ISSN No:-2456-2165

Correlation Coefficients in Capital Market Analyses

Luisa Trombetta Burić
University of Applied Sciences Aspira Split, Croatia

Abstract:- This paper explored the application of three widely utilized correlation coefficients - Pearson, Spearman, and Kendall in the context of financial data. The Pearson correlation coefficient, Spearman rank correlation coefficient, and Kendall's tau are extensively examined in capital market analyses for their ability to capture relationships between two or more financial variables such as investment funds rankings by different criteria (returns, risk, costs, etc.), market portfolio returns, stock exchange indexes and interest rates. For each coefficient, a hypothetical example was provided. First example was demonstrated application of Pearson's correlation coefficient in analysing the relationship between stock returns and market index movements. Second example was demonstrated application of Spearman's rank correlation coefficient in analysing the relationship between interest rates and bond prices. Finally, third example was demonstrated application of Kendall's rank correlation coefficient in comparing fund rankings by three different criteria. Every hypothetical example provided instructions how these coefficients can be used for making informed business decisions in the capital market.

Keywords:- *Capital market, Pearson correlation coefficient, Spearman's rank correlation coefficient, Kendall's tau correlation coefficient.*

I. INTRODUCTION

Correlation coefficients represent important statistical measures that are often used in analyses and research related to capital markets. They show the degree of linear relationship between two or more variables, that is, to what extent those variables are related and whether they move in the same or opposite direction [1]. The most used correlation coefficients in capital market analyses are Pearson's coefficient for numerical variables and Spearman's and Kendall's coefficients for ordinal variables. They take on values in the range from -1 to +1, where 1 denotes perfect positive correlation, 0 no correlation, and -1 perfect negative correlation between the observed variables [2].

Pearson's correlation coefficient (denoted r) is a parametric measure of linear relationship between two numerical variables [3]. It measures the strength and direction of a linear relationship between two variables [4]. Pearson's correlation coefficient is applied in capital market analyses [5] [6] for:

- Measuring correlation between stock returns and market portfolio returns – enables estimation of beta coefficient and capital asset pricing modelling (CAPM).
- Assessing correlation between a stock's return and movements in certain macroeconomic variables (GDP, interest rates, exchange rate...)
- Testing the efficient market hypothesis – correlation between returns and past returns or publicly available information
- Correlation between stock returns and company's financial ratios (Price-to-Earnings Ratio P/E, Price-toBook Ratio P/B...)
- Determining correlation between liquidity and returns on stocks or bonds.
- Estimating diversification effects – correlation of returns across different asset classes (stocks, bonds, real estate...).

Pearson's r requires normally distributed data and interval measurement scale. It is very widely used in financial analyses due to its statistical properties.

Spearman's rank correlation coefficient is a nonparametric statistical measure of the relationship between two variables. Unlike Pearson's coefficient, it does not require the assumption of normal distribution or interval data [7]. Applications of Spearman's correlation coefficient in capital market analyses include:

- Assessing the correlation between bond credit ratings and their yields to maturity. Credit ratings are ordinal data.
- Comparing rankings and ratings of investment funds assigned by institutional agencies. These are ordinal data.
- Correlation between individual investors' risk preferences (risk ranks) and the structure of their portfolios.
- Examining the agreement between ranks of companies' market capitalization and their liquidity measured by trading volume.
- Determining the correlation between Price-to-Earnings (P/E) ranks and returns of those stocks.

Spearman's rank coefficient is useful when there is deviation from the normal distribution, existence of extreme observations and ordinal data which is common for financial time series and data [5].

Kendall's rank correlation coefficient is a nonparametric measure of the relationship between two ordinal or continuous variables. Unlike Pearson's correlation coefficient, Kendall's τ (tau) is based on the rank order (ranks) of observations rather than their actual values. Kendall's τ is calculated by comparing all possible pairs of observations (x,y) from two variables. If pairs have the same order (concordant pairs), they are positively correlated. If they have opposite order (discordant pairs), they are negatively correlated. The value of Kendall's τ ranges from -1 to 1, just like Pearson's correlation coefficient. A value closer to 1 or -1 indicates a stronger linear relationship between the variables. Kendall's τ is recommended when assumptions for Pearson's test are not met, e.g. for ordinal data or deviation from the normal distribution. Applications of Kendall's rank correlation coefficient in capital market analyses include:

- Examining correlation between stock returns and macroeconomic indicators (GDP, inflation, interest rates). Since financial data often deviates from the normal distribution, Kendall's τ may be more appropriate than Pearson's r.
- Assessing correlation between bond ratings assigned by rating agencies and bond yields to maturity. This involves ordinal data where Kendall's coefficient is appropriate.
- Comparing investment fund rankings by different criteria (returns, risk, costs etc.). With ranks, it makes no sense to calculate Pearson's r, but rather Kendall's τ.
- Correlation between individual investors' risk preferences and characteristics of their investment portfolios. Risk ranking allows calculation of Kendall's τ.
- Examining correlation between companies' market capitalization ranks and their price-to-earnings (P/E) ranks.

In market analyses, correlation coefficients enable investors and analysts to determine the relationship between different financial metrics such as stock, bond and other security returns, interest rates, currency exchange rates, commodity prices etc. [8]. Thus, they gain insight into the structure of relationships in financial markets and can more precisely predict movements of certain financial metrics and make higher quality investment decisions.

II. APPLICATION OF PEARSON'S CORRELATION COEFFICIENT IN ANALYSING THE RELATIONSHIP BETWEEN STOCK RETURNS AND MARKET INDEX MOVEMENTS

Pearson's correlation coefficient (denoted r) is a standard measure of linear correlation between two numerical variables [9]. The formula for calculating Pearson's correlation coefficient is:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{Y})^2}} \qquad (1)$$

where xi and yi for i=1,…,n, are individual values of the two variables, $\bar{x}$ and $\bar{y}$ are their arithmetic means, while n is a number of pairs (xi, yi).

The value of this coefficient ranges from -1 to +1, where 1 represents a perfect positive correlation, 0 means no correlation, and -1 denotes a perfect negative correlation.

Through an example, the correlation between the returns of several companies' stocks and the returns of the certain Stock Exchange Index (named STEXCHI) be analysed using Pearson's correlation coefficient. In this way, we wanted to determine whether there is a significant correlation between the price movements of individual stocks and the movements of the stock market index.

A. *Materials and methods (II.)*

Hypothetical data on the closing prices of three companies' stocks (Company X, Company Y, and Company Z) and data on the closing values of the STEXCHI index in the period from 01/01/2022 to 12/31/2022 were used for the analysis.

Based on the data, daily returns were calculated for all the securities under review. Pearson's correlation coefficient between the returns of individual stocks and index returns was calculated in Excel. The shortest period for which it makes sense to present data on stock prices and stock market index values for the purpose of calculating the correlation coefficient is one year or 252 trading days. Namely, for Pearson's correlation coefficient to be statistically significant and representative, it is advisable to have at least 50 to 100 observations, i.e. pairs of data. Given that trading on stock exchanges usually takes place 5 days a week (sometimes 6, including Saturdays), this would mean at least 50 to 100 trading days. If the period is too short with too few observations, the correlation coefficient may give a distorted picture of the true connectivity of the observed securities. For example, if you only look at one-month, random noise and short-term trends can influence too much the obtained connectivity measure. Therefore, one calendar year, with an average of about 252 trading days, is the shortest reasonable period for presenting data in such an analysis. Of course, the longer the time series that includes more observations, the estimate of the correlation coefficient will be more reliable and stable.

B. *Results (II.)*

Hypothetical data on the closing values of the STEXCHI index and the closing prices of stocks of three hypothetical companies X, Y and Z in the period from 01/01/2022 to 12/31/2022 are presented.

Table 1: Closing values of the fictive STEXCHI index and closing prices of stocks of three hypothetical companies X, Y, and Z (01/01/2022 to 12/31/2022).

Date	STEXCHI	C. X	C. Y	C. Z
01/01/2022	250	11	23	13
01/02/2022	256	11	23	12
01/03/2022	258	11	23	13
…	…	…	…	…
12/30/2022	280	12	25	14
12/31/2022	278	12	25	14

C. Company

From the Table I., the growth in value of all the securities observed over the 1-year period can be noticed. The STEXCHI index increased from the initial value of 250 euros to the final 278 euros, representing a growth of 11.2%. The share price of Company X recorded a growth from 11 to 12 euros, i.e. 8.7 % growth. The share price of Company Y increased from 23 to 25 euros, thus 8.6 % growth in value. The share of Company Z saw the strongest growth, strengthening from 13 to 14 euros, growth of 7.7 %.

The daily return is calculated by the formula:

$$R = (P_t - P_{t-1}) / P_{t-1} \quad (2)$$

where:
R is a daily return,
P_t is a security price on day t and
P_{t-1} is a security price on the previous day (t-1).

Example of calculating daily returns for STEXCHI for 01/02/2022:

$P_t = 256$
$P_{t-1} = 250$
$R = (P_t - P_{t-1}) / P_{t-1} = (256 - 250)/250 = 0.024 = 2.4\ \%$

The same calculation is done for other dates and securities.

Pearson's coefficient of correlation between daily returns for Stock X and STEXCHI is obtained by placing into formula (1) values:

n - number of trading days,
x_i for i=1,…,n - individual daily returns of Stock X
y_i for i=1,…,n - individual daily returns of STEXCHI,
$\bar{x}$ and $\bar{y}$ - the arithmetic means of those returns.

By entering all values, a hypothetical coefficient is obtained, e.g. 0.71 - which shows the strength of connection between the observed securities. Therefore, for each pair of securities the process is repeated to obtain their mutual correlation coefficient showing the strength of linear relationship over the observed time.

D. Discussion (II.)

Based on hypothetical scenarios of correlation coefficients between stock price movements and the STEXCHI index, it is possible to give some general guidelines for investment strategies and portfolio composition:

- All stocks exhibit high positive correlation (0.9 to 1) – In this case, all observed stocks closely follow the movements of the stock market index. The discussion would focus on the issue of portfolio diversification and the inability to reduce risk by investing only in stocks listed on reviewed market. The conclusion would be that it is advisable to include foreign stocks in the portfolio.
- Stocks show low, almost no correlation with the index (-0.2 to 0.2) – Here the discussion would go in the direction that the price movements of these stocks are mostly not related to the STEXCHI movements but depend on internal and sectoral specific factors. The conclusion would be that it is possible to construct a well-diversified portfolio with reviewed stocks.
- Analysis results indicate negative correlation between stocks and the index – In the case of negative coefficients, stocks mostly move in opposite direction from STEXCHI. The conclusion would be that such "defensive" stocks can serve well to balance the overall portfolio risk.

Of course, some combinations of these scenarios are possible, given different outcomes for individual stocks. This information would be useful in constructing an investment portfolio for risk assessment and decision on share ratios of individual stocks based on their correlation with market movements.

III. APPLICATION OF SPEARMAN'S RANK CORRELATION COEFFICIENT IN ANALYSING THE RELATIONSHIP BETWEEN INTEREST RATES AND BOND PRICES

The aim was to investigate the correlation between the interest rate movements of Central Bank X and government bond prices.

More precisely, for the purpose of a hypothetical analysis, the relationship between the average interest rate on Central Bank X's main refinancing operations and the yields of 13-year government bonds over the period from January 2010 to December 2022 was presented. To estimate the strength of the monotonic relationship between the two variables, Spearman's rank correlation coefficient was used. The obtained coefficient value served as the basis for drawing conclusions about the impact of Central Bank X's interest rate policy on government bond price movements, i.e. making investment decisions in that market. The formula for calculating Spearman's rank correlation coefficient is:

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} \quad (3)$$

where rs is the Spearman's rank correlation coefficient, d is the difference between the ranks of corresponding values of the two variables and n is the number of pairs of values.

A. Materials and methods (III.)

For the analysis, hypothetical data from Central Bank X on average interest rates on Central Bank X's main refinancing operations and hypothetical data from the Ministry of Finance on yields of 13-year government bonds for the period from January 2010 to December 2022 were used. Data were collected on an annual basis. The values of the variables were ranked and entered the SPSS program to calculate Spearman's rank correlation coefficient.

B. Results (III.)

Average annual interest rates of Central Bank X on its main operations for the period from 2010 to 2022 are shown in Table II. These are fictional data sets that approximate real central bank interest rate and bond yield time series, presented for the purpose of demonstrating statistical analysis of their correlation. Equal values of the variable are given equal rank, which is calculated as the average value of their ranks.

Table 2: Calculating Spearman's Rank Correlation Coefficient

Year	Average Central Bank X interest rate	Rank	13-year bond yield	Rank	Difference in ranks (d)	d^2
2010	2.5 %	13	4.2 %	12	-1	1
2011	2.25 %	12	4.7 %	13	1	1
2012	2.00 %	11	4.1 %	11	0	0
2013	1.75 %	10	3.9 %	10	0	0
2014	1.50 %	9	3.2 %	9	0	0
2015	1.25 %	8	2.5 %	7	-1	1
2016	1.00 %	7	1.8 %	4	-3	9
2017	0.75 %	6	2.1 %	6	0	0
2018	0.50 %	5	2.7 %	8	3	9
2019	0.25 %	3.5	1.9 %	5	1.5	2.25
2020	0.10 %	2	1.2 %	2	0	0
2021	0.00 %	1	0.9 %	1	0	0
2022	0.25 %	3.5	1.5 %	3	-0.5	0.25

For each year, the difference between ranks is calculated as:

$$d = rank(y_i) – rank(x_i) \quad (4)$$

The d^2 values are then summed for all pairs of data:
$\sum d^2 = 23{,}55$

and used in the formula for Spearman's correlation coefficient (3) to give:
$r_s = 1 - 6 \cdot 23.55/(13 \cdot 168) = 0.935$

The calculated Spearman's rank correlation coefficient r_s is 0.935, indicating a strong positive correlation between interest rate and yield.

C. Discussion (III.)

The obtained value of Spearman's rank correlation coefficient indicates a strong positive correlation between the ranks of interest rates and the ranks of yields. The coefficient value is closer to 1 than 0, meaning the correlation between the two variables is very strong. An increase in the interest rate rank is accompanied by an increase in the yield rank and vice versa – a decrease in the interest rate rank follows a decrease in the yield rank.

The ranks of the observed variables move in the same direction. In periods of expected interest rate hikes, it is recommended to reduce exposure to bonds, while in conditions of falling interest rates and rising bond prices, their share in the portfolio should be increased.

The obtained values of Spearman's rank correlation coefficient can vary between -1 and +1, and their interpretation depends on the sign and absolute size of r_s.

- $r_s = 1$
  This indicates a perfect positive correlation where the ranks of the two variables completely match. The increase of one variable fully follows the increase of the other variable. This is an ideal case of complete association.

- $0.7 \leq r_s < 1$
  Strong positive correlation. The increase of one variable is mostly followed by an increase in the other variable. Deviations are small. A statistically significant association can be claimed.

- $0.4 \leq r_s < 0.7$
  Moderate positive correlation. The ranks of variables mostly overlap in the direction of movement but there are considerable deviations. There is some association present, but it is not perfect.

- $0 < r_s < 0.4$
  Weak positive correlation. The direction of rank movement partially coincides but there are many deviations. No strong statistical association can be claimed.

- $r_s = 0$
  There is no correlation between the variables whatsoever. Ranks are completely independent.

- $-0.4 < r_s < 0$
  Weak negative correlation... (and analogously for other negative values).

The interpretation and conclusions strongly depend on the obtained r_s value and the strength and direction of correlation it indicates.

IV. APPLICATION OF KENDALL'S RANK CORRELATION COEFFICIENT IN COMPARING INVESTMENT FUND RANKINGS BY DIFFERENT CRITERIA (RETURNS, RISK, COSTS)

The aim was to investigate the correlation between rankings of investment funds according to returns, risk, and management costs. More precisely, the goal was to determine if there is a match in the fund ranks across these key performance indicators.

A. Materials and methods (IV.)

For the analysis, hypothetical data was used on returns, risk (measured by return standard deviation), and expense ratios for a sample of 4 equity funds. The funds were ranked according to each individual criterion, with rank 1 assigned to the top performing fund. Kendall's rank correlation coefficient was calculated in SPSS.

B. Results (IV.)

Table III. shows fund ranks by return, risk, and costs.

Table 3: Returns, risks, and expense ratios for 4 equity funds and their ranks

Fund	Return	Return rank	Risk	Risk rank	Expense ratio	E. ratio rank
A	7.5 %	1	2.95 %	3	1.0 %	1
B	6.8 %	2	3.10 %	2	1.5 %	4
C	6.3 %	3	2.85 %	4	1.3 %	3
D	5.9 %	4	3.20 %	1	1.2 %	2

Fund A has rank 1 based on the return criterion, Fund D has rank 1 based on the risk criterion, and Fund A also has rank 1 based on the expense ratio criterion. Based on these ranks, Kendall's correlation coefficient τ can be calculated to assess the agreement in fund rankings across the different criteria.

Formula for calculating Kendall's rank coefficient between two rank variable is:

τ = (Number of concordant pairs – Number of discordant pairs) / Total number of pairs (5)

- *Calculating coefficient between return rank and risk rank*

  Concordant pairs are: (A, C), (B, C)

  Discordant pairs: (A, B), (A, D), (B, D), (C, D)

For calculating numbers of concordant and discordant pairs first we must be sure that funds are sorted ascending according to first rank – return rank. Next, we must check for each pair of funds the relation of ranking order regarding the second rank - risk rank. For example, pair (A, B) is discordant because 3 > 2 and pair (A, C) is concordant because 3 < 4.

Total number of pairs (combinations) is

$n (n-1) / 2 = 4 \cdot 3 / 2 = 6$

Kendall's rank coefficient between return rank and risk rank is:

$\tau = (2 - 4) / 6 = -0{,}33$

- *Calculating coefficient between return rank and fee rank:*

  Concordant pairs are: (A, B), (A, C), (A, D),

  Discordant pairs: (B, C), (B, D), (C, D)

$\tau = (3 - 3) / 6 = 0$

Therefore, we obtained the following Kendall rank correlation coefficients:

τ (return, risk) = -0,33 and τ (return, fee) = 0

C. Discussion (IV.)

The Kendall's rank correlation coefficient between returns and fund risk is -0.33, indicating a weak negative linear correlation between return and risk criteria. This means that the concordance in the movement of these ranks is small and negative, i.e. that better ranked funds by return criteria do not match higher ranks by risk criteria. There is even a slight tendency of inverse movement of these ranks. The rank correlation coefficient between return risk and cost criteria is 0, which shows that no linear correlation exists between return rank and fee rank. The obtained low values of correlation coefficients indicate that there are certain disagreements between the fund ranking criteria. A fund's rank according to one criterion (e.g. return) is not necessarily aligned with its rank according to other criteria (risk and costs).

Investors should carefully analyse funds' rankings and ratings with respect to different performance benchmarks, as they are not necessarily mutually compatible, and funds may rank differently on individual benchmarks.

V. CONCLUSION

Correlation analysis is an indispensable statistical tool for gaining insight into relationships between variables in finance. As demonstrated through examples in this paper, correlation coefficients assist investors in constructing optimal portfolios, evaluating investment alternatives, forecasting security returns, and quantifying risks. By determining the strength of connections between individual stock returns and market movements, investors can better understand systematic and unsystematic risk exposures and manage them appropriately through diversification. The paper explains the methodology for calculating and interpreting the most important measures of correlation. Appropriate statistical testing procedures are also elaborated. The presented examples and applications provide a good foundation for utilizing correlation analysis in real-world investment analysis and decision-making.

REFERENCES

[1]. Fabozzi, F. J., Modigliani F., & Jones, F. J. (2010). Foundations of financial markets and institutions (4th ed.). Prentice Hall.

[2]. Woerheide, W., & Persson, D. (1993). An index of portfolio diversification. Financial Services Review, 2(2), 73-85.

[3]. Bodie, Z., Kane, A., & Marcus, A. (2014). Investments. McGraw-Hill Education.

[4]. Brooks, C. (2014). Introductory Econometrics for Finance (3rd ed.). Cambridge University Press.

[5]. Campbell, J. Y., Lo, A. W., & MacKinlay, A. C. (1997). The econometrics of financial markets. Princeton University Press.
[6]. Alexander, C. (2008). Market risk analysis: Value-at-risk models. Wiley.
[7]. Kokkonen, J. (2015). Financial modeling for market risk analysis using R. Packt Publishing Ltd.
[8]. Levine, D. M., Krehbiel, T. C., & Berenson M. L. (2005). Business Statistics: A First Course (4th ed.). Pearson Education Canada.
[9]. Elliott, A. C., & Woodward, W. A. (2007). Statistical Analysis Quick Reference Guidebook. SAGE Publications, Inc.