECTS

	(1) Productivity	(2) Productivity	(3) Productivity
HH Education	0.0112***	0.0103***	0.0111***
	(0.00205)	(0.00207)	(0.00205)
Neighbors' Education	0.0302***		
	(0.00497)		
Caste-weighted Neigh's Educ level		0.0302***	
		(0.00440)	
Occup-weighted Neigh's Educ level			0.0296***
			(0.00485)
Days worked	0.0502***	0.0510***	0.0502***
	(0.0107)	(0.0107)	(0.0107)
Agr. equipment	0.0835***	0.0814**	0.0830**
	(0.0252)	(0.0252)	(0.0252)
Irrig. land	0.414***	0.412***	0.414***
	(0.0213)	(0.0213)	(0.0213)
Fertilizer p.a.	0.000112***	0.000112***	0.000112***
	(0.0000189)	(0.0000188)	(0.0000189)
Distance to town	-0.00309***	-0.00307***	-0.00307***
	(0.000722)	(0.000721)	(0.000721)
% of HH with elect	0.00176***	0.00177***	0.00177***
	(0.000306)	(0.000304)	(0.000306)
Credit access	0.0122	0.0113	0.0130
	(0.0160)	(0.0159)	(0.0159)
Nb of seasons	0.0228	0.0215	0.0228
	(0.0177)	(0.0177)	(0.0177)
Regional Fixed-Effects	YES	YES	YES
Constant	4.502***	4.504***	4.506***
	(0.186)	(0.186)	(0.186)
Observations	11062	11062	11062
r2	0.398	0.398	0.398

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

year of education in the neighborhood increases productivity by 3%.

In the three specifications, almost all the control variables are statistically significant at a 0.1% level and have the expected signs. Not surprisingly, productivity increases with agricultural equipment, number of days worked on the field, irrigation and the amount spent in fertilizers. Being far from a city decreases productivity whereas having electricity in the village increases productivity. Nevertheless, the number of seasons where the land is cultivated is not significant whereas it was in the first two specifications. It may be due to the fact that this variable reflects the region where the household is implanted and is absorbed by the regional fixed-effects when added. The village-level variable "credit access" is not significant. It seems that having a bank branch office or a credit cooperative in the village has no impact on productivity.

5 Robustness check

The results obtained are based on hypotheses which may bias the results if they are not true. To check the robustness of our results, we test these hypotheses. As the three specifications bring the same results, we only conduct our tests on the estimation with the caste-weighted matrix because the spillover measured with this estimate has the lowest variance.

5.1 Migrations

The first hypothesis is that migration is too small to be taken into account. Actually, even if the rate of migration is very low, if people choose to migrate close to people they look like according to factors influencing farms' productivity, it conducts to an overestimation of the neighborhood influence.

As our data let us know since when each household lives in the village, we can differentiate between "old households" and "new households". So we estimate our regression without new households. The caste-weighted education of neighbors variable is re-calculated excluding "new households". Column (1) of table 4 in the appendix is the result of the benchmark regression. Column (2) excludes people who arrived in the last two years before the survey. Column (3) takes a larger definition of "new people": it excludes people who arrived in the last ten years in the village. The results show that whatever the definition of new households is, when we eliminate these people from the sample, the results stay the same. The spillover effect is between 3% and 3.1%.

But this specification takes only into account the situation where the whole household has moved. Actually, the principal source of migration from village to village is women migration for marriage purpose [Rosenzweig and Stark, 1989]. To check if this kind of migration can create biases in our results, we evaluate the knowledge spillover without taking into account women's level of education with the reduced sample without "new households"[10].

The results are shown in column (4) of table (4). The Neighbors' education coefficient is slightly lower: the spillover is evaluated at 2.4%. But it remains significant showing that the spillover effect measured is not only a reflect of people with the same level of education migrating in the same place.

5.2 Moulton effect

Another hypothesis which we have made is the independence of the error terms. However, as underlined by Moulton [1990], the introduction of aggregate level data to estimate micro data with ordinary least square can lead to downward biased standard errors. Nevertheless we cannot correct this potential spatial autocorrelation of residuals by clustering our data because our sample contains villages with less than ten households. To test the robustness of our results, we estimate our model on a reduced sample. We only keep households in villages with at least ten households in the village and the standard errors are clusterised.

[10]New households here are households who are in the village since less than three years.

The sample has now only 7,193 observations.

The results are given in column (1) of table 5 in the appendix. With this reduced sample, having agricultural equipment is not anymore significant. The number of days worked is now only significant at a 5% level and the amount of fertilizer used at a 1% level. Village level-variables are also not significantly different from zero except the variable which measures the pourcentage of having electricity in the village. However, the variable education of neighbors is still significant at a 5% level and its level is close to the previous estimates: the spillover effect if estimated at 2.4%.

5.3 Endogeneity

The endogeneity problem due to omitted variables is taken into account by adding regional dummies and village-level variables. As mentioned earlier, the much easier solution of adding village fixed-effect is not possible because of the small number of observations per village and the small within village variance of the neighbors' level of education variable.

To check the efficiency and consistency of our solution, we use the same reduced sample with only villages where there are more than ten observations and we run the same regression with village fixed-effects. As there is more than ten households per village, the data are clusterised. The results are reported in column (2) of table 5 in the appendix.

The results confirm our previous findings. Even if the standard error has hugely increased for the neighbors' variable, it is still significant at a 1% level. The measured spillover effect is even higher.

Nevertheless, as mentioned earlier, the variance of the education of neighbors' variable is small within village. So it may not be estimated efficiently with fixed-effects. To check this issue, we use the fixed effects vector decomposition proposed by Plumper and Troeger [2007]. This methodology consist in an estimation in three stages. First, the baseline model is estimated with fixed effects. Then, the fixed effect is regressed on the invariant and rarely changing variables within village. The goal is to decompose the fixed effect into two parts:

its explained part and its unexplained part. The third step consist in estimating again the baseline model but this time with the invariant and rarely changing variables and the unexplained part of the fixed effect. We apply this methodology to our model on the reduced sample.

The results in column (3) of table 5 shows the result of the second step. Column (4) shows the third step estimates. It confirms the importance of the spillover effect. The education of neighbor variable is significant at a 0.1% level. The level of the spillover is around 2.6%, which again confirms the previous results.

5.4 Control variables and specification

Finally, the robustness of our specification is tested by adding control variables and changing some variables specification.

First, we test the model by adding a variable which indicates if there was a primary school in the village at the time the most educated man of the household had the age to go to school[11]. As people who live in the same village have correlated levels of education depending on the education supply, we add this variable to control for this correlation.

We also add the caste of the household's head[12]. According to Anderson [2005] people have access to certain essential resources for agriculture such as water depending on their caste group and caste position in the village. So the caste membership may be a determinant of farm productivity.

We also test the robustness of our results by testing different variable specifications for the level of education of the household and neighbors and for the productivity calculation.

The results are reported in table 6 and 7 in the appendix. In each table the column

[11] we calculate the time passed since the most educated man would had finished primary school. We fix the end of primary school at 11 years old. And we compare this time with the date of opening of the first primary school. If it opened after the date the most educated man was supposed to finish school, we consider that there was no school in the village. If it opened before we consider that there was a school.

[12] In our survey people are classified in five groups: Brahmin, OBC, ST, SC and others. We take this classification to proxy the caste group of the household head.

(1) reports the baseline estimates. As expected, as shown in column(2) of table 6, having a school in the village has a positive impact on productivity and is significant at a 5% level. The variable of concern, the educational level of neighbors is still significant and has the same level.

When we include the caste group, being a SC or a ST has a negative and significant impact on productivity, and the spillover effect is similar to the one with the main specification: it is significant at a 1% level and its level is 2.3%.

Table 8 confirms the robustness of our results. Column (2) and (3) report estimates where the crop price to calculate household productivity is respectively calculated at the district level without excluding extreme values, and at the village level. Column (4) to (6) report estimates with different definition of education: column (4) takes into account only the education of the most educated adult in the household ; column (5) only considers the household's chief education ; finally column (6) reports estimates where the level of education of neighbors is taken into account by adding the proportion of more than 15 years old in the village having completed primary school. Whatever the definition of price or education we use, the spillover effect is significant at a 0.1% level and is evaluated between 2% and 3.1%.

6 Conclusion

Education spillovers are a concern for education policies, especially when there is a trade off with other policies. Nevertheless, education spillovers have been empirically tested only inside firms or in an urban context.

In this paper, our goal is to test and evaluate the existence of education spillovers in a rural context in a developing country.

We use cross-sectional data from the India Human Development Survey of 2005, to evaluate the impact of education spillovers in Indian villages on farm productivity. We find that education spillovers do occur: one additional year in the mean educational level of neighbors increases farm productivity by 3%. We further explore the relationships between neighbors to have a more precise measure of education spillover. In a first time we differentiate between neighbors on account of their caste. The results show that the spillover effect is stronger between people from the same caste than between people from other castes. In a second time we differentiate between people on account of their occupation. We find that the spillover is stronger with neighbors who have agricultural as their main activity. But on the overall, whatever the specification is, the global spillover effect is around 3%. The robustness checks confirm these results.

Our findings confirm the literature on education spillovers as well as the literature on neighborhood effects in rural areas in developing countries. It underlines the fact that education policies' output in developing countries should not be evaluated only on account of private returns expected but also on social returns.

Nevertheless many ways are still to explore in this field. For example, further research could look to the channels through which spillovers arise in a rural context. Moreover, as underlined by Manski [1993], "social effects might be transmitted by distributional features other than the mean". This could also be a room to explore.

References

Daron Acemoglu and Joshua Angrist. How large are the social returns to education? evidence from compulsory schooling laws. Technical report, National Bureau of Economic Research, Inc, December 1999.

Siwan Anderson. Caste as an impediment to trade. *University of British Columbia Mimeo*, 2005.

Abhijit Banerjee, Lakshmi Iyer, and Rohini Somanathan. Public action for public goods. Technical report, National Bureau of Economic Research, Inc, February 2007.

Timothy Besley and Anne Case. Diffusion as a learning process: Evidence from HYV cotton. Technical report, Princeton University, Woodrow Wilson School of Public and International Affairs, Research Program in Development Studies., May 1994.

Anne Case. Neighborhood influence and technological change. *Regional Science and Urban Economics*, 22(3):491–508, 1992.

Antonio Ciccone and Giovanni Peri. Identifying Human-Capital externalities: Theory with applications. *Review of Economic Studies*, 73(2):381–412, 2006.

Timothy G. Conley and Christopher R. Udry. Learning about a new technology: Pineapple in ghana. *American Economic Review*, 100(1):35–69, 2010. URL `http://ideas.repec.org/a/aea/aecrev/v100y2010i1p35-69.html`.

Timothy G Conley, Fredrick Flyer, and Grace R Tsiang. Spillovers from local market human capital and the spatial distribution of productivity in malaysia. *The B.E. Journal of Economic Analysis & Policy*, 3(1), December 2003.

Robert Delige. *Les castes en Inde aujourd'hui*. Presses Universitaires de France - PUF, June 2004.

Andrew D Foster and Mark R Rosenzweig. Learning by doing and learning from others: Human capital and technical change in agriculture. *Journal of Political Economy*, 103(6): 1176–1209, 1995.

Julie LE GALLO. Econométrie spatiale 1 -Autocorrélation spatiale. Technical report, LATEC, Laboratoire d'Analyse et des Techniques EConomiques, CNRS UMR 5118, Université de Bourgogne, June 2000.

Dominique Goux and Eric Maurin. Close neighbours matter: Neighbourhood effects on early performance at school. *Economic Journal*, 117(523):1193–1215, 2007.

Ashok Gulati and Anil Sharma. Subsidy syndrome in indian agriculture. *Economic and Political Weekly*, 30(39):A93–A102, September 1995.

Eugene Kandel and Edward P Lazear. Peer pressure and partnerships. *Journal of Political Economy*, 100(4):801–17, 1992.

Michael Kremer. The O-Ring theory of economic development. *The Quarterly Journal of Economics*, 108(3):551–75, 1993.

Robert Jr. Lucas. On the mechanics of economic development. *Journal of Monetary Economics*, 22(1):3–42, 1988.

Charles F Manski. Identification of endogenous social effects: The reflection problem. *Review of Economic Studies*, 60(3):531–42, 1993.

Alfred Marshall. *Principles of economics*. Macmillan and co., limited, 1890.

Pedro S. Martins. Firm-Level social returns to education. Technical report, Centre for Research into Industry, Enterprise, Finance and the Firm, February 2004.

Alexandre Mas and Enrico Moretti. Peers at work. Technical report, National Bureau of Economic Research, Inc, September 2006.

Enrico Moretti. Human capital externalities in cities. Technical report, National Bureau of Economic Research, Inc, April 2003.

Enrico Moretti. Workers' education, spillovers, and productivity: Evidence from Plant-Level production functions. *American Economic Review*, 94(3):656–690, 2004.

Brent R Moulton. An illustration of a pitfall in estimating the effects of aggregate variables on micro unit. *The Review of Economics and Statistics*, 72(2):334–38, 1990.

Kaivan Munshi. Social learning in a heterogeneous population: technology diffusion in the indian green revolution. *Journal of Development Economics*, 73(1):185–213, 2004.

M. Murthi, P. V. Srinivasan, and S. V. Subramanian. Linking indian census with national sample survey. *Economic and Political Weekly*, page 783792, 2001.

Guy Navon. Human capital spillovers in the workplace: Labor diversity and productivity, June 2009.

OCDE. Politiques agricoles des conomies mergentes 2009, suivi et valuation. Technical report, OCDE, 2009.

Thomas Plumper and Vera E. Troeger. Efficient estimation of Time-Invariant and rarely changing variables in finite sample panel analyses with unit fixed effects. *Political Analysis*, 15(2):124–139, April 2007.

George Psacharopoulos and Harry Anthony Patrinos. Returns to investment in education : a further update. Technical report, The World Bank, September 2002.

Mark R Rosenzweig and Oded Stark. Consumption smoothing, migration, and marriage: Evidence from rural india. *Journal of Political Economy*, 97(4):905–26, 1989.

Mysore Narasimhachar Srinivas. Caste in modern india and other essays, 1962.

Appendix

Table 4: MIGRATION ROBUSTNESS CHECK

	(1) Productivity	(2) Productivity	(3) Productivity	(4) Productivity
HH Education	0.0103***	0.0101***	0.0104***	
	(0.00207)	(0.00207)	(0.00207)	
High. man education				0.00712***
				(0.00170)
Caste-weighted Neigh's Educ level	0.0302***			
	(0.00440)			
Neig'educ (more than 2 years)		0.0305***		
		(0.00448)		
Neig'educ (more than 9 years)			0.0308***	
			(0.00450)	
Neig'educ (men/more than 2 years)				0.0236***
				(0.00368)
Days worked	0.0510***	0.0511***	0.0524***	0.0497***
	(0.0107)	(0.0107)	(0.0107)	(0.0110)
Agr. equipment	0.0814**	0.0825**	0.0810**	0.0896***
	(0.0252)	(0.0252)	(0.0253)	(0.0254)
Irrig. land	0.412***	0.415***	0.416***	0.416***
	(0.0213)	(0.0213)	(0.0213)	(0.0215)
Fertilizer p.a.	0.000112***	0.000113***	0.000113***	0.000112***
	(0.0000188)	(0.0000189)	(0.0000191)	(0.0000188)
Nb of seasons	0.0215	0.0177	0.0147	0.0137
	(0.0177)	(0.0177)	(0.0177)	(0.0178)
Distance to town	-0.00307***	-0.00301***	-0.00305***	-0.00330***
	(0.000721)	(0.000721)	(0.000722)	(0.000729)
% of HH with elect	0.00177***	0.00178***	0.00180***	0.00183***
	(0.000304)	(0.000305)	(0.000305)	(0.000308)
Credit access	0.0113	0.0122	0.0119	0.0130
	(0.0159)	(0.0160)	(0.0159)	(0.0162)
Constant	4.504***	4.510***	4.503***	4.583***
	(0.186)	(0.186)	(0.186)	(0.184)
Observations	11062	11037	10977	10719
r2	0.398	0.399	0.399	0.396
F	102.1	102.0	101.6	97.06
ll	-12356.4	-12319.4	-12217.7	-11943.6

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: SAMPLE WITH AT LEAST TEN PEOPLE PER VILLAGE

	(1)	(2)	(3)	(4)
	Productivity	Productivity	fixed-effect residu	Productivity
HH Education	0.00937***	0.0105***		0.00861***
	(0.00240)	(0.00229)		(0.00205)
Caste-weighted Neigh's Educ level	0.0238*	0.0421**	0.0194	0.0261***
	(0.0103)	(0.0153)	(0.0107)	(0.00301)
Days worked	0.0439*	0.0454**		0.0452***
	(0.0204)	(0.0145)		(0.0110)
Agr. equipment	0.0627	0.0767*		0.0796**
	(0.0355)	(0.0304)		(0.0273)
Irrig. land	0.418***	0.333***		0.334***
	(0.0391)	(0.0343)		(0.0227)
Fertilizer p.a.	0.000142**	0.0000904**		0.0000905**
	(0.0000534)	(0.0000341)		(0.0000297)
Nb of seasons	0.00696	0.0352		0.0360
	(0.0335)	(0.0266)		(0.0200)
Distance to town	-0.00240		-0.00242	-0.00236***
	(0.00210)		(0.00210)	(0.0000621)
% of HH with elect	0.00209*		0.00231**	0.00222***
	(0.000816)		(0.000818)	(0.0000522)
Credit access	0.0298		0.0332	0.0291***
	(0.0424)		(0.0434)	(0.00214)
Residuals				0.998***
				(0.0108)
Constant	4.313***	5.238***	-1.169	4.239***
	(0.666)	(0.112)	(0.647)	(0.0772)
Observations	7193	7193	7193	7193
r2	0.409	0.0759	0.565	0.599
F	.	26.63	.	.
ll	-7672.5	-6269.1	-3705.0	-6274.5

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6: ADDITIONNAL CONTROL VARIABLES

	(1) Productivity	(2) Productivity	(3) Productivity
HH Education	0.0103***	0.00988***	0.00906***
	(0.00207)	(0.00207)	(0.00208)
Caste-weighted Neigh's Educ level	0.0302***	0.0303***	0.0225***
	(0.00440)	(0.00441)	(0.00454)
Days worked	0.0510***	0.0512***	0.0522***
	(0.0107)	(0.0107)	(0.0107)
Agr. equipment	0.0814**	0.0821**	0.0702**
	(0.0252)	(0.0252)	(0.0251)
Irrig. land	0.412***	0.413***	0.404***
	(0.0213)	(0.0213)	(0.0212)
Fertilizer p.a.	0.000112***	0.000113***	0.000112***
	(0.0000188)	(0.0000189)	(0.0000186)
Nb of seasons	0.0215	0.0211	0.0155
	(0.0177)	(0.0177)	(0.0177)
Distance to town	-0.00307***	-0.00313***	-0.00292***
	(0.000721)	(0.000721)	(0.000720)
% of HH with elect	0.00177***	0.00171***	0.00166***
	(0.000304)	(0.000309)	(0.000304)
Credit access	0.0113	0.00971	0.0151
	(0.0159)	(0.0159)	(0.0159)
School		0.0384*	
		(0.0190)	
OBC			0.0130
			(0.0366)
SC			-0.0796*
			(0.0402)
ST			-0.159***
			(0.0440)
Other caste groups			0.0496
			(0.0381)
Constant	4.504***	4.487***	4.570***
	(0.186)	(0.186)	(0.188)
Observations	11062	11042	11062
r2	0.398	0.399	0.402
F	102.1	101.9	99.98
ll	-12356.4	-12339.3	-12325.5

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 7: VARIABLE SPECIFICATION CHECKS

	(1) Productivity	(2) Productivity	(3) Productivity	(4) Productivity	(5) Productivity	(6) Productivity
HH Education	0.0103***	0.0107***	0.00927***			0.0108***
	(0.00207)	(0.00222)	(0.00209)			(0.00205)
Highest Adult education				0.00821***		
				(0.00168)		
HH chief education					0.00849***	
					(0.00172)	
Caste-weighted Neigh's Educ level	0.0302***	0.0308***	0.0305***			
	(0.00440)	(0.00465)	(0.00451)			
Neigh's Educ (heighest adult)				0.0195***		
				(0.00339)		
Neigh's HH chief educ					0.0231***	
					(0.00427)	
Prop of primary schooled						0.393***
						(0.0583)
Days worked	0.0510***	0.0577***	0.0510***	0.0485***	0.0563***	0.0517***
	(0.0107)	(0.0115)	(0.0110)	(0.0108)	(0.0107)	(0.0107)
Agr. equipment	0.0814**	0.0716**	0.0700**	0.0854***	0.0950***	0.0828**
	(0.0252)	(0.0265)	(0.0251)	(0.0251)	(0.0249)	(0.0252)
Irrig. land	0.412***	0.421***	0.413***	0.414***	0.415***	0.415***
	(0.0213)	(0.0233)	(0.0224)	(0.0213)	(0.0213)	(0.0213)
Fertilizer p.a.	0.000112***	0.000126***	0.000117***	0.000113***	0.000112***	0.000112***
	(0.0000188)	(0.0000220)	(0.0000210)	(0.0000189)	(0.0000189)	(0.0000188)
Nb of seasons	0.0215	0.0285	0.0306	0.0226	0.0207	0.0216
	(0.0177)	(0.0194)	(0.0186)	(0.0177)	(0.0176)	(0.0177)
Distance to town	-0.00307***	-0.00244**	-0.00292***	-0.00311***	-0.00319***	-0.00317***
	(0.000721)	(0.000764)	(0.000736)	(0.000722)	(0.000721)	(0.000721)
% of HH with elect	0.00177***	0.00210***	0.00207***	0.00190***	0.00201***	0.00169***
	(0.000304)	(0.000308)	(0.000307)	(0.000302)	(0.000303)	(0.000310)
Credit access	0.0113	-0.00574	0.0000427	0.0159	0.0220	0.0120
	(0.0159)	(0.0171)	(0.0165)	(0.0159)	(0.0159)	(0.0159)
Constant	4.504***	4.429***	6.822***	4.527***	4.544***	4.418***
	(0.186)	(0.188)	(0.191)	(0.186)	(0.187)	(0.189)
Observations	11062	11062	11062	11062	11037	11062
r2	0.398	0.407	0.397	0.397	0.397	0.398
F	102.1	98.78	97.21	102.1	101.3	103.0
ll	-12356.4	-13088.9	-12719.5	-12368.6	-12340.1	-12356.7

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$