AN APPROACH OF PLS METHOD APPLIED TO MODEL THE RICE SELF-SUFFICIENCY OF PEASANT HOUSEHOLDS IN ATSINANANA MADAGASCAR

Manase Bezara, Salima Taibi

Proceedings of the 2018 International Conference "ECONOMIC SCIENCE FOR RURAL DEVELOPMENT" No 48
Jelgava, LLU ESAF, 9 11 May 2018, pp. 321-327
DOI 10.22616/ESRD.2018.101

AN APPROACH OF *PLS* METHOD APPLIED TO MODEL THE RICE SELF-SUFFICIENCY OF PEASANT HOUSEHOLDS IN ATSINANANA MADAGASCAR

Manase Bezara[1,] Dr., **Salima Taibi**[2,] Dr.Hab
[1,2] Ecole Superieure d'Agro Developpement International ISTOM.

Abstract. The aim of our study is to build a predictive model using data from survey studies. As the predictors were mostly qualitative variables, we applied an extension of PLS method. This work responds also to the need expressed by the Atsinanana agricultural region of Madagascar to strengthen the decision-making process in the fight against "no-self-sufficiency in rice" and the poverty in this locality.The study consists of constructing a simple scientific tool to improve the comprehension and to explain the links between permanent poverty and the main socio-economic factors: schooling, illiteracy, fragmentation of land, insecurity, isolation, demography) and to make effective the measures taken.The data is a random sample from a database collected in the 82 rural communes of Atsinanana by the students of the University of Toamasina and from the Campus Paysan project. We propose a Partial Least Square regression model of rice self-sufficiency.This study has helped to update the information on the subject and to deepen the knowledge on the issue. It is part of a search for improvement of statistical prediction methods by the PLS regression. We worked mainly with free software R.

Key words: peasant, multicolinearity, PLS method, model, self-sufficiency in rice.
JEL code: C38; Q01

Introduction

In this 21st century, the poverty rate is more than 70 % (*Dabat et al., 2001*). A situation has degenerated, as described in the 2016 World Bank report, which indicated for the principle Island, an alarming poverty rate of 92 %. Most of these poor people live in rural areas, depend on agriculture for their survival and are not self-sufficient in staple food, rice *(UPDR-FAO-CIRAD, 2001).*

In Madagascar, rural households face the scourge of widespread poverty. The enclaves of prosperity are rare. Studies have shown a certain correlation both between poverty and self-sufficiency in rice *(Dabat et al., 2001).*

In 2004, at the beginning of the Peasant Campus project *, the demand for rice for Malagasy national consumption indicated a gap of 200 thousand tons *(campus paysan project).* Then as a result of national effort, official data indicated an increase in production from 2007 to 2011, but the effort was not sustained. However, in 1970, Madagascar exported its rice production surpluses. Domestic supply satisfied domestic demand.

In 2017, the main problem in Madagascar was still the inadequacy of rice production, which appeared in several aspects: economic, political and religious. The records of the Ministry of Agriculture indicate that the country had 30 million hectares of uncultivated farmland in *June 2017.* Experts pointed out that the country has potential in terms of rice production but it's handicapped by the lack of control over the means of production.

In December 2017, Madagascar was on the verge of a socio-political crisis related to rice. Domestic production was in sharp decrease. The market price has seen an extremal peak of 2450 Ar / kg = 0.78 euro, with a *smic** equals 35 euros, and an average consumption 200 kg / year / Malagasian.

Questions have been raised: how to make Malagasy rice farming efficient? How to deal with this peasant hysteresis in terms of rice self-sufficiency?

The following two previous studies represent a real interest.

- The national agricultural census *(RNA, 2005)* establishes a diagnosis of the situation of farms in the Malagasy countryside. In fact, there is a permanent degradation of the areas resulting from

[1]*Corresponding author. E-mail address: m.bezara@istom.fr,*
[2]*Corresponding author. E-mail address:salima.taibi@unilasalle.fr*

parcels fragmentation, the use of hand tools such as spades, machetes and sickles for more than 90 % of farmers. We can observe little use of hitching equipment, so on the one hand, the problem of insecurity and theft operations and the decrease of rural areas has finally damaged the desire of investment.

- The analysis of rural households in Madagascar focuses on such issues as isolation and poverty and analyses the reasons why there is this rural Malagasy poverty, and how to relate the poverty level of the rural household with certain factors including isolation?

These questions represent the basis of our study focused mainly on rice farmers in the eastern region of Madagascar, formerly well-off and dynamic.

We collected, processed and analysed data on the peasantry of the Atsinanana region and sought to build a model explaining the rice self-sufficiency of rural households by projection on latent structures.

1. Data collection

We used data from two sources:

- The survey carried out as part of the Master2-MIA-FacSc / Univ Toamasina 2015-2016, and carried out by groups of students from the Faculty of Science, in collaboration with the regional statistics service of the Atsinanana region;
- Data from the peasant Campus survey (2005-2006), see *Taïbi et al. 2008*.

We applied a probabilistic approach, the cluster sampling. From the 82 rural communes of the Atsinanana region, we drew a random sample of 3 communes from which we sent to all households practicing irrigated rice a questionnaire to be completed face-to-face.

We recorded an average return rate of about 35 % covering about 150 people, or around 30 households.

The variables measured are the income (or wealth) of the household in a year, the area harvested, the size of the household, the number of children (schooling, out-of-school), rice production, the level of self-sufficiency in rice, index of family literacy, landlockedness, isolation, security, ...).

2. Partial least square regression (PLS)

In fact, prediction problems often face the phenomenon of multicollinearity.

Correlations decrease the performances of modelling processes. Methods have been developed to overcome this problem Breiman (2001), Taibi *et al.,* (2010). Projection regression on latent structures or also by partial least squares is one of the most successful.

PLS is a method which performs to explain a set of Y target variables by a set of X predictors, using a construction-iterative process *(Wold et al., 2001)* that produces two sets of synthetic variables.

- One $(t_k)_k$, $k < min\ (card\ (X), card(Y))+1$ from predictive variables, two by two orthogonal.
- The other $(u_n)_n$ resulting from the predictive variables Y, on which the condition of orthogonality is not necessary.

The point of articulation of the iterative process is to group at each step the pairs (t_k, u_k) such as Max cov (t_k , u_k). These (t_k, u_k) we call "*pls components*" or latent structures.

[1]*Corresponding author. E-mail address: m.bezara@istom.fr,*
[2]*Corresponding author. E-mail address:salima.taibi@unilasalle.fr*

3. Algorithm

Step 1: The first component pls, t1, is constructed from a specific linear combination of p centred explanatory variables x_j.

The coefficients of the linear combination are chosen in order to better summarize the explanations for a better explanation of the target Y.

This implies:

$$t_1 = \sum_{k=1}^{p} \omega_{1k} x_k$$
$$\omega_{1k} = \frac{\mathrm{cov}(x_k, y)}{\sqrt{\sum_{k=1}^{p} \mathrm{cov}^2(x_k, y)}} \quad (1)$$

We perform a model of simple linear regression of y on t_1

$$y = \frac{\mathrm{cov}(t_1, y)}{\sigma_{t1}^2} . t_1 + y_1 , y_1 \quad (2)$$

is the residual vector.

Then:

$$y = c_1.\omega_{11}.x_1 + ... + c_1.\omega_{1p}.x_p + y_1 = a_{11}x_1 + ... + a_{1p}x_p + y_1 \quad (3)$$

The translation to a matrix writing gives us.

$$\omega_1 = \frac{X^t.y}{\|X^t.y\|}; t_1 = X.\omega_1; c_1 = \frac{y^t t_1}{t_1^t t_1}; a_1 = c_1.\omega_1 \quad (4)$$

Step 2: We construct the second component pls t_2, uncorrelated to t_1, which explains the residual y_1 by a linear combination of x_{ij} regression residuals x_j variables on t_1.

The procedure is identical to that of step 1 to have (t_2, w_2, c_2 , a_2).

Step k: This procedure can be continued by iteration on the residuals. The number k of pls t_k components to retain is determined by cross-validation.

4. Treatment and comments

Our comments are based on the indications of work of *Tenenhaus (1998, 2001)* and Chavent *et al.* (2003), which will be focused on five pronunciations.

a-Total quality of regression and proportion of explained variance.

The first output of results points out the explicative performance of explicative and target variables in the building of latent variables, the number of which is limited by that of explicative variables.

Nevertheless, for targets, the calculated percentages express the explicative power of the model.

As part of our study, the first both constructed latent variables explain only 65.84 % of the information brought by all explicative variables. What implication for the four remaining latent variables with only 34.16 % of information.

This first result also shows that the model is of a rather good quality 72.74 % changeability of targets. Therefore, our research question on poverty is explained by explicative variables. If both first latent variables are kept, this performance is still equal to 67.17 %.

Let us note that, the consideration of all latent variables identifies this performance with coefficient of determination R^2 of the linear regression. In our case, R^2= 0.72.

[1]Corresponding author. E-mail address: m.bezara@istom.fr,
[2]Corresponding author. E-mail address:salima.taibi@unilasalle.fr

Table 1

Performance of the variables in explanation of latent factors

Latent Factor	Input variables (X)		Target Variables (Y)	
	Current X (%)	Cumulative X (%)	Current Y (%)	Cumulative Y (%)
1	43.442	43.442	55.256	55.256
2	22.407	65.849	11.92	67.176
3	19.599	85.448	1.152	68.329
4	3.937	89.384	3.293	71.621
5	5.402	94.786	0.699	72.32
6	5.214	100	0.053	72.374

Source: author's calculations, R software

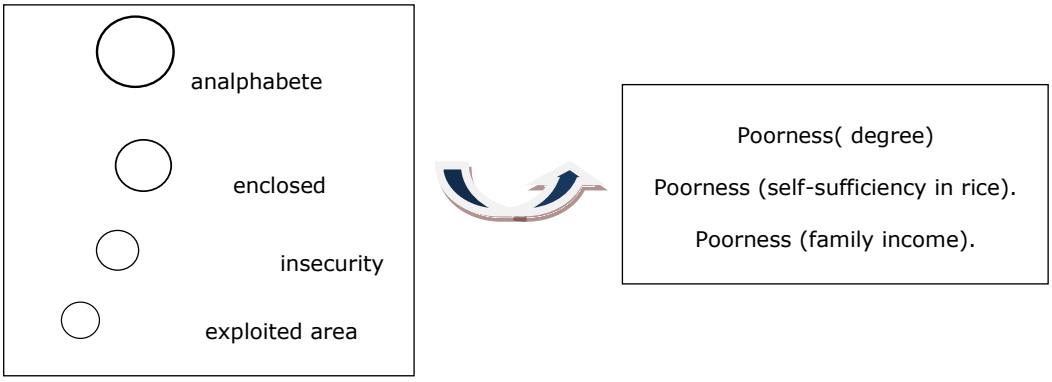
b- Raw contribution in latent variables

The second output result, points out *raw link* (square correlation) of every attribute in latent variables. In our case, we see that four variables: *enclosed- insecurity- alphabetisation- exploited area* are strongly linked to the first latent variable. Their roles are therefore determinants in the explanation of this first Pls component.

Nevertheless, it's important to realise that at this level of the study, we can't state the real nature of the connection yet. We speak about absolute intensity of the link.

As for targets, we see that the variables: *degree of poverty, family ease, family income, self-sufficiency in rice* are explained well. Indicators in the output show us it: 77.55 % (degree of poverty), 68.48 % (family income), 64.75 % (self-sufficiency in rice).

Then, it was possible to quantify correspondence.



The second latent variable contains 11.4 % available information. The variables "*age of the head of household*, *size of household* are influential. For the target variable, *rice-rowing/ ha* is well explained. It is explicative of the poverty at the level of 11.92 %.

Concerning the third latent variable, the proportion of explained variance is 19.6 %. Some determination of the variable indication of schooling is pointed out. But third axis doesn't give enough explanation of poverty problem.

c- Sense of links and weight of variables

Loadings bring the sense of links in the logic of correlation sign. We think here that the upper absolute value 0.40 indicates a significant correlation.

At this level, we can comment *Pls component* more correctly. The first Pls component is formed by a conjunction of in *alphabetisation* and *exploited area* and this, in opposition to variables *insecurity, is enclosed*. So, a household taught to read and write, exploiting a proper area living in a village not enclosed without too much problem of security is fortunate enough to be self-sufficient

[1]Corresponding author. E-mail address: m.bezara@istom.fr,
[2]Corresponding author. E-mail address:salima.taibi@unilasalle.fr

DOI 10.22616/ESRD.2018.101

in rice and boost its family income. The second Pls component is procreated by no opposition. A conjunction of *the age of the head of household* with the *size of household* for a better access to a proper *rice-growing* can also be seen.

It is also possible to comment on the indicator *weight*. For target variables, *weight* reflects their correlation with Pls component (target scores) to be predicted. It allows to estimate what is explained well by the Pls component .We cannot claim that that the first Pls component explains the conjunction principally*, family income, self-sufficiency in rice*, and displays opposition with variable *degree of poverty*.

On the other hand, the *weight* of explicative variables indicates the role of each variable in the explanation of six Pls components. We can see that it's a duplication of X- loadings.

Variable importance in projection (VIP)

VIP indicator indicates the important level of explicative variables in the explanation of the target.

- VIP ≥ 1, high level of explanation.
- VIP < 0.70 + coef(regress) ~ 0 , less importance.

In the case of our study, we have aggregated the VIP indicators as shown in Table 2.

Table 2

VIP indicator

Input	*Factor 1*	*Factor 2*	*Factor 3*	*Factor 4*	*Factor 5*	*Factor 6*
Age household	0.5551	0.7094	0.7115	0.6949	0.7152	0.715
Size household	0.5558	0.7045	0.7084	0.746	0.754	0.7537
Ind.Alphab	1.3869	1.2582	1.2476	1.2193	1.2134	1.2142
School Indic	0.4172	0.4367	0.5056	0.5517	0.5505	0.5503
Exploited area	1.4506	1.4679	1.4562	1.4277	1.4209	1.4207
Encl insecur	1.0868	1.0352	1.0292	1.0635	1.0609	1.0608

Source: author's calculations, R software

We can notice that, every column gives an account of the evolution of the role of corresponding attribute. The evolution of stocks in horizontal shows the influence combined by the Pls components. In our case, we kept 6 Pls components. We see easily across Pls component, the evolution and weight of 3 variables *(Alphabetisation indicator, Enclosed and insecurity, Exploited area).*

At the same time, the irrelevance of variable of schooling indicator is determined. This study has perhaps brought a partial explanation of the inflexible rate of descholarisation" recorded in Madagascar and it, in spite of efforts and plans of type «*ept*» (education for all people).

Model of prediction of self-sufficiency in Rice

We arrive at the level which allows us to establish a model of prediction by target and variables. It can be presented in *standardised coefficients* form. The second case applies to the variables of origin without any prior transformation. We note that standardization allows an interpretation convenient. It is possible to put attribute themselves in comparison with others by level of standard deviation.

[1]*Corresponding author. E-mail address: m.bezara@istom.fr,*
[2]*Corresponding author. E-mail address:salima.taibi@unilasalle.fr*

Table 3

Parameters of the predictive model

	Target variable(s)					
Average			1.5909	0.3182	33.7273	3.0909
Sd			0.8541	0.4239	7.1991	1.269
Input	**Average**	**Sd**	**Ricgro (t/ha)**	**SelfSufRice**	**Income/P**	**Degpoorness**
Age household	37.6818	13.4142	0.15661	-0.007493	0.17437	-0.04387
Size household	4.5455	1.4385	0.062203	-0.108666	-0.564393	0.525006
Ind.Alphab.	0.4118	0.2777	0.137124	0.343618	0.302385	-0.245969
School Indic	0.3818	0.3049	-0.100946	-0.028718	0.437846	-0.335946
Exploi area	0.73	0.4826	0.564254	0.670596	0.68931	-0.54884
Encl.insecur	3.4091	1.2596	0.06307	0.023467	0.288747	-0.071989

Source: author's calculations, R software

Where the form of model :

SelfSufRice = 0.670596. Exploi. area + 0.343618. Ind Alphab + 0.023467.

Encl-Insecur – (0.108666.Sizehousehold + 0.028718.School Indic + 0.007493.Agehousehold)

It is possible to see that:

- In order of importance variables exploited area , alphabetisation indicator , enclosed - insecurity, have impacts on variable self-sufficiency in rice.
- Variables *age of the head of household, schooling indicator, number of children*, destabilise the *self-sufficiency in rice*.
- On predictive plan, if they augment for example the *exploited area* by 0.4826 ha, then the coefficient of *self-sufficiency in rice* grows for 0.670296*0.4239.

Conclusions

The established *PLS* model allowed the leaders of the 82 communal localities to have a more rational approach to the problem of *no-self-sufficiency in rice*, a main source of rural poverty in the Atsinanana region, with a view to improving the decision-making performance. Three main factors have been identified: illiteracy, isolation, land fragmentation with some ambiguity about the behaviour of the schooling indicator. A judicious choice of communal samples, and an adapted technique of data collection the elaborated PLS model would have been otherwise efficient. The difficulties are many, insofar as our data are not able to point out particular problems related to each specific municipality, but also the inexistence of statistical archives that would have helped our work. Finally, the work triggered some form of awareness-raising process among all economic and strategic actors currently seeking to set up the *Atsinanana Region Observatory of Rurality* or the *University of Toamasina's Multidisciplinary Laboratory*.

Bibliography

1. Breiman, L. (2001). *Random Forests. Machine Learning*, 45,pp. 5-32.
2. Chavent, M., Patouille, B., (2003), *Calcul des coefficients de regression et du Press en regression PLS*. Revue Modulad. Vol. 30, pp. 1-9.
3. Dabat, M., Jenn-Treyer, O., Razafimandimby, S., Bockel, M.H, (2008*), l'histoire inachevée de la régulation du marché du riz à Madagascar*. Economie rurale. 303-305, pp. 75-89.
4. Doob, J.L. (1953) *Stochastic process*, Wiley, New York., p. 654.
5. Hubert, P. J. (1981) *Robust Statistics*, Wiley, New York.

[1]*Corresponding author. E-mail address: m.bezara@istom.fr,*
[2]*Corresponding author. E-mail address:salima.taibi@unilasalle.fr*

6. Taibi-Hassani, S., Bezara, M., Rajaonarivelo, R., (2006).*Rapport de synthèse campus paysan . Université de Toamasina*, Esitpa, Région Atsinanana, Région Normandie, p. 28.
7. Taibi Hassani, S., Bezara, M. (2010). *La méthode des forêts aléatoires appliquée à l'observatoire de la ruralité dans le cadre du projet campus paysan à Madagascar. Pratiques et méthodes de sondage*, Dunod, pp. 209-213.
8. Tenenhaus, M.,(2006) *Statistique : méthodes pour décrire , expliquer et prévoir* , Dunod.
9. Tenenhaus, M., (1998*), la regression PLS, théorie et pratique*. Edition technip, p. 254.
10. Tremblay, M.E.,Lavallée, P. ,Tirari, M.E.H, (2011). *Pratiques et méthodes de sondage*.Dunod. p. 382.
11. Wold, S., Sjostrom, M., Eriksson, L., (2001). PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems Vol. 58 pp. , 109–130.

[1]*Corresponding author. E-mail address: m.bezara@istom.fr,*
[2]*Corresponding author. E-mail address:salima.taibi@unilasalle.fr*