International Journal of Economics & Strategic Management of Business Process (ESMB)
Vol. 10-pp. 53-57

Multiple regression-based models for accurate credit risk management

Marwa HASNI*, Safa BHAR LAYEB
University of Tunis El Manar, National Engineering School of Tunis
UR-OASIS: Optimization & Analysis of Service and Industrial Systems
BP 37 Le Belvédère, 1002, Tunis, Tunisia
*Marwa.Gharbi@enit.rnu.tn

***Abstract*— In banking system, credit risk represents the expectation of losses stemming from inability of a borrower to repay a loan. In the academic literature, it has been shown that inadequate management of credit risk is a driving factor of financial crises. To accurately control them, banks seek for developing portfolios of information on their customers. To achieve that, the econometric theory advocates using mathematic models (e.g. scoring), which are not only restricted to collecting information on borrowers characteristics, but also, anticipates the quantitative credit risk. In this work, we show that multiple regression-based models are accurate tools to be explored. Specifically, we focused on a real life case study of a private Tunisian bank. In fact, we designed a multiple regression model which enables the bank system to predict the total turnover of a company (i.e. a customer), within a specific time horizon, with regard to eventual changes in financial, macroeconomic and microeconomic data/variables. Accuracy of our approximation is established due to extensive simulation experimentation.**

***Keywords*—Forecasting; Management of Credit Risk; Banking System; Loan; Multiple Regression Models.**

I. INTRODUCTION

Credit risk represents eventual loss resulting from non-performance of financial contracts. For commercial banks, several driving factors of credit risk may be the bonds, short-term debt securities and derivatives. Amongst, loans are typically viewed as the mostly influencing. Besides, according to Mačcimskein et al. [1], country risk and settlement risk, which are macroeconomic indicators, are also regarded as credit risks. [2] stated that it is essential for a banking organization to effectively control credit risk in order to ensure its long-term success and evenly, to better cope with inevitable defaults due to global financial crisis. Traditional econometric tools for credit risk management may include rating methods [3], expert systems or neural networks. Relevant examples of these approaches may be found in [4]. All of these rely upon qualitative data analysis (e.g. assessment of future business strategies or appraisals of a business data) and produce estimates based on knowledge gained in the past. A significant example may be found in [5], where a multi-criteria model for credit risk assessment is developed, based on the capability of incorporating value judgements.Mačcimskein et al. [1] established that it was not until 2007, date of the global financial crisis in the USA, that researches started to advocate incorporation of quantitative analysis methods to optimize the risk/return ratio. The relevance of quantitative criteria was firstly highlighted in the work of Saunders and Allen[6], who developed new approaches such as the optional pricing models (e.g. KMV), the VAR models (e.g. CreditMetrics) and time varying models (e.g. CreditPortfolio View).

Following this line of research, the concern of our study is to design a new statistical analysis model based on multiple regressions, which provides a private Tunisian bank with accurate information on its credit market including 18 companies. Specifically, for each company, the total turnover is modelled with regard to eventual changes in financial and macroeconomic data/variables. Then, underlying model is used to predict future total turnovers for three consecutive years.

The reminder of this paper is organized as follows. In section 2, we briefly describe the case study, then, we sketch the methodological approach being processed in section 3. Details on the experimental design and underlying computation results are included in section 4 and the final conclusions are outlined in section 5.

II. THE CASE STUDY: A BRIEF DESCRIPTION

The company under study consists of a private Tunisian commercial bank in which the credit process generates, annually, about 60% of its net banking income. Yet, this process involves developing portfolios of information on all its customers as a way to efficiently allocate funds and avoid repayment defaults. For each company, underlying portfolio of information is constructed by means of judgmental estimations of its future financial solvency (i.e. ability to meet its long-term financial obligations), its future liquid assets (i.e. funds, such as stocks or bonds that can easily be converted into cash to meet financial obligations), its rate of return and its margin rate. Thereby, the decision rule of whether a loan should be leaded or not, relies upon comparison of estimates of these ratios with minimal prefixed respective, thresholds. It should be noted, however, that accurately collecting this information is a challenging task, especially because the risk of borrowers is hazard and their performance is difficult to

ISSN 2356-5608

monitor. Thereby, criticality of the credit process, together with the difficulty in controlling it, imply that more robust analysis should be performed in order to allow greater rating accuracy. For this bank, estimating both rates of returns and margin, involve subjective assumptions upon upcoming turnover rate of the related company. With respect to that, our mission is to determine a quantitative statistical tool that allows decision makers to efficiently predict future turnover rates instead of simply assuming it by "hazard". For that purpose, we explored multiple linear regressions and developed an econometric model which assumes the turnover rate of a particular company to be directly dependent on financial and macroeconomic factors.

III. Proposed Approach

Our work is divided into three phases. *In the first phase,* we checked the validity of usage of the general linear regression equation to generate our econometric model. To do that, we initially defined macroeconomic and financial variables and obtained their respective data series. Then, we conducted an ANOVA test to verify whether dependent variables (i.e. macroeconomic and financial) are in relationship with their independent counterpart (i.e. turnover rate). In the second phase, we firstly constructed our linear model following the descendent mode which we sketch below:

1. Write the general linear model using all available exogenous variables.
2. Apply a p-value test to assess statistical significance of each variable.
3. Dismiss the least significant variable and obtain the final general regression model.

In the second phase, we estimated the parameters of the population regression line using the method of least squares. Accuracy of the linear adjustment is given by the coefficient of determination which is mathematically evaluated as R-squared, where R is the coefficient of correlation. Accordingly,$R^2 = \frac{SPR}{SPT}$, whereby SPR measures the influence of exogenous variables in the total endogenous variable; and SPT is the sum of squared errors. Here, we shall note that R^2 should be comprised between 0 and1. Particularly, one would wishR^2 to be as nearest as possible from the upper bound to assert that his developed model is satisfactory. For example, if$R^2 = \alpha\%$,we conclude that exogenous variables capture $\alpha\%$ of the variability of the endogenous variable. Afterwards, it is important to investigate the residuals to determine whether or not the fitted model meets the hypothesis the multiple linear regressions. Indeed, residuals should be normally distributed with mean zero and unchanged variance. The latter feature refers to their homoscedasticity property. Besides, residuals are required to be uncorrelated between each other ($\mathrm{COV}_{i\neq j}(\varepsilon_i, \varepsilon_j) = 0$) and with exogenous variables as well. By asserting these hypotheses the regression line model is considered valid. Finally, *in the third phase*, we explore this model to forecast future turnovers, within three consecutive years.

IV. Experiment and Underlying results

A. Data details

In this real life case study, we are interested in modeling the turnover of 18 Tunisian companies operating in several sectors such as the automobile, equipment distribution and agro alimentary.Thereby, we concentrate on designing 18 multiple linear regression models. Recall from section 2 that estimates of the turnover should accommodate macroeconomic and financial variables. Macroeconomic variables consist in Inflation Rate (IR), Unemployment Rate (CR), the Growth Rate (GR) and the External Balance of Trade balance (EBT). Selection of these factors is inspired from the "magic square" of the economist Kaldor [5], which is a graphical representation linking the four main objectives of conjectural economic policy of a country. In economy studies, it is believed that this representation is successful and it is labeled by the expression of "magic square" to highlight how the simultaneous achievement of the four goals yields the economic performance of the country. In our sense, using these factors allows the bank to align with the conjectural economic strategy of Tunisia. To end with, we notice that numerical data being used are historical records laying from 2006 to 2015 praised by the National Statistical Institution (NSI)in Tunisia, as well as estimated data for years 2016, 2017 and 2018 which are provided from the worldwide bank. Now, Turning on to financial factors, we noticed that our commercial bank is considerably influenced by four indicators, namely: the Equity Ratio (ER), the Current Asset (CA), the Current Liability (CL) and the Net Income (NI). Underlying data series, laid from 2006 until 2015, and their macroeconomic counterparts are reported in Table I.

From now on, we consider that, for each customer (i.e. company); variability of its turnover is explained by eight exogenous variables according the following model equation:

$$TR_t = \alpha_0 + \alpha_1 GR_t + \alpha_2 IR_t + \alpha_3 CR_t + \alpha_4 EBT_t + \alpha_5 ER_t + \alpha_6 CA_t + \alpha_7 CL_t + \alpha_8 NI_t \quad (1)$$

In practice, we separately focused on data of each company and determined underlying model by means of multiple regression, then, by applying the descendent approach described in section 3, we maintained the most influencing variables. Table I shows the explanatory variables to be considered for each company. For example, variabilities of the turnover rate of the customer R are explained by its equity ratio, its current liability, the growth rate and the unemployment rate of Tunisia. Hence, its underlying linear model is given by:

$$TR_t = \alpha_0 + \alpha_1 GR_t + \alpha_2 CR_t + \alpha_3 ER_t + \alpha_4 CL_t \quad (2)$$

ISSN 2356-5608

TABLE I
LINEAR MULTIPLE REGRESSION RESULTS

Company	*ER*	*CL*	*NI*	*CA*	*GR*	*EBT*	*IR*	*CR*
A	×	×	×	×	×	×	×	×
B		×	×	×	×	×	×	×
C	×	×	×	×		×	×	×
D	×		×	×	×	×		
E	×	×	×		×	×	×	×
F					×			×
G		×	×	×	×		×	
H							×	
I	×		×	×	×	×	×	×
J			×	×			×	
K				×	×	×	×	×
L			×					
M		×	×	×	×	×	×	×
N	×	×	×		×	×	×	×
O				×				
P				×				
Q						×		
R	×	×				×		×

B. The model validation

With respect to our methodology described in section 3, we now carry on a statistical analysis on residuals to ensure that our developed models meet the assumptions of multiple linear regressions. In the following, we report investigations of three particular cases, where the turnover rate may depend on only financial factors, only macroeconomic factors, or else, on variables affiliated to both factors. Here, we shall note that we picked one example of each case in order to avoid redundant information.

Case1: the turnover rate depends only from financial variables

Both companies, labeled "O" and "P" are examples of this case. Indeed, their respective turnover rates depend only from the current assets. Here, we only report analysis related to the company "O". The regression model parameters yield by multiple linear regression and presented in table II, lead to the following model:

$$TR_t = 8925.553 + 3.697 * CA_t \quad (3)$$

This adjustment is viewed as highly accurate since the coefficient of adjustment R^2 is equal to 0.992 as shown in Table III.

TABLE II
PARAMETERS OF THE MODEL

Source	*Constant*	*ER*	*CL*	*NI*	*CA*	*GR*	*EBT*	*IR*	*CR*
Value	8925.55	0	0	0	3.7	0	0	0	0

To check statistical significance of the model, we use the Fisher test which involves the P-value calculation [7]. Here, the P-value is the probability that the dependence between exogenous and endogenous variables is simply a chance finding. The decision rule, recommended by Fischer, consists in asserting the significance of the model if P-value is below 0.02. However, if it is comprised between 0.1 and 0.9, then, all observations are purely due to chance.

TABLE III
QUALITY OF THE ADJUSTMENT

Observations	10.000
R²	0.992

Results of the Fisher's test for the company "O" are illustrated in Table IV and confirm the significance of the associated statistical model since the P-value given by the last column is approximately null (<0.0001). To test normality of residuals, we conducted the Kolmogorov-Smirnov test, whereby the null hypothesis suggests that residuals are normally distributed with $(1-\alpha)\%$ of certainty. In our case, for a 5%-error risk (i.e. $\alpha = 0.05$, we obtained a P-value of 97.4% which confirms the null hypothesis. Finally, the White test asserts the homoscedasticity property of residuals as associated probability is superior to 0.05. Related details on numerical findings are reported in TableIV.

TABLE IV
NORMALITY AND HOMOSCEDASTICITY VALIDATIONS

Source	Model	Error	Corrected total
DDL	1	8	9
Total Squares	581574195832.5	4412045961.1	585986241793.6
Mean Squares	581574195832.5	551505745.2	
F	1054.5		
Pr > F	**< 0.0001**		

TABLE V
NORMALITY AND HOMOSCEDASTICITY VALIDATIONS

Tests	*P-value*
Kolmogorov-Smirnov test	0.975
White test	0.792

Case2: the turnover rate depends only from macroeconomic variables

This case was observed with company "Q" where the turnover rate depends only on the external balance of trade balance (EBT). Underlying model is given by:

$$TR_t = 45902.202 + 2321.993 * EBT_t \quad (4)$$

By evaluating the determinant coefficient R^2, we conclude that external balance of trade balance explains 95.5% of variability of the turnover rate, which is a good fit. Additionally, the adopted linear model is found to be valid since the Fischer test

ISSN 2356-5608

yields a P-value, approximately equal to zero. The Kolmogorov-Smirnov test asserts that residuals are likely to be normally distributed in 73% of cases and the White test shows that their variance remains unchanged in about 58% of cases.

Case3: the turnover rate depends only from financial and macroeconomic variables

The company "A" is the only company where variabilities of the turnover rate depend on all available variables from both financial and macroeconomic factors. Its underlying equation is given by:

$$TR_t = 1847574.910 - 39647.910 * GR_t - 19006.62 * IR_t - 98979.221 * CR_t - 100238.23 * EBT_t - 60.115 * ER_t + 7.012 * CA_t + 13.042 * CL_t + 79.491 * NI_t \quad (5)$$

The determinant coefficient R^2 is equal to 1. Hence, the total variance of the turnover rate is explained by the eight variables. This model is found to be significant since the Fischer probability is 0.048, inferior to the risk probability. Besides, normality and homoscedasticity of residuals are confirmed with respective Kolmogorov-Smirnov and White probabilities: 0.73 and 1.

C. Estimation of future turnover rates

So far, this paper has focused on how to model turnover rate variations using linear multiple regression. In this section, we address a way of anticipating its future values. Note that the obtained models include time series of the explicative variables. Thus far, being at period*n*, estimates of the values of these variables at period*n+1* are required for predicting the turnover of the same period. With regard to our case study, we wish to anticipate turnover rates for the years 2016, 2017 and 2018. However, for models including financial explicative variables, available historical data are only laid from 2006 to 2015. Therefore, prior to predicting turnovers of these models, underlying financial exogenous variables should be estimated. Since the latter are time series, we thought of exploring an extrapolative forecasting method to obtain accurate forecasts. The relevant literature provides a wide range of methods to accommodate several characteristics of available data series. Here, we shall note that we engaged in discussions with the bank for the purpose of approximating the series using exponential smoothing methods because they allow capturing trend and seasonality phenomenon within fixed periods. As decisions effectiveness depends on the accuracy of forecasts, then a suitable model should be selected. For that purpose, we used time series modeling package of SPSS®[8] to identify the adequate exponential smoothing method that should be associated to a particular time series according to its key aspects, which are, priory, diagnosed by SPSS® [8]. Yet so, we used the suggested fitted modelto generate estimates of future amounts of exogenous variables which we injected in the corresponding linear model to forecast the turnover of the associated company, within the three years 2016, 2017 and 2018. In the following, we report details on numerical results emanating from companies raised in both cases displayed in section 4.3.1 and 4.3.3. Recall from the company example in case 1 (see section 4.3.1) that its underlying linear model depends on only one financial variable which is the current asset. Descriptive analysis of related time series shows the existence of trend. To allow forecasting such data, the SPSS recommends the Holt's linear trend model.

TABLE VI
FORECASTING MODEL FOR COMPANY "O"

Model description			
			Model Type
Model ID	*CA*	*Model*	*Holt*
Parameters of the Exponential smoothing method			
Model			*Estimation*
CA		*Alpha (Niveau)*	0.800
		Gamma (Tendance)	4.699^{E}-5
Forecasts			
Model	2016	2017	2018
CA	217392	236488	255585

Further numerical details on the parameters of the model and resulting estimates are depicted in Table VI, which represents the statistical report obtained from SPSS. To explain the relevance, let us illustrate the case where one wishes to predict, the turnover rate of the year 2016. As can be noticed, the predicted current asset is equal to 217392. It turns out that, for company "O" the predicted turnover for the year 2016 is expected to be equal to:

$$TR_{2016} = 8925.553 + 3.697 * 217392 = 812539.8\ TND \quad (6)$$

For the company in the third case, the turnover depends on macroeconomic and financial variables. Again, we used SPSS to analyze each time series related to each financial variable and deduced underlying estimate. Statistical details are outlined in table VII.

TABLE VII
FORECASTING MODEL FOR COMPANY "O"

Model description			
			Model Type
Model ID	*ER*	Model_1	Brown
	LC	Model_2	Simple
	NI	Model_3	Simple
	CA	Model_4	Simple
Parameters of the Exponential smoothing method			
Model			**Estimation**
ER-Model		Alpha (level and trend)	1.000
LC-Model		Alpha (level)	1.000
NI-Model		Alpha (level)	1.000
CA-Model		Alpha (level)	0.781

ISSN 2356-5608

Forecasts			
Model	***2016***	***2017***	***2018***
ER	32277.5	27996.9	23716.3
LC	35702.0	35702.0	35702.0
NI	-6096.9	-11656.8	-13354
CA	81119.6	87009.5	92899.4

As for the predicted values of the macroeconomic factors, they are praised by the worldwide bank (as mentioned in section 4.1). In fact, the turnover of this company for 2018 is equal to 27886.9 TND. To end with, we summarize all predicted turnovers for the whole companies within the three year horizon time in Table VIII.

TABLE VIII
FINAL TURNOVER FORECASTS

Company	2016	2017	2018
A	120 039.1	29 535.2	278 860.9
B	579 949.2	553 028.4	557 090.9
C	1 397 903	979 204.0	560 504.2
D	360 820.3	368 575.2	372 189.3
E	37 164.3	40 273.9	38 655.8
F	31 742.5	29 292.5	28 085.2
G	1 397 903.5	979 204.0	560 504.2
H	60 659.8	61 477.6	61 477.61
I	121 918.8	114 617	11 4954.1
J	45 014.5	49 205.6	53 333.1
K	24 249.9	23 102.6	22 540.8
L	70 568.4	72 506.9	74 445.4
M	52 930.3	60 201.9	64 941.0
N	47 664.1	64 485.8	70 924.0
O	36 098.5	34 704.8	33 311.1
P	812 539.8	883 131.6	953 723.3
Q	76 064.9	78 874.5	81 684.1
R	24 745.6	25 716.5	26 687.4

V. CONCLUSION

This study investigates developing a survey statistical tool for the practice of a loan officer from a Tunisian commercial bank, using multiple linear regressions. Our proposed model can be readily used and is of direct practical relevance. Mainly, it analyzes business credit applications for financing eventual investments through evaluation of the key features of a borrower (productive company) and production of accurate estimates of its future turnover, within three successive years. As such, our model allows for more reasonable credit risk management according to which the decision maker takes measures to prevent from potential problems rather than solving them when they arise. Finally, we state that the improvement of credit risk management should not only restrict to quantitative analysis. Indeed, further research into incorporation of qualitative criteria is necessary because, it translates the real intensions of the decision maker.

REFERENCES

[1] Mačerinskienė, I., Ivaškevičiūtė, L. and Railienė, G., 2014. The Financial Crisis Impact on Credit Risk Management in Commercial Banks. KSI transactions on knowledge society, 7(1).

[2] Basle Committee on Banking Supervision and Bank for International Settlements, 2000. Principles for the management of credit risk. Bank for International Settlements.

[3] Thonabauer, G. and Nösslinger, B. eds., 2004. Guidelines on Credit Risk Management: Rating Models and Validation. Oesterreichische Nationalbank (OeNB).

[4] Saunders, A. and Allen, L., 2002. Credit risk measurement. *New approaches to VaR and other paradigms*.

[5] Van Greuning, H. and Brajovic-Bratanovic, S., 2009. Analyzing banking risk: a framework for assessing corporate governance and risk management. World Bank Publications.

[6] Kaldor, N., 1971. Conflicts in national economic objectives. *The Economic Journal*, *81*(321), pp.1-16.

[7] Foucart, T., 1991. Introduction aux tests statistiques: enseignement assisté par ordinateur. Éditions Technip.

[8] Green, S.B. and Salkind, N.J., 2010. Using SPSS for Windows and Macintosh: Analyzing and understanding data. Prentice Hall Press.

ISSN 2356-5608