Financial Risk Disclosure Return Premium: A Topic Modeling Approach

1st Beichen Zhang
School of Business
Stevens Institute of Technology
Hoboken, United States
bzhang61@stevens.edu

2nd Steve Yang
School of Business
Stevens Institute of Technology
Hoboken, United States
syang14@stevens.edu

2024 IEEE Symposium on Computational Intelligence for Financial Engineering and Economics (CIFEr) | 979-8-3503-5483-6/24/$31.00 USG | DOI: 10.1109/CIFER62890.2024.10772758

***Abstract*—We examine the risk factors disclosed in the 10-K financial statement section 1A across 9 years with over 500 hundred companies. We propose a financial disclosure risk factor to extend the Fama-French 3 factor model and Fama-MacBeth cross-section regression. Using the risk factors data from 2015 to 2023, we find the average risk-return premium across nine sectors is significant after controlling for other risk factors from the Fama-French 3-factor model. The premium is measured by monthly return series on risky-minus-less risky stocks or by the coefficient of stock risk factor estimated from cross-section Fama-MacBeth regressions. These text risk factors can potentially be used to construct portfolios that can generate significant returns across different sectors.**

***Index Terms*—risk factors, financial statement, Fama-French 3 factor model, Fama-McBeth cross-section regression**

I. Introduction

Financial risk is frequently associated with business and is crucial for corporations. It is the risk that the company loses the ability to meet its obligations to pay back its debts, resulting in a loss for potential investors. Qualitative data contains information that is not fully reflected in quantitative data. Analyzing unrevealed information such as 10-K financial disclosures can help investors and analysts better understand firms' performances, including financial risks. This paper aims to study the effect of financial risk factors on expected return as a characteristic, focusing on the U.S. market. Financial common risk factors in returns on stock and bonds have been previously studied by Fama and French [8]. They construct the three-factor asset pricing model, identify three common risk factors on stocks: RM, SMB, and HML, represent the market excess returns, the size (small-minus-big, market capitalization) and value (high-minus-low, book-to-market) factor returns respectively. Bank finds that after 1993, SMB and HML risk premium contributions are strongly related to sentiment [3]. Many other researchers, such as Amihud, use the three common risk factors as the basis to construct return premiums [1]. In this paper, we examine the risk factors disclosed in 10-K financial statement section 1A across 9 years, in a sample of over 500 companies. We propose a financial disclosure risk factor to extend the Fama-French 3 factor model and Fama-MacBeth cross-section regression. Our findings reveal that the financial risk return premium is not only significant but also varies at different levels after controlling for common risk factors and firm characteristics at the sectors level. Our Fama-French regression results suggest that shareholder interest risk is one of the most critical risks for companies of varying sizes. Additionally, the potential significance of financial condition risk emerges as a critical factor in the market, assuming all other factors remain constant. Furthermore, the Fama-MacBeth regression suggests that reliance on a few large customers is one of the most critical risks for companies of varying values. This underscores the crucial role of financial risk in determining expected returns.

We describe two procedures for estimating the risk-return premium. One measure of risk-return premium is the differential return between the riskiest and less risky stock quintile portfolios, denoted as $RFactor$ (1, 2, and 3), generated from three risk disclosure factors. We find that across sectors, the average return-weighted portfolio return $RFactor1$, $RFactor2$, and $RFactor3$ are 0.8%, -1.4%, and -0.7%, respectively. The risk-adjusted risk premium, $\alpha_{RFactor1}$, $\alpha_{RFactor2}$ and $\alpha_{RFactor3}$ are 1.2%, -1.8% and 0.1% respectively, after controlling for three sector risk factors. Financial risk is significantly priced across 2015 to 2023 within the U.S.

The second measure of risk-return premium is the mean coefficient (denoted $b1$, $b2$, and $b3$) from cross-section regressions [7] of stock returns, controlling for firm characteristics (size, value, year, sector, and volatility). The mean $b1$, $b2$, and $b3$ are calculated for each portfolio and averaged across sectors. Examining the 10-K financial disclosure statement and risk factors section from 2015 to 2023, we find the average risk-return premium across eight sectors is significant after controlling for other common risk factors from the Fama-French 3-factor model.

The risk-return premium, a key focus of our study, can be measured by monthly return series on risky-minus-less risky stocks or by the coefficient of stock risk factor estimated from cross-section Fama-MacBeth regressions. Our research demonstrates that the risk-return premiums are robust across the measures of the premium employed in this study. This has significant implications in financial analyses and investment strategies, as it provides concrete evidence of the value of considering financial disclosure risk factors in decision-making process.

The paper proceeds as follows. In the section on data and methodology, we thoroughly explain the construction of our sample, portfolios, and variables. Our rigorous approach ensures the reliability and validity of our findings, providing a solid foundation for financial analyses and decisions. In the section design and experiment, we propose a financial disclosure risk factor to extend the Fama-French 3-factor model and Fama-MacBeth cross-section regression. Furthermore, we describe two procedures for estimating the risk-return premium and provide estimates of the financial risk-return premium. Following are the results and concluding remarks.

II. Data and methodology

A. Sample construction

The sample focuses on the U.S. market 10-K financial statement, section 1A risk factors with nine years of data, from 2015 to 2023. Data are obtained from Calcbench Data Query on the NYSE/AMEX markets. This dataset, including company (ticker, CIK), fiscal year, sector, and risk factors, disclosure from the 10-K document for textual analysis. Additionally, it contains shares outstanding, end-of-the-period stock price, assets, and liability. All the stock's daily and monthly adjusted closing prices are obtained from Yahoo Finance for calculating stock returns. We restrict the sample to the company with the valid ticker and CIK number.

Our final sample consists of 49,213 observations over the fiscal years between 2015 and 2023, followed by the classification of 8 sectors within each year. The eight sectors are material, industrial, consumer discretionary, consumer staple, health care, financial, information technology, and communication services.

B. Measuring risk

To prepare the risk factors from the 10-K financial statements, we use the Latent Dirichlet allocation (LDA) method to extract the ten topics from the overall risk factors textual, which can present the overall document focus. This unsupervised learning total probability generative model (LDA) eliminates subjective human data selection and has frequently been utilized by researchers for identifying abstract topics within collections of documents, such as news articles and other texts. Researcher Bybee has further derived an interpretation of the estimated systematic risk factors from news text. [5] Researcher Wang applied the NLP-based model to financial documents and communications, demonstrating its effectiveness in detecting, identifying, and predicting financial risks [16] [17]. We then label each topic to correspond to risk factors using Huang's multi-label text classification algorithm for risk factors in SEC form 10-K [10]. Ten topics are presented as the following:

TABLE I
10 TOP TOPICS EXTRACTED FROM SAMPLE STATEMENTS

Topic	Risk factors
1	Funding risk
2	New product introduction risk
3	Intellectual property risk
4	Input prices risk
5	Rely on few large customers
6	Shareholder's interest risk
7	Restructuring risk
8	New product introduction risk
9	Funding risk
10	Financial condition risk

Each of these topics represents a risk factor; thus, ten risk factors were extracted from the text document. Moreover, the LDA model also presents the percentage of topics in each document. We then choose to use the Principle Component Analysis method to reduce the dimension of these risk factors without losing the most valuable part of all of the variables. It is a linear dimensional reduction method that uses the Singular Value Decomposition of the data to project it into a lower-dimensional space. One can see this as a feature extraction process. It combines all ten risk factors (variables) in a specific way, then drops the 'least important' variables while keeping the characteristics of all. New variables are formed in this process as a combination of the old variables. Each of the new variables, labeled as $PC1$, $PC2$, and so on, is independent of one another.

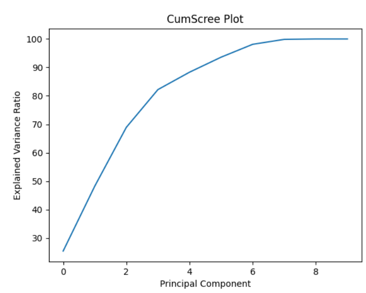

Fig. 1. Cumulative scree plot

The cumulative scree plots resulted in us using $PC1$, $PC2$, and $PC3$ as risk measurements, which capture 85% of the financial risk among all risk factors from overall documents. Thus, each company (document) is correlated with one $PC1$ value, one $PC2$ value, and one $PC3$ value. The final sample includes over 500 stocks traded in 8 sectors over nine years.

C. Formation of risk portfolios

After dividing all stocks by their fiscal year and sectors, we evaluate the effect of risk on stock returns by forming risk-based portfolios of stocks in each sector. Noting financial risk and market volatility are positively correlated [4]. To control the effect of volatility on stocks' expected return, we first sort stocks into three portfolios by their standard deviation of daily return during their fiscal year. We then sort stocks within each volatility group on their $PC1$ value and aggregate them into quintile portfolios. This provides ranking by risk over (controlling) volatility. We then obtain 15 (3x5) portfolios with equivalent stocks in each sector. As a result, three portfolios consist of high-risk and low-risk quintiles across three levels of volatility. Next, we calculate the average return for each of the 15 portfolios. The risk premium is present by the return difference between the three riskiest and three least risky portfolios.

This methodology resembles HML factor (high-minus-low, book-to-market stocks) formation from the Fama-French three-factor model [9]. They construct their portfolios by sorting stocks based on their market capitalization (size) and within each sized portfolio, then sorting stocks by their book-to-market ratio. This procedure captures the book-to-market effect on stock returns while controlling for the size effect. Amihud and his co-author also used this methodology to generate their illiquidity premium in their research [1].

Finally, we construct the zero-investment risk premium factor $RFactor1$, the return on the riskiest-minus-less risky quintile portfolios (across the volatility portfolios). Since we use $PC1$, $PC2$, and $PC3$ as risk measurements, this portfolio formation process is repeated three times. Hence, three risk premium factors, $RFactor1$, $RFactor2$, and $RFactor3$, are constructed.

III. Design of Experiments

A. Year and sector risk factors

The $RFactor$ excess return is obtained as the intercept $\alpha_{RFactor}$ from a regression of $RFactor$ on the Fama-French 3 factors model's common risk factors [9]. The factors in Fama and French are RM, HML (high-minus-low), and SMB (small-minus-big). Factor RM presents the market excess returns, HML presents the value (book-to-market) factor returns, and SMB represents the size factor returns. Since we divide the dataset by sectors before market capitalization and book-to-market ratio, we use sector indexes for each factor. The empirical models are:

$$\begin{aligned} RFactor1_{sec_t} =& \alpha_{RFactor1} + \beta_1 RM_{sec_t} + \beta_2 SMB_{sec_t} + \\ & \beta_3 HML_{sec_t} + \epsilon_t \end{aligned} \tag{1}$$

$$\begin{aligned} RFactor2_{sec_t} =& \alpha_{RFactor2} + \beta_1 RM_{sec_t} + \beta_2 SMB_{sec_t} + \\ & \beta_3 HML_{sec_t} + \epsilon_t \end{aligned} \tag{2}$$

$$\begin{aligned} RFactor3_{sec_t} =& \alpha_{RFactor3} + \beta_1 RM_{sec_t} + \beta_2 SMB_{sec_t} + \\ & \beta_3 HML_{sec_t} + \epsilon_t \end{aligned} \tag{3}$$

The factors are sector (sec) at time t. The market factor (RM_{sec_t}) is the average return on portfolios with the same sector minus the risk-free rate of the corresponding year. The sectors SMB_{sec} and HML_{sec} factors are constructed following the procedures of Fama and French [9]. We first divide the stocks (company) by their fiscal year (9 years) and each fiscal year into eight sectors. These eight sectors are the same as the previously mentioned sample construction part. Then, we divide each sector into two equal-size groups (small and big) based on the median market value (capitalization), where the equity market capitalization is for the end of the fiscal year. Finally, we divide each sized group into three portfolios based on the value of the book-to-market ratio of equity (B/M). The three book-to-market groupings use $35th$ and $70th$ percentiles as breakpoints for all stocks within a sized portfolio. These three groups are growth, neutral, and value companies based on small to large book-to-market ratios. Therefore, we obtain 6 (2x3) portfolios ranked on size and value for each sector, each fiscal year, giving a total of 432 (6 x 8 x 9) portfolios. We calculate the average return for each of the 432 portfolios. The sector's SMB_t factor is the average return difference between the three small and three big portfolios within a sector for each fiscal year. The sector's HML_t factor is the difference between the two value portfolios and the two growth portfolio's average return within each sector.

B. The risk return premium, estimated from cross-sectional regressions

As an alternative estimation of the risk premium, we use individual stock returns and employ the Fama-MacBeth cross-sectional regression approach [7]. First, we perform individual regressions for each stock across different years and sectors. The stock's return serves as the dependent variable, while SMB_t, HML_t, $logSTDEV_t$ as independent variables in the regression. The resulting coefficients (β) indicate the sensitivity of each stock's returns to these factors. We then aggregate these sensitivities by regressing the stock returns on the estimated factor βs across all stocks within each sector and year. Our goal is to find out how much return the market demands for each unit of risk associated with the factors. We estimate the risk premia (average returns) associated with each factors. These premia tell us what the market compensates investors for bearing the risk associated with each factor during the period, accounting for variations over time.

$$\begin{aligned} R_t =& b0 + b1 * RFactor1_t + b2 * RFactor2_t + \\ & b3 * RFactor3_t + b4 * SMB_t + b5 * HML_t + \\ & b6 * logSTDEV_t + b7 * sector_t + b8 * year_t + \epsilon_t \end{aligned} \tag{4}$$

$RFactor1_t$, $RFactor2_t$, and $RFactor3_t$ are financial risk factors regressed on the Fama-French 3-factor model. SMB_t and HML_t (book-to-market) are size and value factors following the Fama and French 3-factor model, obtained according to the procedure in the Fama and French 3-factor model [8]. $logSTDEV_t$ is the logged stock return standard deviation, calculated from the monthly portfolio returns of each sector. Taking the log will normalize the variable's distribution.

Table II presents the original data's sample statistics along with the cross-sectional regression's independent variables.

Table III represents the correlation of the independent variables in this cross-sectional regression model. The cross-section Fama-MacBeth regression model is estimated for each sector per fiscal year by the return-weighted method proposed by Asparouhova, Bessembinder, and Kalcheva [2]. This method is constructed to correct the potential bias from microstructure noise.

TABLE II
DESCRIPTIVE STATISTICS OF VARIABLES

Variable	Obs.	Mean	STD	Min.	Max.
market cap	4202	2.445e+11	7.889e+12	6496	3.968e+14
BTM	4202	624.714	3.453e+04	-9.603	2.198e+06
$PC1$	49213	2.483e-17	2.260	-4.676	5.642
$PC2$	49213	1.527e-16	2.139	-3.981	7.106
$PC3$	49213	-8.728e-17	2.037	-5.170	6.392
$RFactor1$	72	0.008	0.041	-0.193	0.174
$RFactor2$	72	-0.014	0.073	-0.470	0.181
$RFactor2$	72	-0.007	0.071	-0.487	0.230
SMB	72	-0.001	0.028	-0.065	0.124
HML	72	-0.012	0.034	-0.097	0.106
$\log STDEV$	72	-4.663	0.552	-5.827	-3.568
$sector$	72	5.500	2.307	2	9
$year$	72	2019	2.600	2015	2023

TABLE III
CORRELATION OF INDEPENDENT VARIABLES IN CROSS-SECTIONAL REGRESSION [a]

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1)	1							
(2)	0.231	1						
(3)	-0.096	-0.094	1					
(4)	0.211	-0.080	0.077	1				
(5)	0.182	0.093	0.096	0.029	1			
(6)	0.011	0.099	0.166	-0.015	-0.044	1		
(7)	-0.075	-0.035	-0.140	-0.022	-0.100	-0.525	0	
(8)	-0.179	-0.078	-0.137	0.072	0.056	0.102	0	0

[a] The variables presented in the table are defined as follows: 1 - RFactor1; 2 -RFactor2; 3 -RFactor3; 4 -SMB; 5 -HML; 6 -logSTDEV; 7 -sector; 8 -year

IV. RESULT ANALYSIS

A. The risk return premium

In this section, we present the risk disclosure premium results. First, we examine the return premium associated with the risk disclosure factors using the Fama-French 3-factor model. Then, we explain how the different disclosure risk factors contribute to such a risk-return premium. Finally, we use cross-sectional analysis to show the existence of such a risk-return premium.

Table IV presents the regression statistics on the risk-return premiums ($RFactor1$, $RFactor2$, and $RFactor3$) and the standard risk-adjusted risk-return premium – the intercept ($\alpha_{RFactor1}$, $\alpha_{RFactor2}$, and $\alpha_{RFactor3}$) estimated from the empirical model. The three risk-return premiums are the dependent variable of these regressions, while the common market access return, size and value factors are the independent variables. The results from the first regression are shown in the top section of this table. The risk-adjusted risk premium $\alpha_{RFactor1}$ is positive and statistically significant with a level of 0.05. For the other coefficients, we use the significant level of $p - value$ of 0.1. The SMB (size) factor is positive and significant, with a p-value of 0.091, indicating that $RFactor1$ reacts more to changes in market capitalization. Moreover, the adjusted R square value is 0.04, indicating that these variables can explain 4% of the financial risk in the textual document.

TABLE IV
REGRESSION RESULT OF $RFactor$'S ON FAMA-FRENCH 3 FACTOR MODEL

variable	coefficients	p-value [a]	t-statistic	SE
$\alpha_{RFactor1}$	0.012	0.023**	2.318	0.005
RM	0.148	0.539	0.617	0.240
SMB	0.287	0.091*	1.712	0.168
HML	0.184	0.215	1.252	0.147
R^2	0.081			
Adj.R^2	0.040			
F-Statistic	1.985			
$\alpha_{RFactor2}$	-0.018	0.073*	-1.819	0.010
RM	-0.720	0.103*	-1.655	0.435
SMB	-0.171	0.577	-0.560	0.304
HML	0.338	0.208	1.270	0.266
R^2	0.054			
Adj.R^2	0.012			
F-Statistic	1.284			
$\alpha_{RFactor3}$	0.001	0.890	0.138	0.009
RM	0.641	0.134	1.515	0.423
SMB	0.148	0.619	0.500	0.296
HML	0.078	0.765	0.300	0.259
R^2	0.047			
Adj.R^2	0.005			
F-Statistic	1.119			
Observations	72			

[a] ** and * denote significance at the 5% and 10% levels respectively

The middle section in table IV shows the result from the second regression model. The risk premium is economically meaningful. The risk-adjusted risk premium $\alpha_{RFactor2}$ is -0.018 and statistically significant on the significant level of 10%. The market access return, RM, is statistically significant at the same level. However, $RFactor2$ shows low sensitivity to the size and value factor, as shown by the insignificance of these two parameters. The adjusted R square value is 1.2%, implying that 1.2% of the financial risk in the textual document can be explained by these variables.

The last section in table IV shows the result from the third regression model, $RFactor3$. The risk-adjusted risk premium $\alpha_{RFactor3}$ 0.1% is statistically insignificant with a p-value of 0.89. The market access return, size, and value factors, namely RM, SMB, and HML, are all statistically insignificant with very high p-values. The overall adjusted R square value of 0.5% implies that these variables can explain only 0.5% of the financial risk in the textual document. The regression using $RFactor3$ has a weak overall significant level.

B. Risk disclosure factors

We used $PC1$, $PC2$, and $PC3$ as our risk disclosure factors to divide our portfolios and construct $RFactor1$, $RFactor2$, and $RFactor3$, respectively. The significant level of variables are weak in the regression result of $RFactor3$ on the Fama-French three factors model, whereas the regression of $RFactor1$ and $RFactor2$ presents a more significant result. We present the correlations between the topics and the principal components. The loading score for $PC1$, $PC2$, and $PC3$ as shown below

TABLE V
LOADING SCORES FOR $PC1$, $PC2$ AND $PC3$ [a]

PC1		PC2		PC3	
topic	loading score	topic	loading score	topic	loading score
6	0.640	10	0.798	5	0.707
7	-0.604	6	-0.454	7	-0.613
5	-0.419	5	-0.385	10	0.279
10	0.218	7	0.073	6	-0.209
8	0.040	4	0.048	4	-0.048

[a] Topic number is matched with risk factors in table I

These PCs share some of the same components but with different magnitudes and directions of correlation with the topics. Comparing with the 10-risk factor generated previously (table I), we find that $PC1$ has a relatively strong positive correlation with shareholder's interest risk while negatively related to the restructuring risk (topics 6 and 7). $PC2$ has a strong positive correlation with the financial condition risk while having a moderate negative relationship with both shareholder's interest risk and the risk of relying on a few large customers (topics 10, 6, and 5). Lastly, $PC3$ has a strong positive correlation with the risk of relying on a few large customers yet has a relatively strong negative relationship with the restructuring risk (topics 5 and 7). By combining the interpretation of the loading scores with the regression results, we observe the following connections: Both Risk Factor 1 and the SMB factor demonstrate significance in the first empirical model, with shareholder's interest risk shows a positive relationship and contributing the most to the principal component used to segment the portfolios. This suggests that shareholder's interest risk is one of the most critical risks for companies of varying sizes, assuming all other factors remain constant. In the second empirical model, both the RM factor and Risk Factor 2 show significance. The portfolios in this model are segmented based on the principal component exhibiting the strongest positive correlation with Financial Condition risk. This underscores the potential significance of Financial Condition risk as a critical factor in the market, assuming the constancy of all other factors.

C. The exposure of the financial risk return premium to risk disclosure factor risk

We focus on $RFactor1$ and $RFactor2$ regression models as they provide a higher significance level of the regression. Both risk factors are regressed on three sector Fama-French risk factors, each producing three slope coefficients, β_1 *to* β_3. The most consistent relation is $RFactor2$ and market returns (RM_{sec}). Both regression results show that the coefficients β_1 of the sector excess returns, RM_{sec}, are optimistic. That is, the risk premium rises in times with the market conditions.

Moreover, Fig. 2 presents the relationship between $RFactor1$, $RFactor2$ and RM_{sec} over the time period from 2015 to 2023. These three-time series mostly move in the same direction. The overall correlation between $RFactor1$ and $RFactor2$ to RM_{sec} are 0.139771 and -0.163847, respectively. The result shows a relatively weak relation between the risk premium and the sector market portfolio return.

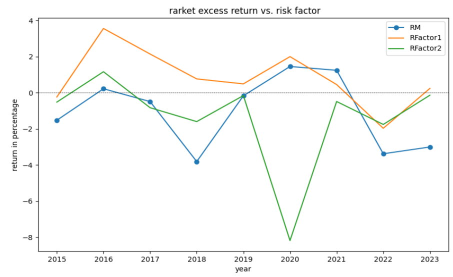

Fig. 2. Correlation of $RFactor1$, $RFactor2$ and RM_{sec}

The coefficient of the SMB_{sec} factor in the regression result (Table IV) is positive and significant for $RFactor1$ because firm size is associated with shareholder's interest risk and the size premium is partly due to this risk premium. The coefficient of the HML_{sec} factor in both regressions is positive and statistically insignificant, indicating that the relation between both shareholder's interest risk premium and financial conditional risk premium to the value premium is weak.

D. Result of risk return premium, estimated from cross-sectional regressions

Using the cross-sectional regression model, we calculate b_k, where $k = 2, 3, \ldots, 9$, the monthly coefficients for each sector. The analysis includes eight sectors per fiscal year for nine fiscal years. The statistics of the regression are presented as follows:

TABLE VI
FAMA-MACBETH CROSS-SECTIONAL REGRESSION ESTIMATION RESULT

variable	coefficients	p-value[a]	t-statistics	SE
intercept	0.0039	0.784	0.275	0.014
$RFactor1$	0.0367	0.401	0.846	0.043
$RFactor2$	-0.0329	0.167	-1.399	0.024
$RFactor3$	0.0503	0.040**	2.092	0.024
SMB	-0.0043	0.944	-0.071	0.060
HML	0.1504	0.004***	3.027	0.050
$\log STDEV$	-0.0021	0.492	-0.691	0.003
R^2	0.224			
Adj.R^2	0.152			
F-Statistic	3.129			
Observations	72			

[a] ** and * denote significance at the 5% and 10% levels respectively

We run this cross-sectional regression across all stocks, with the stock returns as the dependent variable. We focus on the coefficients of $RFactor1$, $RFactor2$, and $RFactor3$, the monthly risk premium after controlling for other firm characteristics. The coefficients of $RFactor1$ and $RFactor2$ are insignificant with large p-values. However, the coefficient of $RFactor3$ is positive and significant (with p-value of 0.04). In other words, for each additional unit of $RFactor3$'s risk taken, the market pays an extra 5.03% in return.

As for the other coefficients, the coefficient of HML (book-to-market ratio, value factor) is positive 0.1504 and significant (p-value=0.004). This means that for every additional unit of value risk taken, the market pays an extra 15.04% in return, indicating a higher reward for value stocks. The coefficient of the size factor SMB and return volatility $\log STDEV$ are insignificant.

The overall regression's adjusted R-square value is 0.152, indicates that approximately 15.2% of the variability in stock return can be explained by the independent variables included in the regression model. By combining the interpretation of the loading scores with the cross section regression results, we observe that $RFactor3$ and the HML factor demonstrate significance in this model. There are two major influencers of $RFactor3$: Topic 5 and Topic 7. This suggests that relying on few large customers and restructuring risks are the critical risks for companies of varying values, assuming all other factors remain constant.

V. Conclusions

In this research, we employ various methods to calculate the financial risk premium, using portfolio-based estimates ($RFactor1$, $RFactor2$, and $RFactor3$) and estimates based on cross-section regressions ($b1$, $b2$, and $b3$). The time-series regression we performed in table IV, where the $RFactor1$, $RFactor2$, and $RFactor3$ are the dependent variable for each, and the market excess return, size, and value factors are the independent variables. We can tell the sensitivity of each portfolio return are to these factors by its beta coefficients. We find that portfolios with high financial risk stocks generate higher risk-adjusted returns than portfolios with less financial risk stocks. The portfolio risk-adjusted annual risk-return premium $\alpha_{RFactor1}$ is 1.2%, and $\alpha_{RFactor2}$ is -1.8%, statistically significant with various levels, after controlling for the three general risk factors (market, size, and value factors) in the sector. This positive and significant risk-adjusted risk-return premium is scaled by the financial risk factor $RFactor$'s, defined as the difference between the high and low-risk portfolios across each volatility level for each sector.

As we take the beta coefficients calculated for each portfolio from the time-series regression, and use them in the new regression. We find the risk premium $b3$ significant when it is measured by cross-sectional Fama-MacBeth regression of individual stock returns on stock financial risk while controlling for other firm characteristics. This methodology resembles Amihud et al. (2013) illiquidity risk premium [1]. In conclusion, we document a significant return premium for the risks disclosed in the 10-K financial statement section 1A across nine years. We conclude that the textual risk factors can generate return premium across different sectors.

References

[1] Amihud, Y., Hameed, A., Kang, W., Zhang, H., The Illiquidity Premium: International Evidence. SSRN Electronic Journal. doi: 10.2139/ssrn.2207810. 2013.

[2] Asparouhova, E., Bessembinder, H., Kalcheva, I., Liquidity Biases in Asset Pricing Tests, Journal of Financial Economics 96, 215-237.2010.

[3] Bank, M., Insam, F.. Risk Premium Contributions of the Fama and French Mimicking Factors. Finance Research Letters, 29, 347–356. doi: 10.1016/j.frl.2018.08.017.

[4] Brownlees, C. T., Engle, R. F.. Volatility, Correlation and Tails for Systemic Risk Measurement. SSRN Electronic Journal. doi: 10.2139/ssrn.1611229. 2011.

[5] Bybee, L.,Kelly, B., Su, Y., Narrative Asset Pricing: Interpretable Systematic Risk Factors from News Text. Johns Hopkins Carey Business School Research Paper No. 21-09, SSRN: 3895277. 2022

[6] Daniel, K., Titman, S. Evidence on the Characteristics of Cross Sectional Variation in Stock Returns. doi: 10.3386/w5604. 1996.

[7] Fama, E. F., Macbeth, J. D. Risk, Return, and Equilibrium: Empirical Tests. Journal of Political Economy, 81(3), 607–636. doi: 10.1086/260061. 1973.

[8] Fama, E. F., French, K. R. The Cross-Section of Expected Stock Returns. The Journal of Finance, 47(2), 427–465. doi: 10.1111/j.1540-6261.1992.tb04398.1992.

[9] Fama, E. F., French, K. R. Common Risk Factors in the Returns on Stocks and Bonds. Journal of Financial Economics, 33(1), 3–56. doi: 10.1016/0304-405x(93)90023-5. 1993.

[10] Huang, K.-W., Li, Z. A Multi-Label Text Classification Algorithm for Labeling Risk Factors in SEC Form 10-K. SSRN Electronic Journal. doi: 10.2139/ssrn.1916044. 2011.

[11] Lawrence, A. Individual investors and financial disclosure. Journal of Accounting and Economics, 56(1), 130–147. doi: 10.1016/j.jacceco.2013.05.001. 2013.

[12] Lehavy, R., Li, F., Merkley, K. The Effect of Annual Report Readability on Analyst Following and the Properties of Their Earnings Forecasts. The Accounting Review, 86(3), 1087–1115. doi: 10.2308/accr.00000043. 2011.

[13] Li, F. Annual Report Readability, Current Earnings, and Earnings Persistence. SSRN Electronic Journal. doi: 10.2139/ssrn.887382. 2006.

[14] Loughran, T., Mcdonald, B., Yun, H. A Wolf in Sheep's Clothing: The Use of Ethics-Related Terms in 10-K Reports. Journal of Business Ethics, 89(S1), 39–49. doi: 10.1007/s10551-008-9910-1. 2008.

[15] Loughran, T., Mcdonald, B. Textual Analysis in Accounting and Finance: A Survey. Journal of Accounting Research, 54(4), 1187–1230. doi: 10.1111/1475-679x.12123. 2016.

[16] Wang, L.,Cheng, Y. Application of Natural Language Processing in Financial Risk Detection. doi: 10.48550. 2024.

[17] Wang, C., Gu, X. Risk Element Identification in Bulk Stock Electronic Based on the Text of Domain News. Procedia Computer Science, 214, 1293-1300. doi: 10.1016/j.procs. 2022.

[18] Wallach, H. M. Topic modeling. Proceedings of the 23rd International Conference on Machine Learning. doi: 10.1145/1143844.1143967. 2006.