THE INTERNATIONAL JOURNAL OF BUSINESS & MANAGEMENT

A Research on Comparison of Regression Models Explaining the Profitability Base on Financial Data

Serkan ETİ
Graduate Student, Department of Quantitative Methods, Marmara University, Istanbul, Turkey
Dr. Mehmet Nuri İNEL
Lecturer, Department of Quantitative Methods, Marmara University, Istanbul, Turkey

Abstract:
Profitability is always a main consideration for businesses. In this study, return on assets is modeled and compared with two different methods. Multiple linear regression and logistic regression are established with financial statements. The two models in the study were evaluated. Their strengths to each other were determined.

Keywords: *Profitability, financial ratios, modeling*

1. Introduction

With industrial revolution, the importance of the technologies developed for serial production began to increase day by day in businesses. Also, businesses began to benefit from technology about reduction of their costs and improve profitability with globalization of the world and they follow technological developments. Share of the technology sector in the economy began to increase each passing day with the importance of technology. These businesses have traded on stock exchanges in the world and Turkey by evaluating (Bayrakdaroğlu & Ege, 2007).

One of the main purpose of the business is to make a profit. Profit is benefit or gain money by provided from the shopping. In terms of economic and financial, profit is the share of entrepreneurs from production. In trade, it is difference between the cost price and sale price (TDK, 2016). Profitability is situation of profit. Profit is a quantity according to the currency, but profitability is rate (Baykara, 1994). Financial analysis contains process of establishment, measurement and interpretation of relationship between items in the financial statements. By examining performance of a business in the current year and past with analysis, it provides the opportunity to make predictions for the future and can be obtained information about planning the future (FinansalYönetim, 2012; Aydın, Şen &Berk, 2012). In this study, using multiple linear regression and logistic regression analyzes, profitability models will be established. As a result, established models will be compared.

2. Financial Structure and Profitability

Financial analysis is used for selection, evaluation and interpretation of financial data with other information in investment and financing decision-making process. Financial analysis also can be used to assess issues such as employee performance, operation productivity and credit policies (Oğuzlar, 2005).

Financial statement analysis is a vital factor in the species diversity of professions including commercial loans, investment management and corporate finance. A primary objective of financial reporting is to measure a company's profitability and financial position correctly with financial statements. Objective of financial reporting is also to obtain cheap capital. In this sense, there are many studies in the literature (Martin, 2008).

Chen and Zhang (2005) has identified that the natural logarithm of total sales has a positive effect on profitability. Akhtar (2005) has determined that the growth rates are important determinants of profitability, leverage and firm size for multinational and national businesses with horizontal section tobit regression analysis. DehuanveJin (2008) has used simple multiple regression in their study. The annual stock returns were used as dependent variable. The total asset turnover, changes in earnings per share, profit margin, return on assetsand return on equity sales were also used as independent variables in this model.İlhan and Bayrakdaroğlu (2009) have used logistic regression analysis in their study. Financial ratios that affect stock returns have been identified. The companies are divided into low and high-income in this analysis. Stephan Dreiseitl and LucilaOhno-Machado (2002) compared established models with logistic regression and neural network. Ali SaitAlbayrak and RamazanAkbulut (2008) examined financial ratios effected profitability of traded industrial and service sectors using multiple linear regression analysis. İlhanEge and Ali Bayrakdaroğlu (2009) determined the success of companies' stock returns using logistic regression analysis. RamazanAkbulut (2011) used multiple linear regression analysis in the manufacturing sector.

3. Comparative Methods

3.1. Multiple Linear Regression

When we examine the relationship between variables, first of methods coming to mind is regression. The relation that causal relationship or linear function between dependent and one or more independent variables examine with regression. Mathematical representation of the linear relationship is,

$$\mu_y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \varepsilon$$

Whereas x_n are n units' independent variables. $\beta_0$is the constant of the equation. β_i $(i = 1,2, \dots, n)$parameters are coefficients of the independent variables. ε is the error term (Armutlulu, 2008).

3.1.1. Stepwise

A variable is added or removed with partial F tests in each step of this method. The algorithm starts with adding independent variable that has the highest correlation with the dependent variable. In the second step, individual partial Fj values are calculated for the remaining variables. Variable has the greatest partial Fj value is added to model. Whereas Fj,

$$F_j = \frac{SS_R(\beta_j|\beta_1, \beta_0)}{MS_E(x_j, x_1)}$$

After new variable is added to the model, first variable status is checked. Depends on the situation, removal of these variable may be required. In next steps, other variables are added to model as second steps. These steps are repeated until there are no significant variables (Armutlulu, 2008).

3.2. Logistic Regression

Linear regression analysis is applied due to the dependent variable is categorical. Reason of this is the least squares method using method for estimation of parameters (coefficients). The least squares method cannot express the real situation. Logistic regression is a type of analysis based on the estimated that dependent variable is probability one of the categories. General logistic regression model is (Muzır&Çağlar, 2009),

$$\ln\left(\frac{E(Y)}{1 - E(Y)}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n + \varepsilon$$

Whereas, E(Y) is the expected value of the categorical variable. There are two procedures for determining of the optimal model in logistic regression analysis: Standard and stepwise. Stepwise procedure also divides as forward and backward methods (Çokluk, 2010).

3.2.1. Forward Likelihood Ratio

The first step in this procedure is started with a model that has an only constant term. Then, according to the test score, variable that has the most important statistics score is added at each step. This process continues until there is no variable that has a significant statistics score. In this process, added variable is tested in each steps. According to results of test, variable is removes or remain. Wald statistic is used for test. Logistic regression begins with an uncertain outcome for estimate coefficient with likelihood ratio. The process that based on improvement of the model iterative continues and finally determined the best results. In process estimation is based on testing and re-estimation. $\beta_i$are coefficients. When iteration solution is searched with Newton-Raphson method,

$$\beta^{(t+1)} = \beta^{(t)} + \{XDiag\,[n_i E(y_i)^{(t)}(1 - E(y_i)^{(t)})X\}^{-1} X(y - (n_i E(y_i)^{(t)})^{(t)})$$

The iteration continues until being $\beta^{(t+1)} - \beta^{(t)} < \varepsilon$(Alan, 2002).

4. Practice

We have used financial data for two models. There are ten companies in technology sector on BIST. We selected nine companies and took the financial data of their 2007-2015 years. Because a company's financial data is not attainable and homogeneous data set is being necessary. Linear regression and logistic regression were compared on same sample. In the next chapter we will touch on these two methods and their results will be discussed.

There are 17 variables in Table 1. One of them is the dependent variable of the study. Spss.20 program were used in the analysis.

	Name of Variable	Formula of Variable
Dependent Variable	Return On Assets	Net Income / Assets
Independent Variables	Sales Size	Ln(Sales)
	Liquidity Ratio	Current Assets - Stocks / Short-Time term Foreign Resources
	Leverage Ratio	Total Foreign Resources / Total of Liabilities (Assets)
	Current Ratio	Current Assets / Short-Time termForeign Resources
	Growth of Asset	Asset at t-time - Asset at (t-1)-time / Asset at (t-1)-time
	Financial Expenses	Financing Expenses / Total Liabilities
	Short Term Leverage Ratio	Short-Term Foreign Resources / Total Liabilities (Assets)
	Long Term Leverage Ratio	Long-Term Foreign Resources / Total Liabilities (Assets)
	Asset Turnover	Net Sales / Total of Assets
	Capital / Total Source	Equity / Total Source
	Average Collection Period Receivables	365x (Short-Term Trade Receivables / Net Sales)
	Debt Ratio	Total Debt / Total Assets
	Net Sales / Assets	Net Sales / Real Assets
	Earning Power in Shareholder's Equity	Net Profit / Equity
	Equity Turnover	Net Sales / Equity
	Business Size	Ln(Assets)

Table 1: Dependent and Independent Variables and Theirs Formulas

4.1. Establishment Model with Multiple Linear Regression

In multiple regression analysis, α=0,05 was taken for selection and significance of variables. All independent variables are involved in the analysis. The variables that identified in this result of analysis are earning power in shareholder's equity, equity turnover, average collection period receivables, long term leverage ratio and liquidity ratio.

Model	R^2	S.E.
1	0,749	0,04517
2	0,867	0,03304
3	0,889	0,033044
4	0,903	0,02866
5	0,912	0,02738

Table 2: R^2 on Each Steps of Stepwise

In Table 2, first model has only independent variable. This independent variable is earning power in shareholder's equity. The explanatory value of this model(R^2) is 0,749. Second model has equity turnover and earning power in shareholder's equity variables. The explanatory value of this model is 0,867. After that, average collection period receivables are added to model. The explanatory value of the third model is 0,889. Then, long term leverage ratio is added to model. The explanatory value of this model is 0,903. Finally, liquidity ratio is added to model. The explanatory value of the last model is 0,912. This value is explanatory value of the established model with multiple linear regression analysis. In multiple linear regression analysis, VIF value of each variable in model must be smaller than 10(Albayrak, 2005). In this case, it indicates that has no linear connection between variables. VIF values of established model is in Table 3. These values are smaller than 10.

Independent Variables	VIF
Liquidity Ratio	1,443
Long Term Leverage Ratio	1,097
Average Collection Period Receivables	1,308
Equity Turnover	1,914
Earning Power in Shareholder's Equity	1,242

Table 3: VIF Values of Independent Variables in Model.

In 5% significance level, sig. value in Anova Table is smaller than 0,05. Therefore, multiple regression coefficients are viable. Additionally, $F_{0,05;75;80}$ is 1,454798 in this significance level. F value of our model is 156,238. Therefore, H_0 is rejected.

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_n = 0$$

In this case, the model is significant in this significance level.

ANOVA TABLE					
	Sum of Squares	**df**	**Mean Square**	**F**	**Sig.**
Regression	0,585	5	0,117	156,238	0
Residual	0,056	75	0,001		
Total	0,642	80			

Table 4: Anova Table of the Multiple Linear Regression Model

Table of coefficients in model is as follows. All significate values are smaller than 0,05. Therefore, it shows that all variables in the model are significant.

Variables	**Unstandardized Coefficients**		**Standardized Coefficients**	**t**	**Sig.**
	β	Std. Error	β		
Constant	,027	,011		2,564	,012
Liquidity Ratio	,003	,001	,102	2,405	,019
Average Collection Period Receivables	,001	,000	-,098	-2,426	,018
Earning Power in Shareholder's Equity	,684	,027	1,011	25,649	,000
Equity Turnover	-,007	,001	-,378	-7,729	,000
Long Term Leverage Ratio	-,280	,073	-,142	-3,843	,000

Table.5: Coefficients of Established Model with Multiple Linear Regression

$$\mu_{\text{Return On Assets}} = 0{,}027 + 0{,}003 * x_{\text{Liquidity Ratio}} + 0{,}001 * x_{\text{AAverage Collection Period Receivables}} + 0{,}684 * x_{\text{Earning Power In Shareholder's Equity}} - 0{,}007 * x_{\text{Equity Turnover}} - 0{,}280 * x_{\text{Long Term Leverage Ratio}}$$

4.2. Establishment Model with Logistic Regression

Return on assets in data sets of study is continuous variable. Dependent variable should be categorical variable in logistic regression analysis. Therefore, variable is changed as

$$lojik = \begin{cases} 0, & \text{Return on assets} \leq 0 \\ 1, & \text{Return on assets} > 0 \end{cases}$$

In new variable, mean of 1 is that business is profitable, mean of 0 is that business is not profitable. Numerators of return on assets and earning power in shareholder's equity are same. This situation leads to full compatibility problems. Therefore, logistic analysis was done with other variables. The logistic regression model was established by five steps. Values of -2 Log likelihood, Cox & Snell R square and Nagelkerke R square in five steps were shown in Table 6

Steps	**-2 Log likelihood**	**Cox & Snell R Square**	**Nagelkerke R Square**
1	61,481	,149	,248
2	56,835	,197	,327
3	51,974	,244	,405
4	45,614	,301	,500
5	47,637	,283	,470

Table 6: R^2 Values in Logistic Regression Analysis

According to Hosmer and Lemeshow test, H_0 hypothesis is accepted. Because Lemeshow test's significate values are bigger than 0,05 in Table 7. $H_0$is "there is no lack of cohesion in the models". In the case, compliance with model and data set is good with 95 percent confidence (Tatlıdil, 2005).

Steps	**Chi-square**	**df**	**Sig.**
1	13,818	8	,087
2	8,406	8	,395
3	6,736	8	,565
4	8,942	8	,347
5	11,778	8	,161

Table 7: Hosmer and Lemeshow Test

Exp(B) value shows the effect of a unit change in the independent variable over probability. The established model with logistic regression analysis contains current ratio, asset turnover and average collection period receivables variables in Table 8

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1	Sales Size	,564	,171	10,932	1	,001	1,758	1,258	2,456
	Constant	-8,562	2,990	8,200	1	,004	,000		
Step 2	Sales Size	,810	,228	12,600	1	,000	2,249	1,437	3,517
	Average Collection Period Receivables	,012	,006	3,809	1	,051	1,012	1,000	1,023
	Constant	-14,074	4,336	10,535	1	,001	,000		
Step 3	Sales Size	1,250	,368	11,546	1	,001	3,492	1,697	7,182
	Current Ratio	,219	,111	3,912	1	,048	1,245	1,002	1,548
	Average Collection Period Receivables	,021	,009	5,670	1	,017	1,021	1,004	1,038
	Constant	-23,687	7,502	9,970	1	,002	,000		
Step 4	Sales Size	,612	,448	1,863	1	,172	1,844	,766	4,442
	Current Ratio	,293	,128	5,228	1	,022	1,340	1,043	1,723
	Asset Turnover	1,861	,907	4,207	1	,040	6,429	1,086	38,051
	Average Collection Period Receivables	,030	,011	7,517	1	,006	1,031	1,009	1,054
	Constant	-16,490	8,139	4,105	1	,043	,000		
Step 5	Current Ratio	,231	,112	4,266	1	,039	1,260	1,012	1,568
	Asset Turnover	2,777	,774	12,865	1	,000	16,066	3,523	73,255
	Average Collection Period Receivables	,029	,011	7,322	1	,007	1,029	1,008	1,051
	Constant	-6,406	2,357	7,390	1	,007	,002		

Table 8: Variables and Coefficient in Steps of Logistic Regression Analysis

5. Conclusions and Recommendations

Profitability is one of the objectives for the business. Nowadays with the increased competition, to understand and increase the profit gain great importance. In this study, multiple linear regression and logistic regression were focused. Profitability is explained with these two models. It has benefited from nine technology companies to explain profitability. These are traded companies in the BİST. Effective items on profitability are determined, using the items of financial statements of these companies. Profitability is explained with five variables in multiple linear regression analysis and with three variables in logistic regression analysis. The methods in two analyzes can provide the opportunity to reduce variable selection. According to Table 9, model has the best performance is multiple linear regression analysis. Generally, when we evaluated in models, the type of analysis should be determined the advantages and disadvantages of them by looking over. It would be more useful. Optimal model is determined according to purpose of analysis on choosing model.

Models	R^2
Multiple Linear Regression	0,912
Logistic Regression	0,47

Table 9: R^2 Values of Models

Consequently, the choice of the developed model is important in determining the method to be used in analysis of profitability. If the profitability situation of company is discussed, logistic regression may be preferred. Multiple linear regression can explain the change in return on assets. When variables investigated how effective on return on assets, multiple linear regression may be preferred. As a practice, technology sector was discussed. As a proposal for future studies, these study may be applied to other sectors.

6. References

i. Akbulut R. (2011). İMKB'deİmalat Sektöründeki İşletmelerde İşletme Sermayesi Yönetiminin Kârlılık Üzerindeki Etkisini ÖlçmeyeYönelik BirAraştırma, İstanbul Üniversitesi İşletme Fakültesi Dergisi, 40(2)

ii. Aktan B. &Bodur B. Oranlar Aracılığı İle Finansal Durumunuzu Nasıl Çözümlersiniz?, Journal of Yasar University, 1(1)

iii. Alan A. (2002). Categorical Data Analysis, Second Edition, New Jersey:Wiley

iv. Albayrak A.S. &Akbulut R. (2008). Kârlılığı Etkileyen Faktörler: İMKB SanayiveHizmetSektörlerindeİşlemGörenİşletmelerÜzerineBirİnceleme, ZKÜ Sosyal Bilimler Dergisi, 4(7)

v. Albayrak A.S. (2005). ÇokluDoğrusalBağlantıHalindeEnKüçükKarelerTekniğininAlternatifiYanlıTahminTeknikleriVeBirUygulama, ZKÜ Sosyal BilimlerDergisi, 1(1)

vi. Armutlulu İ. H. (2008).İşletmelerde Uygulamalı İstatistik, İstanbul:AlfaYayınları
vii. Aydın N., Şen M. &Berk N. (2012). Finansal Yönetim-1, Eskişehir
viii. Bajkowski (1999). Financial Ratio Analysis: Putting The Numbers To Work, The American Association Of Individual Investors
ix. Baykara B. (1994, July).Döner Sermayeli İşletmelerde KârveKârlılığın Anlamı, Mali HukukDergisi
x. Bayrakdaroğlu A., Ege İ. (2007, December). Teknolojik Gelişme, İMKB ve NASDAQ'daİşlem Gören Teknoloji Şirketlerinin Finansal Analizi, 6. Bilgi Ekonomive Yönetim Uluslararası Kongresi, İstanbul, ss. 84-100
xi. Büyükşalvarcı A. (2010). Finansal Oranlar İle Hisse Senedi Getirileri Arasında kiİlişkinin Analizi: İMKB İmalat Sektörü Üzerine Bir Araştırma, Muhasebeve Finansman Dergisi (MUFAD), 48
xii. Çokluk Ö. (2010). Lojistik Regresyon Analizi: KavramveUygulama, Educational Sciences: Theory & Practice, 10(3)
xiii. Dehuan J. & Jin Z. (2008). Firm Performance And Stock Returns: An Empirical Study Of The Top Performing Stocks Listed On Shanghai Stock Exchange, Academy Of Accounting And Financial Studies Journal, 12(1)
xiv. Demir Y. (2001). Hisse Senedi Fiyatını Etkileyen İşletme Düzeyindeki FaktörlerVe Mali Sektör Üzerine İMKB'deBir Uygulama, Süleyman Demirel Üniversitesi İ.İ.B.F. Dergisi,2
xv. Dönmez D., Altuntaş G. &Birgören T. (2011) Girişimci Yönetim Biçimive Finansal PerformansİlişkisiÇanakkaleBölgesi'ndeFaaliyetGösterenOtelİşletmelerindeBirAraştırma, İstanbul Üniversitesi İşletme Fakültesi Dergisi, 40(2)
xvi. Doron N. & Stephen H. P. (2016. March). Financial Statement Analysis of Leverage and How It Informs About Profitability and Price-to-Book Ratios, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=292725
xvii. Ege İ. &Bayrakdaroğlu A. (2009). İMKB ŞirketlerininHisseSenediGetiriBaşarılarınınLojistikRegresyonTekniği İle Analizi, ZKÜ SosyalBilimler Dergisi,5(10)
xviii. Finansal Yönetim (2012). Sermaye Piyasası FaaliyetleriİleriDüzeyLisansıEğitimi
xix. İnel M.N., Eti S., Yıldırım H. (2016). "A Comparison Of Artificial Neural Network And Decision Tree For Profitability In Technology Sector", International Journal of Development Research, 6(7), pp.8417-8421
xx. Karadeniz E. &İskenderoğlu Ö. (2011). İstanbul MenkulKıymetlerBorsası'ndaİşlemGörenTurizmİşletmlerininAktifKârlılığınıEtkileyenDeğişkenlerinAnalizi, Anatolia: Turizm Araştırmaları Dergisi, 22(1)
xxi. Kayalıdere K. (2013). Hisse Senedi Piyasasında Muhasebe BilgilerininRolü: İMKB-Mali Sektör ÜzerineBir Uygulama, İşletme Araştırmaları Dergisi, 5(1)
xxii. Korkmaz Ö. &Karaca S.S. (2014). Üretim İşletmelerinde Firma Kârlılığının Finansal Belirleyicilerive BİST İmalatSanayi Uygulaması, EgeAkademikBakış, 14(1)
xxiii. Koyuncugil A.S. &Özgülbaş N. (2008). İMKB'deİşlem Gören Kobi'lerin GüçlüVeZayıfYönleri: CHAID Karar Ağacı Uygulaması, Dokuz Eylül Üniversitesi İ.İ.B.F. Dergisi,23(1)
xxiv. Long C. &Xinlei Z. (2005). Profitability, Mean Reversion of Leverage Ratios, and Capital Structure Choices
xxv. Martin P. W. (2008) Neural Networks And Their Application To Finance, Business Intelligence Journal, 1(1)
xxvi. Muzır E. &Çağlar N. (2009). Finansal Başarsızlık Tahmin Modellerinin Türkiye' de Geçerliliği: Basit Model Önerileriyle Karşılaştırmalı BirAraştırma, Anadolu Üniversitesi SosyalBilimler Dergisi, 9(2)
xxvii. Oğuzlar A. (2005). Lojistik Regresyon Analizi Yardımıyla Suçlu Profilinin Belirlenmesi, Uludağ Üniversitesi İ.İ.B.F. Dergisi,19(1)
xxviii. Okuyan H.A. (2013). Türkiye'dekiEnBüyük 1000 Sanayi İşletmesinin KârlılıkAnalizi, Business and Economics Research Journal, 4(2)
xxix. Oruç E. (2010). İMKB'deİşlem GörenİşletmelerinHisse Senedi Getirileri İle Çeşitli Finansal Göstergeleri Arasındakiİlişki, Hitit Üniversitesi Sosyal Bilimler Enstitüsü Dergisi,1(2)
xxx. Shumi A. (2004) Australian Multinational and Domestic Corporations Capital Structure Determinants
xxxi. Stephan D. & Lucila O. (2002). Logistic Regression and Artificial Neural Network Classification Models: A Methodology Review, Journal of Biomedical Informatics, 3
xxxii. Tatlıdil H. &Özel M. (2005). Firma DeğerlendirmeÇalışmalarıKonusundaÇokDeğişkenliİstatistikselAnalizeDayalıKararDestekSistemlerininKullanımı, Bankacılık Dergisi. 54
xxxiii. TürkDilKurumu (15 Mayıs 2016). Bilimve Sanat TerimleriSözlüğü, http://www.tdk.gov.tr/