Determinants of the Net Interest Margin in the Banking Institutions: Contribution of PLS Regression Compared to the Principal Components Regression

Benammou Salwa[1] and Mlayeh Chahrazed[2]

[1]Department of Quantitative Methods, University of Law, Economic and Political Sciences, Erriadh City 4023 Sousse, Tunisia. E-mail: **salwa.benammou@voila.fr**

[2]Department of Quantitative Methods, Higher Institute of Management, Abed Aziz El Bahi Street, BP 763 4000 Sousse, Tunisia. E-mail: **Chahrazed.Mlayeh@isffs.rnu.tn**

Corresponding author: Mahjoubi Dalenda. E-mail address: dalendaisg@yahoo.fr. Address: 14, Rue Abdallah Ibn Raouha. Cité Erriadh. 4023. Sousse. Tunisie. Tel: +21673301165

Abstract: Partial Least Squares regression and Principal Components Regression make possible to relate a set of dependant variables Y to a set of independent variables X, when there is multicollinearity. This paper suggests a new approach for analyzing the net interest margin. After using the PCR method, the determinants of the net interest margin have been viewed through a PLS model.

Keywords: Partial Least Squares Regression, Principal Components Regression, net interest margin.

INTRODUCTION

The exhaustive development requires a deep survey on the set of events that can generate unfavorable variations and that have a negative effect on the national economy. The banking institutions represent an important market that has an efficient role in the development and the creation of a new economic politics. We identify the various risks to which the banking system is exposed. If they are badly managed, these risks can affect the bank's objective and can lead to unfavorable situations. We carry out a statistical and econometric study to identify which principal risks that have an effect on the banking profit and reflect the reality economic of this market.

This research points out the determinants of the net interest margin in the Tunisian banking structure. The article is divided into two sections. In the first part we introduce some tools Partial Least Squares regression (PLS) and the Principal Components Regression (PCR) and we justify the choice of the PLS approach. In the Last part we present an application of the PLS regression to net interest margin data and we compare the results with those obtained by the PCR methods.

DATA AND METHODS

We will explain the net interest margin (NIM), which is measured by the ratio of the difference between returned in interest and the expenditure of interest on the product banking net by eight exogenous variables such as:

Spread (SPREAD): is a variable defined as the difference between the debtor rate and the credit rate.

The implicit interest payment (IMPL): is appreciated by the ratio of the difference of the expenditure except interest and the incomes except interest on the results rough of exploitation.

The credit risk (CRDTR): is measured by the ratio provision on the sum of appropriations.

The solvency risk (SOLVR): is appreciated by the prudential ratio which requires a proportional relation between the own capital stocks of each financial establishment and its balanced credits.

The interest rate risk (IRR): is measured by the ratio of short-term credit on the total credit.

The liquidity risk (LQDR): is estimated by the liquid asset ratios on short-term liability.

The opportunity cost (OPPC): is measured by the ratio of the reserves held on the total credit of the bank.

Management Effectiveness (MEFF): is measured by the ratio of the product banking net on the total credit (Angbazo, L. 1997).

PLS regression

PLS regression is a recent technique that generalizes and combines features from principal component analysis and multiple regression. It is particularly useful when we need to predict a set of dependent variables from a large set of independent variables. It originated in the social sciences (specifically economy, HermanWold1966) but became popular first in chemometrics and in sensory evaluation (Martens and Naes, 1989).But PLS regression is also becoming a tool of choice in the social sciences as a multivariate technique for non-experimental and experimental data a like (Mcintosh, Bookstein, Haxby and Grady, 1996). It was first presented as an algorithm akin to the power method (used for computing eigenvectors) but was rapidly interpreted in a statistical framework. (Frank and Friedman, 1993; Helland, 1990; Hoskuldsson, 1988; Tenenhaus, 1998).

Advantages of PLS regression:

- ✓ handles multicollinearity

- ✓ robust in terms of data noise and missing values
- ✓ balances the two objectives of explaining response and predictor variation thus predictions are more robust
- ✓ calibrations generally more robust
- ✓ single step decomposition and regression
- ✓ can give good insight into underlying relationship between variables.

By contrast, PLS regression finds components from X that are also relevant for Y. Specifically, PLS regression searches for a set of components that performs a simultaneous decomposition of X and Y with the constraint that these components explain as much as possible of the covariance between X and Y. This step generalizes PCA (Tenenhaus M., Gauchi J. P. and Ménardo C. 1995).

PLS regression consists in identifying the vectors u_1 and t_1 by maximizing the identity

$$\operatorname{cov}(t_1, u_1) = cor(t_1, u_1).\sqrt{\operatorname{var}(t_1).\operatorname{var}(u_1)}\ .$$

Where

- ✓ t_1 corresponds to new calculated variables as a linear combination of X vector t_1 summarizes X at the same time to appreciate X but also to predict Y. In the same way, we can identify a vector u_1 which corresponds to the new variables predicting Y. In the particular case where Y_0 corresponds to a scalar (only one independent variable), u_1 is compared to Y_0.
- ✓ P_1 and r_1 respectively indicate the column vector of the coefficients of regression of X_0 and Y_0 on t_1.

PLS regression is an iterative method. It is resumed in three stages (Tenenhaus 1995)

Stage 0: starting by tables X and Y.

Stage 1: construction of components u_1 and t_1 of the columns of X and Y respectively by maximization of covariance between these two components and to build the two equations:

$$\begin{cases} X = t_1 p_1^{'} + x_1 \\ Y = t_1 c_1^{'} + y_1 \end{cases}$$

Stage 2: resumption of stage 1, but we replace tables X and Y by the residues X_1 and Y_1. What gives two new components t_2 and u_2 linear combinations of the columns of X_1 and Y_1 respectively.

From where two new equations:

$$\begin{cases} X = t_1 p_1' + t_2 p_2' + x_2 \\ Y = t_1 c_1' + t_2 c_2' + y_2 \end{cases}$$

The iterations continue until t_h component may explain Y in a satisfactory way (Tenenhaus M. 1999).

Just like the Principal Component Analysis (PCA), the PLS regression makes possible to reduce the dimension of a study. Contrary to the PCA which seeks to reduce only one matrix X of the variables, the PLS will make possible to work at the same time on the matrix X and the matrix Y. The PLS consequently makes possible to model the matrix Y according to the matrix X.

Principal Components Regression (PCR)

The Principal Components Regression is the method of combining linear regression with principal component analysis. The principal component analysis can gather highly correlated independent variables into a principal component, and all principal components are independent of each other, so that all it does is to transform a set of correlated variables to a set of uncorrelated principal components. Then we built the regression equations with a set of uncorrelated principal components and get the 'best' equation according to the principle of the maximum adjusted R^2 and minimum standard error of estimate.

The following approach proceeds in three stages.

Stage 0: Proceed a stepwise regression with a dependent variable Y and all independent variables X for getting the p independent variables with statistical significances and revealing whether the p independent variables have a mutlticollinearity or not.

Stage 1: Proceed a PCA with the p independent variables for transforming a set of correlated variables to a set of uncorrelated principal components and indicating information quantities of different set of principal components.

Stage 2: build the Principal Components Regression equation with the first principal component, then add principal component backwards one by one. Then determine the 'best' equation on the basis of the maximum of adjusted R^2 and minimum standard error of estimate.

RESULTS AND DISCUSSION

Result of PLS regression

Table 1 presents the cross validation, and shows that we preserve only one component.

The choice of the component is based on the cross validation, which should not exceed the limit.We notice that starting from second component, the cross validation becomes negative (-0.32).

Table1: Cross validation

Numbers of Component	Q_h^2	Limit
l	0.218	0.0975
2	- 0.32	0.0975

This component is defined by

$$t_1 = \left[\sum_{j=1}^{8} cor\left(x_j, y\right) x_j \right] \Big/ \sqrt{\sum_{j=1}^{8} cor^2\left(x_j, y\right)}.$$

$$t_1 = -0.0602 CRDTR - 0.3587 SOLVR + 0.1657 OPPC + 0.5223 IRR + 0.5113 LQDR$$

$$0.4336 IMPL - 0.3338 SPREAD + 0.08 MEFF$$

The regression equation becomes:

$$\hat{y} = c_1 \; t_1 + y_1$$

$$\text{N}\hat{\text{I}}\text{M} = \text{c}_1 \; t_1 = 0.437 t_1$$

$$\text{N}\,\hat{\text{I}}\text{M} = 0.437\,\text{t}_1$$

$$= -0.0263 CRDTR - 0.1569 SOLVR + 0.0725 OPPC + 0.2285 IRR$$

$$+ 0.2237 LQDR + 0.1897 IMPL - 0.1460 SPREAD + 0.0351 MEFF.$$

By examining the coefficients of PLS regression, we conclude that the CRDTR, SOLVR and SPREAD are negatively correlated whereas the other variables are positively correlated.

The credit risk

The coefficient of the credit risk is negative. This indicates that the average of the interest margin drops if the bank credits are contracted. If the margin is fall, the establishment is sensitive to the fluctuations of the market interest rate, from which the market credit rates were improved by instruments whose principal role is to ensure a limitation of the credit risks according to the financial standing of the establishment while applying the credit rationing.

The solvency risk

The coefficient of solvency risk is negative. This is due to the fact that the banks substitute their own capital stocks by the debts which generate a reduction of insolvency risk and involves a fall of the costs of loans which requires a weaker allowance for insolvency risk. We also notice that this risk has a weak fluctuation during the period of study. This is due to the fact that the Tunisian banks seek to satisfy the prudential solvency rules.

Opportunity cost

The coefficient of this variable is positive. By defined, it is the output credit profit which gives up the depositors to hold the liquidity; it increases the cost of funds beyond the rate observed. To compensate for it, the bank will be brought to integrate an additional premium in the interest margin.

The rate risk

The rate risk has a positive coefficient. This coefficient means more the rate risk increases more the bank becomes demanding with its customers and requires more of the interest margin. The positive sign of the rate risk can be explained by the fact that the granted appropriations are appropriations of long term, then, the bank becomes concerned since it is not able any more to envisage the economic situation of its customers as well as the evolution of the future interest rate. In this case it runs a potential risk and thus will require a premium of maturity.

The liquidity risk

This risk is the incapacity to provide the withdrawals with the depositors or the granting of credit. This is due to the behavior of Tunisian trade banks which aim at rather holding a position of speculative costs than to a position of prudential liquidity. The bank holds the behavior of amateur to the risk.

The implicit interest payment

This variable is negative during all the study. The resources without interest exceed the expenditure without interest. The implicit coefficient of payment admits a positive sign. This is due to the payment into extra with the depositors through the service of handing-over or another type of transfers

Spread

It is expected that the coefficient of variable is positive. This is due to the effort of the disintermediation undertaken, the recourse to the other forms of debts and the money market, which generates a reduction in the degree of our function of intermediation.

Management effectiveness

The management effectiveness admits a positive coefficient. We generally wait until a good management of the bank influences, in a positive way, its deposit and credit policy like its marketing policy in order to manage improving the interest margin; which is the case.

Result of the principal components regression

The application of the principal components regression is carried out by the software SPSS which provided statistics of collinearity such as the tolerance and variance inflation factor (VIF).

Tolerance = 1- R_i^2, where R_i^2 is squared multiple correlation of i_{th} variable with other independent variables. When its value is small (close to 0), the variable is almost a linear combination of the other independent variables. VIF is reciprocal of tolerance. Variables with low tolerance tend to have large VIF, so variables with low tolerance and large VIF suggest that they have collinearity. The eigen value, condition index and the variance proportions are also indices of collinearity. If the index of collinearity is more than 15, this indicates a possible problem and an index greater than 30 suggests a serious problem with collinearity.

The execution of SPSS gives us seven significant variables which are CRDTR, SOLVR, OPPC, LQDR, IMPL, SPREAD and MEFF. The results are presented in the following table 2.

Table2: Indices of collinearity diagnosis

Model	Coefficient	t	Significance	Collinearity Statistics		Condition Index
	β			Tolerance	VIF	
Constant		2.4 E+07	0.000			
CRDTR	0.337	8529374	0.000	0.201	4.988	5.849
SOLVR	0.304	9660228	0.000	0.316	3.168	6.911
OPPC	0.952	2.2 E+07	0.000	0.16	6.234	9.661
LQDR	0.493	1.8 E+07	0.000	0.431	2.321	10.677
IMPL	0.969	2.5 E+07	0.000	0.213	4.692	31.103
SPREAD	-0.8	-3.5 E+07	0.000	0.600	1.667	36.921
MEFF	-1.028	-1.7 E+07	0.000	0.088	11.36	77.627

According to the values of collinearity statistics and condition index, we conclude the existence of multicollinearity. We will check whether there are multicollinearities among the independent variables. It also displays that condition index of three independent variables is more than 30. The MEFF can be written as a linear combination of the variables since it has a weak tolerance, tends to zero (0.088) and VIF (11.36). After a factorial analysis, we find seven principal components. We built the principal component regression equation with the first component C_1, then add principal component backwards one by one and get 'best' equation according to the principal of the maximum of adjusted R^2 and minimum standard error of estimate mentioned in table3.

Table3: Adjusted $R^2$and standard error of estimate

Standarized Principal Components Regression Equation	Adjusted R^2	Standard Error of Estimate
$N\hat{I}M = f(C_1)$	0.107	0.148
$N\hat{I}M = f(C_1, C_2)$	0.285	0.141
$N\hat{I}M = f(C_1, C_2, C_3)$	0.617	0.112
$N\hat{I}M = f(C_1, C_2, C_3, C_4)$	0.699	0.108
$N\hat{I}M = f(C_1, C_2, C_3, C_4, C_5)$	0.885	0.075
$N\hat{I}M = f(C_1, C_2, C_3, C_4, C_5, C_6)$	0.965	0.047
$N\hat{I}M = f(C_1, C_2, C_3, C_4, C_5, C_6, C_7)$	0.907	0.078

We notice that if we add the seventh component, R^2 decreases and the standard error of the estimate increases. Thus, the best equation function is according to six components.

Table 4 reports the PCR estimate of the regression coefficient of NIM of the six components.

Table 4: Regression coefficient

Component	Coefficient	t	Significance
	β		
C_1	-19.235	-2.451	0.092
C_2	4.938	2.044	0.133
C_3	- 7.184	-2.187	0.117
C_4	- 3.516	-2.242	0.111
C_5	0.070	0.274	0.802
C_6	-16.908	-2.609	0.080

This table shows that the component five is not significant; so we will make the regression by excluding this component. We noticed that R^2 does not change (0.964) but the standard error of the estimate decreases by 0.047 to 0.042 and all the components are significant.

All the components are significant. We will retain five components.
The principal components regression equation will be written

$$\widehat{\mathrm{NIM}} = -20.983C_1 + 5.487C_2 - 7.928C_3 - 3.87C_4 - 18.337C_6$$

The five components which were retained explain more than 96% of the original variance. The interpretation of the axes is specified by the factorial co-ordinates of each variable. We can thus identify the effect of each axis on the interest margin and not on the variables of origins. When according to this result, the use of the principal components regression method does not make possible to check the individual impact of each explanatory variable. Therefore, we deduce the importance of PLS regression compared to the Principal Components Regression.

CONCLUSIONS

The method of PLS regression represents a more immediate approach since the analysis of interest margin is explicitly carried out by the variables of origin. However, the method of Principal Components Regression, presupposes an interpretation of the factorial axes on which the regression will be carried out.

The contribution of PLS regression by the principal components regression is the fact that PLS regression helps us to release the effect of each variables. The PLS regression helps the bank to release the variables on which it act to increase the interest margin, because PLS regression preserves all the initial variables. On the other hand, the principal components regression eliminates some variables by carrying out a principal component analysis, which has a principal interest to reduce the dimension of the space of the variables in order to better visualize them.

REFERENCES

Angbazo L. (1997): Commercial bank net interest margins, default risk, and off-balance sheet banking. Journal of banking and finance, 21.

Frank I.E. and Friedman J.H. (1993): A statistical view of chemometrics regression tools. Technometrics, 35, 109–148.

Helland I.S. (1990): PLS regression and statistical models. Scandivian Journal of Statistics, 17, 97–114.

Hoskuldson A. (1988): PLS regression methods. Journal of Chemometrics, 2, 211-228.

Martens H. and Naes T. (1989): Multivariate Calibration. London: Wiley.

McIntosh A.R., Bookstein F.L., Haxby J.V and Grady C.L. (1996): Spatial pattern analysis of functional brain images using partial least squares. Neuroimage, 3, 143–157.

Tenenhaus M. (1999) : L'approche PLS. Revue de Statistique Appliquée, 47, 2, 5-40.

Tenenhaus M. (1998) : La régression PLS. Paris: Technip.

Tenenhaus M., Gauchi J. P. and Ménardo C. (1995) : Régression PLS et Application. Revue de Statistique Appliquée, 43, 3, 7-63.

Wold H. (1966): Estimation of principal components and related models by iterative least squares. In P.R.Krishnaiaah (Ed.). Multivariate Analysis, New York: Academic Press, 391-420.