

Discriminant Analysis of Listed Commercial Banks based on Mahalanobis Distance

Xianggang Li

Dongbei University of Finance and Economics, Dalian, 116014, China

Abstract. This paper calculates the total rate of return on equity investment of 20 American listed commercial banks from 2016 to 2021, and conducts discriminant analysis using Mahalanobis distance for commercial banks that outperform and commercial banks that underperform. It is found that the combination of financial indicators of loan/shareholder's equity and loan/deposit can conduct better discriminant analysis on the performance of the return on equity investment of listed commercial banks.

Keywords: Mahalanobis Distance; Listed Commercial Banks; Logistic Regression.

1. Introduction

This paper calculates the total return on investment of 20 commercial banks with the largest market value in the United States from 2016 to 2021. The highest 25% of the total return on investment is regarded as outperform, and the lowest 25% is regarded as underperform. Then use Mahalanobis distance to discriminate analysis on the five financial indicators. The Mahalanobis distance is calculated for each two indicators as a combination. Discriminant analysis of 10 pairs of indicators shows that the combination of loan/deposit and loan/shareholder's equity can best classify outperform and underperform, and the Mahalanobis distance is around 0.3. This paper also uses the combination of loan/deposit and loan/shareholders' equity to conduct a logistic regression on these 10 listed commercial banks. There are few studies on the discrimination of the return rate of listed commercial banks' stocks, especially those using Mahalanobis distance. In this paper, we use Mahalanobis distance and Logistic regression to analyze the factors that affect the return rate of listed banks' investment, providing reference for relevant studies. Because the main goal of enterprises is to create value for shareholders, through the research on financial indicators that affect the return rate of listed commercial banks, this paper finds that different indicators have different effects on the return rate of listed banks' stocks, which provides a theoretical basis for commercial banks to formulate business strategies.

2. Literature Review

2.1 Value Investing

Although modern investment theory argues the market is efficient, Many market anomalies indicate that is not quite true. What is more, it's possible to beat the market by value investing. The value of a company is the discounted sum of its future cash flows. Greenwald, B. C. (2020) argue that Value investors use indicators of fundamental value as guidelines albeit no one knows the future. Yet there is such a thing as fundamental value. Buy when the price falls below the fundamental value and sale when the price is higher than the fundamental value. Value investing is the process of computing the fundamental value of a security, and then comparing this value to the market price and buying if the value is significantly higher than the price to provide a cushion against risk. It is not enough to identify undervalued securities. Greenwald, B. C. (2020) suggest an investor needs to assemble a portfolio of securities with prudent attention to risk, including a diversification strategy.

2.2 Discriminant Analysis

Traditional approach is using low P/E as filters. This will result in reduction of investment universe. Because no trade means no loss, it means missing opportunities. Modern approach is using historical

data and applying modern data analysis techniques (possible because of more computing power and data). More efficient use of filters lead to bigger universe of investment. And it lead to more judicious introduction of new filters and expanded universe of investments e.g. IT and Biotech start ups. This article uses Mahalanobis distance for discriminant analysis. Mahalanobis (1930) found this measure in the context of his studies on racial likeness. It has played a important and fundamental role in data and statistics analysis with multiple measurements, has become an important piece in a statistician's repertoire and has found applications in many fields where numerical taxonomy, classification and statistical pattern recognition problems are encountered - from medical diagnosis to archaeology to remote sensing (McLachlan, G. J., 1999). Distance metric is a key issue in many machine learning algorithms. A distance metric with good quality should identify important features and discriminate relevant and irrelevant features (Xiang, S., Nie, F., & Zhang, C., 2008).

2.3 Logistic Regression

In this paper, we use the logistic expression to model probability of whether a listed commercial bank belongs to outperform or underperform. Logistic regression models are used to study influence of predictor variables on categorical outcomes (Nick, T. G., & Campbell, K. M., 2019). Regression methods have become an integral component of any data analysis concerned with describing the relationship between one or more explanatory variables and a response variable. it is often the case that the outcome variable is discrete, taking on two or more possible values, over the last decade the logistic regression model has become the standard method in many fields of analysis in this situation (Hosmer Jr, D. W., 2013). The existing literature on the impact of bank financial indicators on the stock return of listed commercial banks is relatively scant, Cole, R. A. (2008) found evidence of a significant and positive relationship between future economic growth and banking industry stock returns that is independent of the relationship between general market stock returns and future economic growth.

3. Data Description

The total return on investment in commercial banks is the total return from 2016 to 2021. The purchase price P_0 of the stock is the closing price on January 1, 2016, and the sale price P_1 of the stock is the closing price on December 31, 2021. The dividend D of the stock is the sum of the dividends of the stock from 2016 to 2021. Formula for calculating the total return rate: $\text{Total Return} = (P_1 - P_0 + \text{Intermediate Cash Flows}) / P_0$. The stock price and dividend data are from Yahoo Finance website.

The 20 selected commercial banks are JPMorgan Chase (JPM), Bank of America (BAC), Wells Fargo (WFC), Citigroup (C), Goldman Sachs Group (GS), U.S. Bancorp (USB), PNC Financial (PNC), Capital One (COF), Bank of New York Mellon (BK), Charles Schwab (SCHW), State Street (STT), Discover Financial (DFS), BB&T (BBT), SunTrust Banks (STI), Fifth Third Bancorp, (FITB), M&T Bank (MTB), Northern Trust Corp (NTRS), Regions Financial (RF), KeyCorp (KEY), CIT Group (CIT). The market value of these banks accounts for a large proportion in the listed banks and is representative. The data of financial indicators for discriminant analysis are from the annual reports of listed banks in 2015.

1) Deposit Loan Ratio = Total Deposits / Total Loans. From the perspective of commercial banks, the higher the deposit loan ratio is, the higher the benefits will be received, and the best use of the maximum value of funds will be achieved. If a bank has a lot of loans and few deposits, the bank will not be profitable, which means that the bank bears huge costs in the course of operation and has little income. It is an important indicator to measure bank liquidity.

2) Return On Net Assets = Net Profits Obtained by Asset Owners / Total Net Assets Owned $\times 100\%$. The greater the value of the return on equity, the greater the ability of the enterprise to obtain income from its net assets. For investors, the higher this rate of return is, the greater the profits investors will get after investing in assets. In some countries, this indicator is one of the information that listed

commercial banks must disclose to the public, and it is also an important basis for determining whether listed commercial banks can obtain rights issue.

3) Revenue Growth. Revenue Growth is the growth rate of total revenue of commercial banks compared with the previous year. In the stock valuation model, the growth of the company's income is an important factor. The faster the company's income grows, the greater the stock value.

4) Loans/Shareholders' Equity. Loans/Shareholders' Equity is an indicator reflecting the operating capacity of a bank. The larger this indicator is, the stronger the bank's operating capacity is, which can create greater returns for investors. In addition, excessive leverage of commercial banks has played an important role in triggering the outbreak of financial crisis.

5) Net Interest Income Growth. The growth of net interest income can reflect the profitability and growth of commercial banks, which can affect the income of banks and the return on investment of investors. The size of interest income also affects the income structure of commercial banks, which in turn affects the return rate of investors.

4. Conclusion

The total return of these banks from 2016 to 2021 are shown in Table 1. The banks of Outperform are RF, BAC, AXP, SCHW and FITB. The commercial banks of underperform are WFC, C, MTB, USB and BK. The Mahalanobis distance obtained from Discriminant analysis of the combination of the above five indicators is shown in Table 2. The largest Mahalanobis distance is loan/shareholder's equity and loan/deposit, which shows that the combination of these two indicators can make the best distinction between outperform and underperform of total stock return of listed commercial banks. Logistic regression the ten listed banks with loans/shareholders' equity and loans/deposits, The results are shown in Table 3. The data in the table reflects the probability of out performance or under performance of each commercial bank. RF, BAC, AXP, SCHW, FITB. The corresponding probability is the probability that the return rate of these banks' stocks is out of performance. The probability of WFC, C, MTB, USB and BK is the probability that the total stock return of these listed banks is underperform.

Tables

Table 1. Total Return on Equity Investment of Commercial Banks

Commercial bank	GS	USB	AXP	SCHW	PNC
Total return	151.83%	61.43%	222.58%	241.83%	156.48%
Commercial bank	NTRS	FITB	KEY	CFG	RF
Total return	112.84%	205.57%	136.65%	153.22%	201.97%
Commercial bank	BK	COF	STT	DFS	MTB
Total return	77.86%	136.35%	85.70%	172.64%	59.68%
Commercial bank	JPM	BAC	WFC	C	CMA
Total return	194.64%	236.78%	12.31%	62.78%	186.47%

The deposit loan ratio and loan/shareholder's equity are indicators reflecting the liquidity of commercial banks. The larger these two indicators are, the lower the liquidity of commercial banks is, which increases the risk of commercial banks in the business process. According to the CAPM model, commercial banks have therefore generated higher excess stock returns. Therefore, these two indicators perform better in Discriminant Analysis. The higher these two indicators are, the stronger the operating capacity of commercial banks is. More deposits are used for lending, generating more income. As reflected in the stock market, the higher the return on investment of stocks is. The second largest financial indicator portfolio of Mahalanobis distance is Revenue Growth&Loans/Shareholders' Equity, which shows that in 2015, American commercial banks with faster income growth will have higher returns on equity investment in the next five years. Investment suggestion: According to Discriminant analysis and Logistic region analysis, listed banks with large deposit loan

ratio and large loan/shareholder's equity can be selected as investment objects for a five-year investment. In this way, it is more likely to obtain the excess return rate of shares higher than that of other listed banks.

Table 2. The Result of Discriminant Analysis

Combination of financial indicators	Mahalanobis distance
Total deposits/total loans&Return on net assets	0.02585
Return on net assets&Revenue growth	0.00036
Revenue growth&Loans/shareholders' equity	0.27995
Loans/shareholders' equity&Net interest income growth	0.02184
Net interest income growth&Revenue growth	0.07809
Total deposits/total loans&Revenue growth	0.14363
Return on net assets&Total deposits/total loans	0.02585
Net interest income growth&Return on net assets	0.03532
Loans/shareholders' equity&Total deposits/total loans	0.30372
Net interest income growth&Total deposits/total loans	0.00036

Table 3. The Result of Logistic Regression

OUTPERFORM/UNDERPERFORM	Commercial bank	Probability
OUTPERFORM	RF	45.33%
OUTPERFORM	BAC	75.81%
OUTPERFORM	AXP	98.67%
OUTPERFORM	SCHW	49.74%
OUTPERFORM	FITB	24.95%
UNDERPERFORM	WFC	64.20%
UNDERPERFORM	C	54.13%
UNDERPERFORM	MTB	52.51%
UNDERPERFORM	USB	73.82%
UNDERPERFORM	BK	49.84%

References

- [1] P C Mahalanobis, On tests and measures of group divergence, Journal of the Asiatic Society of Bengal, 26, 541-588, 1930.
- [2] McLachlan, G. J. (1999). Mahalanobis distance. Resonance, 4(6), 20-26.
- [3] Xiang, S., Nie, F., & Zhang, C. (2008). Learning a Mahalanobis distance metric for data clustering and classification. Pattern recognition, 41(12), 3600-3612.
- [4] Nick, T. G., & Campbell, K. M. (2007). Logistic regression. Topics in biostatistics, 273-301.
- [5] Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied logistic regression (Vol. 398). John Wiley & Sons.
- [6] Cole, R. A., Moshirian, F., & Wu, Q. (2008). Bank stock returns and economic growth. Journal of Banking & Finance, 32(6), 995-1007.
- [7] Greenwald, B. C., Kahn, J., Bellissimo, E., Cooper, M. A., & Santos, T. (2020). Value investing: from Graham to Buffett and beyond. John Wiley & Sons.