Journal of Innovation in Business and Economics
Vol. 07 No. 02 December 2023 Page 115-128
P-ISSN: 2580-9431 E-ISSN: 2581-2025 http://ejournal.umm.ac.id/index.php/jibe

Received: 24-05-2023 | Revised in Revision: 28-09-2023 | Accepted: 14-11-2023

Clustering-based value investing strategy in the Helsinki Stock Exchange: *k*-means algorithm

Topi Issakainen[1]
Master's Program in Strategic Finance and Business Analytics, School of Business and Management, LUT University, Finland[1]

DOI: https://doi.org/10.22219/jibe.v7i02.26424

Abstract
The purpose of this research is to study the possibility of combining quantitative clustering of stocks and value investing. The feasibility of this approach is tested using Finnish market data from the period 2005 to 2021. The benchmark index used in this research is the OMX Helsinki Growth Index. The strategy is based on the combination of P/E, P/CF, and P/B ratios, which serve as the basis for the *k*-means algorithm. The data is pre-processed by removing stocks that have not generated positive earnings and cash flow during the previous 12 months. The *k*-means algorithm assigns stocks to clusters, and the cluster with the lowest financial ratios is chosen as the value portfolio. The research also includes a sensitivity analysis of value portfolios, where the initial number of clusters in the clustering phase ranges from three to ten. Returns of different value portfolios are compared to each other and the benchmark index. The quality of results is evaluated using the Sharpe ratio and Jensen's alpha. According to the findings, the value portfolio constructed using nine clusters generated the highest risk-adjusted return, with an annual return of 30.27% over the 2005 to 2021 period. Furthermore, the best-performing value portfolio from 2005 to 2017 was compared to the benchmark index from 2018 to 2021. The value portfolio achieved an annual return of 26.05% during the 2018-2021 period, while the corresponding return of the benchmark index was 11.74%.

Keywords: Value investing; k-means clusteting; Helsinki Stock Exchange

Introduction

Value investing constitutes an investment strategy wherein investors acquire stocks with low ratios based on various value metrics like earnings, cash flow, and book value (Bodie, Kane, and Marcus, 2014). Academic literature underscores the crux of value investing as the investment in stocks exhibiting low price-to-earnings ratio (P/E) and price-to-book ratio (P/B) concerning the market (Hanson, 2013). Benjamin Graham, recognized as the progenitor of value investing, delineated diverse rules and techniques for evaluating companies' financial ratios in his seminal work, "Security Analysis" (Graham and Dodd, 1934). Across numerous studies, value investing has been extensively scrutinized in diverse markets and has consistently demonstrated the potential to yield higher returns compared to other stocks.

The initial investigation into the value anomaly was pioneered by Nicholson (1960) who scrutinized the U.S. stock market from 1939 to 1959. His findings indicated that stocks with low P/E ratios yielded superior returns compared to those with high P/E ratios. Basu (1977) echoed similar results in his research spanning the period from 1957 to 1971. Recent studies have corroborated the existence of the value anomaly from 1985 to 2006 (Athanassakos, 2011). Nevertheless, in recent decades, the surge in stock prices of various technology companies and technological advancements have propelled the ascendancy of growth investing. The robust performance of growth stocks has at times rendered value investing seemingly irrational. Notably, from 2002 to 2019, growth stocks outpaced value stocks (Miller and Prondzinski, 2020). This discrepancy underscores the significance and relevance of this research.

[1]E-mail: topi@windowslive.com

There exists limited recent research on value investing in the Finnish markets. Leivo and Pätäri (2009) conducted an extensive inquiry into the value anomaly in the Finnish markets, discovering evidence of the P/E anomaly over the period from 1993 to 2008 when stocks were segregated into terciles. However, anomalies related to P/B or price to cash flow ratio (P/CF) were not observed in the Finnish markets during their studies spanning from 1993 to 2008 (Leivo and Pätäri, 2009, 2011). Conversely, weak evidence of the P/B anomaly was identified from 1991 to 2006 in the Finnish markets (Leivo, Pätäri, and Kilpiä, 2009). Therefore, it is intriguing to explore the existence of the value anomaly in the Helsinki Stock Exchange from 2005 to 2021.

This research endeavors to scrutinize the value anomaly and its manifestation in the Finnish markets from 2005 to 2021 by employing the k-means algorithm for stock portfolio formation. The k-means clustering method stands out in data clustering compared to classical approaches such as mean or quartile-based clustering. Its flexibility in handling various forms of clusters is a key advantage. Moreover, its independence from specific data distribution assumptions allows for the formation of clusters with more complex shapes. Additionally, k-means exhibits superior performance when confronted with high-dimensional data or a substantial number of attributes, as it focuses on calculating distances between data points. The interpretability of k-means clustering results is another advantage, facilitated by the clear representation provided by the centroid of each cluster. Furthermore, this research employs varying numbers of clusters, comparing the outcomes with the benchmark index, namely the OMX Helsinki Growth Index. Additionally, data from 2005 to 2017 serves as a training dataset to ascertain the optimal number of clusters, subsequently applied to a test dataset spanning from 2018 to 2021. The objective is to ascertain whether selecting the optimal number of clusters during the clustering phase empowers investors to generate risk-adjusted outperformance.

This research yields valuable contributions both in practical application and academia. In practical terms, it introduces a data-driven and automated approach to portfolio construction, leveraging the potency of the k-means algorithm. This methodology has the potential to streamline the intricate process of stock selection and management within a value investing strategy. From an academic perspective, this research advances the knowledge base in the realm of finance and investment by investigating value anomalies in the Finnish market—a domain that has received relatively limited attention. Moreover, the application of the k-means algorithm in portfolio formation presents an innovative methodology open for scrutiny and adaptation by fellow researchers in the field. The findings from this research might spark further academic inquiry into the effectiveness of automated portfolio construction techniques and their influence on risk-adjusted returns across diverse market contexts, thereby enriching the scholarly discourse in finance and economics.

Literature Review

The utilization of clustering in stock selection represents a relatively recent method, hence most peer-reviewed research on this topic has emerged in the 21st century. Mirkin (1996) succinctly defines clustering as a mathematical technique designed to unveil classification structures within real-world data. The primary objective of clustering is to group similar data points—here, stocks with analogous financial ratios—into clusters, ensuring that data points within a cluster are akin to each other while being dissimilar to those in other clusters. Effective clustering manifests in high intra-class similarity and low inter-class similarity, denoting high homogeneity within each cluster and minimal similarity between different clusters.

Nanda, Mahanty, and Tiwari (2010) scrutinized the Indian stock market from 2007 to 2008 employing a sample of 106 stocks. These researchers classified stocks based on valuation ratios to construct portfolios that performed well. The clustering methods utilized encompassed *k*-means, self-organizing map (SOM), and Fuzzy C-means. Assessment of these clustering methods was conducted using various validity indices like Silhouette and Calinski-Harabasz. The outcomes suggested that employing five or six clusters was optimal for this dataset. The compactness of each cluster was gauged using the Interclass inertia method, calculating the average squared Euclidean distance between the cluster mean and each observation. Results demonstrated that *k*-means clustering yielded the most suitable clusters, and an efficient portfolio was constructed based on Markowitz's (1952) model. This clustering approach simplifies portfolio construction by enabling investors to select stocks from clusters rather than manually analyzing extensive datasets (Nanda et al., 2010).

Similarly, Bini and Mathew (2016) in subsequent years reported analogous success regarding the efficacy of *k*-means clustering in the Indian markets. These authors compared diverse partition-based, density-based, and hierarchical clustering techniques for stock selection, utilizing a sample of 1232 stocks. The findings indicated that *k*-means and the Expectation Maximization algorithm proved to be the most effective clustering methods based on validity indices like the Rand index or Silhouette method. Additionally, the authors underscored that *k*-means clustering represents an efficient and scalable methodology (Bini and Mathew, 2016).

Moreover, the *k*-means algorithm exhibited success in a study centered on stock forecasting in 2021. In this study, similar stocks were clustered into portfolios, and their returns were forecasted using a Long Short-Term Memory model. Results indicated that employing the *k*-means algorithm did not compromise prediction accuracy. The authors concluded that leveraging the *k*-means clustering approach enables investors to economize time by reducing manual screening efforts for similar stocks (Affonso et al., 2021).

The *k*-means algorithm has seen application in stock selection within emerging markets as well. Al-Augby, Majewski, Majewska, and Nermend (2014) scrutinized the Gulf Cooperation Council (GCC) markets, employing samples of bank and energy stocks from Bahrain, Saudi Arabia, Qatar, Oman, and the United Arab Emirates. Their investigation delved into the impact of news on stock prices, comparing the performance of the *k*-means and fuzzy c-means clustering methods based on stock price reactions. Clustering relied on various financial ratios like P/E, P/B, and market capitalization. The results highlighted that the *k*-means algorithm yielded favorable outcomes, as companies' stocks across different clusters exhibited analogous reactions to news events (Al-Augby et al., 2014).

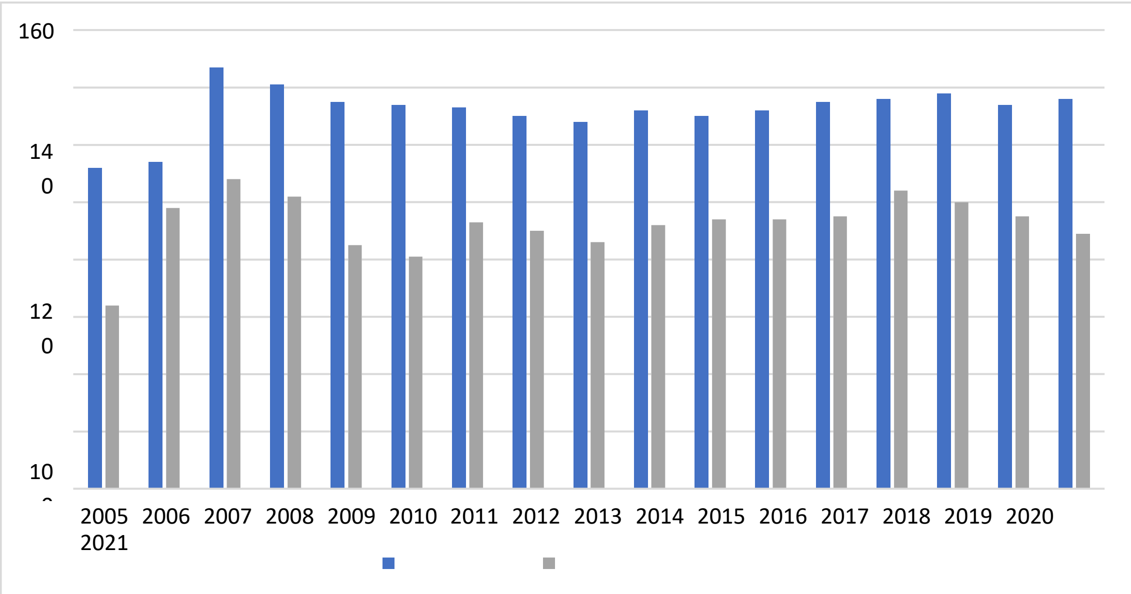

Figure 1. Initial sample size and the sample size after stocks with negative P/E or P/CF ratios and outliers have been removed from 2005 to 2021

Similarly, proven efficacy of *k*-means clustering has been documented in Thailand's stock markets. Muangprathub, Intarasit, Boongasame, and Phaphoom (2020) successfully utilized the *k*-means algorithm to construct diversified portfolios spanning the years 2015-2017. In the Indonesian markets from 2016 to 2019, Zuhroh, Rofik, and Echchabi (2021) applied *k*-means clustering to categorize banks based on their average rate of return. Subsequent regression analyses on macroeconomic factors such as inflation, exchange rate, and interest rate unveiled distinctive traits within the clusters. Notably, the cluster exhibiting the highest rate of return demonstrated greater resilience to diverse macroeconomic factors compared to others. Consequently, these clusters were identified as appealing options for long-term investments due to their stability amidst changing macroeconomic scenarios (Zuhroh et al., 2021).

Further evidence supporting the utility of the *k*-means algorithm comes from Cheong, Kim, Buyn, Oh, and Kim (2017) in Korean markets. Their research aimed to partition stocks into portfolios based on available investor information. *k*-means clustering facilitated stock selection tailored for different investor types (institutional, foreign, individual). Additionally, the study integrated a genetic algorithm to determine optimal stock weights within portfolios. By aligning portfolios with investor objectives based on information, the results indicated that employing the *k*-means clustering algorithm led to enhanced stock-picking performance and improved portfolio returns (Cheong et al., 2017).

Moreover, *k*-means clustering has found application in portfolio formation in another study. El, Guennon, and Hamza (2012) utilized *k*-means clustering to group stocks exhibiting similarity in terms of their value at risk (VaR) and expected return. This clustering approach facilitated subsequent steps, where an optimization algorithm was applied to minimize risk and maximize expected return. The result was the identification of three distinct clusters, enabling scholars to pinpoint attractive stocks for their optimization algorithm (El et al., 2012). This chapter's literature underscores various applications of clustering within finance, notably in portfolio formation. As previous studies indicate, the *k*-means algorithm has demonstrated robust performance in stock selection (Affonso et al., 2021; Bini and Mathew, 2016; Nanda et al., 2010).

Research Method

Data

The research dataset spans from 2004 to 2021, encompassing crucial data from 2004 necessary for calculating financial ratios pertaining to the initial return calculation year of 2005. It comprises comprehensive information, including daily returns, P/E (Price-to-Earnings), P/B (Price-to-Book), and P/CF (Price-to-Cash Flow) ratios for all stocks listed on Nasdaq OMX Helsinki from 2005 to 2021. The dataset is sourced from the Eikon database, and data manipulation and calculations are conducted utilizing Microsoft Excel. Furthermore, the implementation of the *k*-means clustering technique is executed using RStudio.

The computation of P/E, P/B, and P/CF ratios involves dividing a stock's price on the first trading day of the year by the most recent 12-month data available, excluding the preceding year's Q4 data. This methodology is employed since companies typically report Q4 data between late January and March. To accommodate this reporting pattern, adjustments for stock splits and dividends are factored in using the stock's total return index, assuming reinvestment of dividends and stock splits in the same company. Throughout this research, no transaction costs are assumed.

Figure 2. OMXHGI return from 2005 to 2021

For the clustering phase, only companies demonstrating positive cash flow and earnings within the last 12 months are considered viable options. Consequently, companies with negative P/E and P/CF ratios are filtered out, aligning with the objective of constructing a high-quality value portfolio. Companies failing to generate positive earnings and cash flow are consistently excluded from the dataset annually. Additionally, to optimize the efficiency of *k*-means clustering, evident outliers are eliminated. Observations deviating from the mean by more than three times the standard deviation are excluded. Figure 1 provides a visual depiction of the initial sample size by year and the resultant sample size post-application of these adjustments.

The average sample size observed from 2005 to 2021 stood at 132. However, when companies commenced the year with negative P/E or P/CF ratios, the average sample size decreased to 92. It's noteworthy that preceding studies addressing value anomalies in the Finnish markets (Leivo and Pätäri, 2009, 2011; Leivo, 2012b; Leivo et al., 2009; Pätäri and Leivo, 2009) did not employ filters excluding companies unable to generate positive earnings and cash flow. The selected benchmark index for this research was the Helsinki Stock Exchange Growth Index (OMXHGI), chosen due to its inclusion of dividend returns, rendering it the most appropriate benchmark. Figure 2 provides a visual representation of the benchmark index returns. Over the period spanning 2005 to 2021, the annual return of the benchmark index averaged 7.30%, culminating in a cumulative return slightly exceeding 250% across the sample duration.

Methodology

The empirical segment of this research is focused on crafting value portfolios using the *k*-means clustering algorithm and subsequently contrasting their returns against those of the benchmark index. This analysis also entails a comparative assessment of returns across various cluster numbers. The initial phase of the research involves computing P/E, P/CF, and P/B ratios for individual stocks. These ratios are derived based on each stock's data on the first trading day of the year, utilizing the most recent 12-month data as the denominator. Stocks exhibiting negative P/E or P/CF ratios are excluded from the sample.

The research employs the *k*-means clustering method following Zuhroh, et al. (2021). Formally, the method can be presented as follows. Assuming there is a set of observations $(x_1, x_2, x_3 \ldots, x_n)$ where each observation is a real vector with d-dimension, *k*-means clustering aims to divide n observations into k cluster $(k \leq n)$. The set $S = \{S_1, S_2, S_3, \ldots, S_k\}$ thus minimizing the within-cluster sum of square-like variants. The purpose of minimizing this variant can be stated.

$$\underset{S}{\arg\min}\sum_{i=1}^{k}\sum_{x\in S_i}\|x-\mu_i\|^2 = \underset{S}{\arg\min}\sum_{i=1}^{k}|S_i|\,Var\,S_i \qquad (1)$$

Where μ_i is the mean of points in S_i. This is equivalent to minimizing the paired square deviations of points in the same cluster.

$$\underset{S}{\arg\min}\sum_{i=1}^{k}\frac{1}{2|S_i|}\sum_{x,y\in S_i}\|x-y\|^2 \qquad (2)$$

The equivalence can be deduced from identity $\sum_{x\in S_i}\|x-\mu_i\|^2 = \sum_{x\neq y\in S_i}(x-\mu_i)(\mu_i-y)$. Since the total variance is constant, it is equivalent to maximizing the sum of the squared deviations between points in different clusters following the law of total variance.

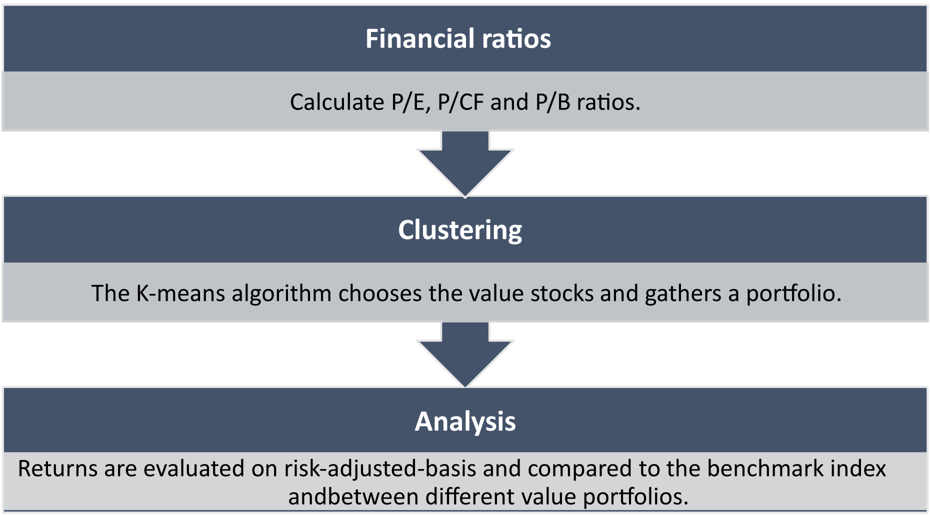

Figure 3. Methodology of research

The subsequent step, outlined in Figure 3, entails the application of the *k*-means algorithm. Each year, the algorithm processes stocks possessing positive P/E, P/CF, and P/B ratios. Diverse cluster numbers, ranging from three to ten, are tested to discern the most optimal value portfolio concerning risk-adjusted returns. In contrast to prior studies such as Nanda, Mahanty, and Tawani (2010) and Bini and Mathew (2016), which evaluated clustering performance using validity indexes like Calinski-Harabasz and Silhouette, this research adopts a distinct approach. This research compares different cluster numbers based on risk-adjusted value portfolio performance. In cases where the *k*-means algorithm generates multiple options for the value portfolio, selection is based on the portfolio exhibiting the two out of three lowest financial ratios. For instance, if Portfolio X displays a lower P/E ratio compared to Portfolio Y, while Portfolio Y showcases lower P/CF and P/B ratios, Portfolio Y is chosen as the value portfolio.

Following the determination of the value portfolio by the *k*-means algorithm, returns and performance metrics are computed, integrating dividends. Stocks are evenly allocated within the value portfolio, ensuring an equitable distribution among all stocks. The Sharpe ratio and Jensen's alpha serve as assessment tools to evaluate risk-adjusted performance (Jensen, Johnson, & Mercer, 1997; Jensen, 1967; Sharpe, W.F., 1964; 1966). The selected risk-free rate for this research is the 10-year German government bond, considered representative of the risk-free rate across Europe. Portfolio rebalancing occurs at the beginning of each year, spanning from 2005 to 2021, covering a 17-year sample period deemed adequate for producing valid and reliable results. Once value portfolios are constructed for each year, their performance is compared among themselves and against the benchmark index.

Result and Discussion

The penultimate section of this research delineates the attained results, primarily focusing on addressing the second research question. It entails a comparative analysis of returns from diverse value portfolios constructed with varying cluster numbers, juxtaposed against both each other and the benchmark index. Additionally, previously introduced performance metrics are computed for these portfolios. Furthermore, based on the observed outcomes across the sample period, the chapter determines the most optimal number of clusters for this dataset. Subsequently, this chosen cluster number is utilized to gauge the performance of the value portfolio against the benchmark index specifically for the duration spanning 2018 to 2021.

Value portfolios' performance

The results subsection provides an overview of the performance exhibited by the value portfolios throughout the sample period. Illustrated in Figure 4 are the annual returns of these portfolios, where a numerical value, such as "3," denotes the creation of three clusters derived from the sample data spanning 2005 to 2021. The cluster characterized by the lowest financial ratios was subsequently selected as the value portfolio.

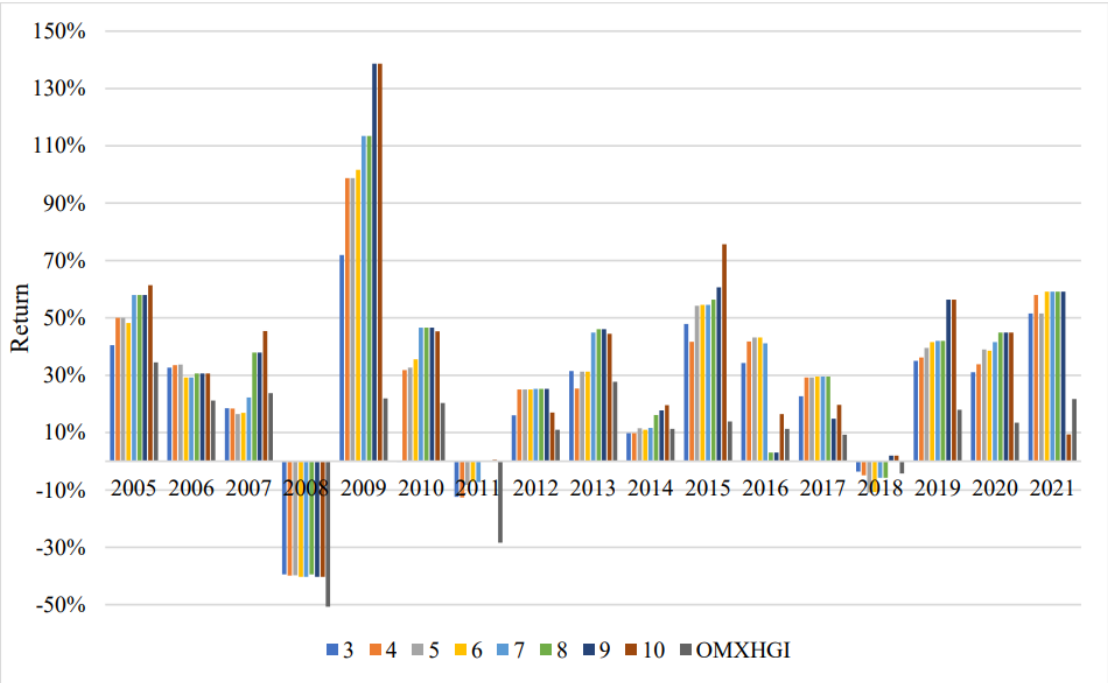

Figure 4. Yearly returns from 2005 to 2021 of the benchmark index and value portfolios when 3-10 clusters are used in the clustering phase

Notable years highlighted within Figure 4 are 2008 and 2009. The global financial landscape grappled with the repercussions of the subprime crisis in 2008, resulting in a substantial 51% decline in OMXHGI. However, the subsequent year experienced a notable recovery, particularly within the realm of value investing, with the average return of the value portfolios in 2009 reaching 109.41%. Remarkably, the two smallest portfolios, consisting of 9 and 10 clusters, achieved returns of 138.68% during this period. Additionally, Figure 4 accentuates 2011 and 2018 as relatively poor performing years. In 2011, Europe confronted a severe debt crisis, especially in Southern European nations, marked by a significant surge in bond rates due to escalating debt levels and unwise economic policies. Similarly, 2018 witnessed lackluster stock performance due to multifaceted factors including Brexit and the US-China trade war, culminating in OMXHGI declines of 28% in 2011 and 4% in 2018.

Table 1. Annualized value portfolio returns by the number of clusters

Clusters	3	4	5	6	7	8	9	10
Return	19.68 %	24.22 %	25.35 %	25.78 %	28.87 %	28.79 %	30.27 %	29.38 %

Another remarkable year for the value portfolios was 2016. Portfolios emphasizing value stocks (constructed from 8-10 clusters) demonstrated an average return of 7.54% during that period. In contrast, portfolios with greater diversification (constructed from 3-7 clusters) showcased an average annual return of 40.68%. Across other years, returns remained relatively consistent. It's essential to note that the sample solely comprised companies exhibiting positive earnings and cash flow in the previous 12 months. As highlighted, the least favorable years for value portfolios were 2008, 2011, and 2018, while the most favorable were 2009, 2015, and 2021.

Figure 4 consistently demonstrates a right skew, indicating that more concentrated portfolios yielded superior returns. This suggests a positive correlation between portfolios primarily composed of value stocks and higher returns. For further details on the annualized returns of the value portfolios, please refer to Table 1.

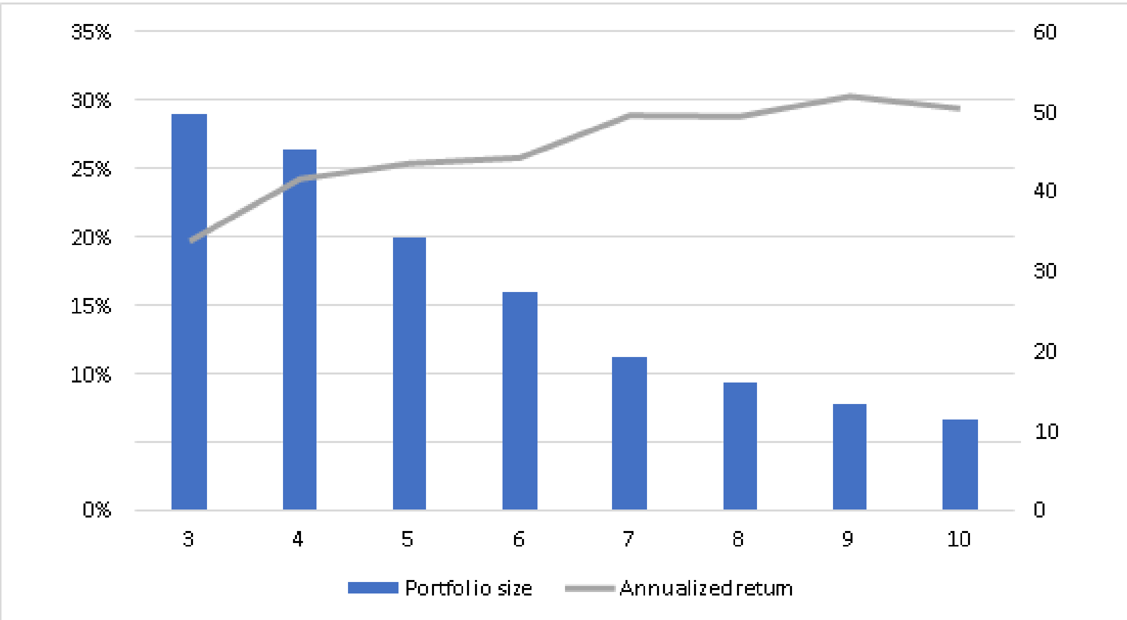

Figure 5. Average portfolio size and annualized return for value portfolios during 2005-2021

Based on raw returns, the optimal choice for a value investor appears to be between 7-10 initial clusters. Remarkably, minimal differences exist in returns when employing 7, 8, 9, or 10 clusters. Notably, all returns demonstrate significant levels of performance compared to prior literature focusing on value investing in the Finnish markets. For instance, Leivo (2012) reported that the top value portfolio yielded an average annual return of 18.449% across 1993-2009. However, it's crucial to note the disparity in the time periods analyzed between this research and Leivo's study. An important determinant impacting returns is that this research exclusively incorporates companies that exhibited positive earnings and cash flow within the 12 months preceding portfolio construction.

Analyzing cumulative returns, the value portfolio constructed from 9 clusters stands out remarkably, boasting a cumulative return of 8856.27% from 2005 to 2021. The second-best performing portfolio was formulated using 10 clusters, closely followed by the portfolio derived from 8 clusters during the clustering phase. Notably, all value portfolios have significantly surpassed the benchmark index in terms of cumulative returns. Conversely, the value portfolio utilizing 3 clusters achieved the smallest cumulative return at 2019.12%, while the OMXHGI achieved a cumulative return of 255.22%.

Table 2. Portfolio sizes, Sharpe ratios and Jensen alphas by the number of clusters

Clusters	3	4	5	6	7	8	9	10
Portfolio size	50	45	34	27	19	16	13	11
Sharpe ratio	0.64	0.78	0.78	0.85	0.93	0.88	0.94	0.93
Jensen's alpha	19.51 %	24.92 %	26.56 %	27.20 %	31.46 %	30.63 %	35.17 %	35.33 %

Table 3. Portfolios ranked by returns during and average returns from 2005 to 2021

Cluster	**Average return in year**
10	34.53 %
8	33.16 %
9	35.38 %
7	33.28 %
5	29.25 %
4	28.00 %
6	29.83 %
3	22.80 %

Table 4. Most common stocks from 9 clusters over the years 2005-2021

Stock	**Tmes**
Telia Company	15
Neles	14
Cramo	12
Nordea Bank	11
Digia	9
Sievi Capital	7
Saga Furs	7
Elecster A	6
PKC Group	5
SSAB B	5

An observable trend in returns indicates that as the initial cluster count rises, returns also tend to increase. This phenomenon likely occurs due to the heightened efficiency of the *k*-means algorithm in identifying superior value stocks with a larger cluster count. Consequently, value portfolios become more focused on stocks with lower valuation metrics, excluding those with moderate financial ratios. As a result, returns escalate over the 2005-2021 period, as depicted in Figure 5. Additionally, the average portfolio size diminishes as the sample is divided into a larger number of clusters.

An intriguing finding of this research suggests that achieving optimal returns necessitates a relatively higher number of clusters. Beyond a certain point (7 clusters in this research), further increases in the cluster count don't notably enhance returns, as the *k*-means algorithm has already identified the most promising value stocks.

Moreover, Figure 5 validates the efficacy of the *k*-means clustering approach in constructing a diversified, high-quality value portfolio that yielded robust returns. Even with a substantial initial cluster count, the average size of the value portfolio retained 11 stocks across 2005-2021. This effectively mitigated the risk of the value portfolio becoming excessively concentrated, as 11 stocks ensure a reasonable level of diversification. Presuming low correlation among stock returns, this diversification likely reduced the overall risk of the value portfolio. It's pivotal to acknowledge that the average sample size from 2005-2021 was 92, relatively small in scale. Therefore, these findings might have substantially differed if a larger market segment had been encompassed.

Based on raw returns and further substantiated by Table 2, there is evidence of a value anomaly spanning the 2005 to 2021 period. Table 2 reveals notable trends in average portfolio sizes and risk-adjusted performance metrics. As indicated, both the Sharpe ratio and Jensen's alpha show an upward trajectory as the value portfolio becomes more concentrated. This pattern strongly implies the existence of a value anomaly within the Finnish markets across the sample period. Moreover, it suggests that as a portfolio attains a certain size, its unsystematic risk diminishes, aligning with the principles outlined in the Capital Asset Pricing Model (CAPM).

Table 5. Value portfolio's return, Sharpe ratio and Jensen's alpha over the years 2018-2021

Year	Return	Sharpe ratio	Jensen's alpha
2018	1.92 %	0.21	1.17 %
2019	56.33 %	1.72	56.79 %
2020	44.88 %	0.82	45.85 %
2021	9.37 %	0.09	1.84 %

Based on the evaluation of the Sharpe ratio and Jensen's alpha, the optimal number of clusters for this dataset appears to be 9, with a marginal difference between 9 or 10 clusters. While 10 clusters slightly outperformed in Jensen's alpha, the portfolio generated a marginally higher Sharpe ratio with 9 clusters. However, the annualized returns were 0.89% higher with 9 clusters. Consequently, the most advantageous number of clusters for the Helsinki Stock Exchange during the 2005-2021 period, employing the *k*-means clustering method, is determined to be nine.

The selected value portfolio faced a decline only once, notably in 2008, during the sample period. It managed to produce a modest positive return in 2011 and 2018. This consistency is particularly significant in terms of capital compounding, as negative returns can substantially impact long-term compounding. Additionally, in 8 out of 17 years, this portfolio yielded the best or equally strong returns compared to other value portfolios. Table 3 ranks the portfolios based on the number of times they generated the highest returns during the 2005-2021 period. Although portfolios constructed from 8 and 10 clusters performed nearly as well as the one from 9 clusters, the value portfolio created from 9 clusters exhibited the highest average return during the 2005-2021 period, despite portfolios from 8 and 10 clusters achieving higher returns more frequently.

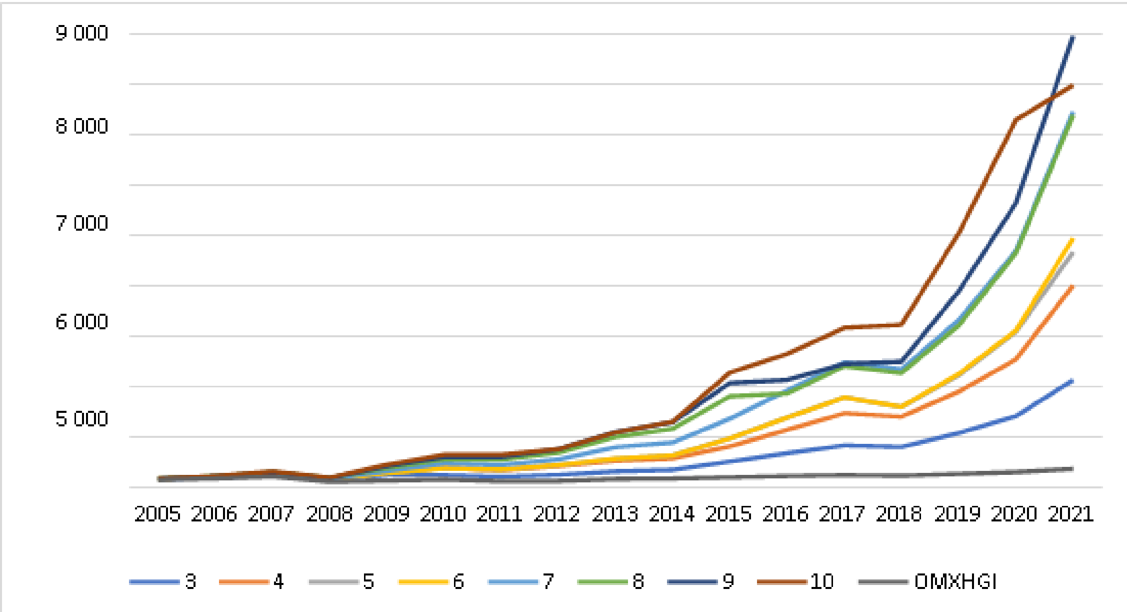

Figure 6. Cumulative returns from 2005 to 2021 of the benchmark index and value portfolioswhen 3-10 clusters are used in the clustering phase

Table 4 showcases the 15 most common stocks in the best-performing value portfolio from 2005 to 2021. The Swedish telecommunications company Telia appeared most frequently in the best value portfolio, consistently chosen by the *k*-means algorithm for inclusion almost every year. Furthermore, traditional industrial companies such as Neles and SSAB were among these stocks. They share the characteristic of being cyclical in terms of their valuation. Notably, growth-oriented companies with high valuation metrics were absent from Table 4, highlighting the *k*-means algorithm's success in identifying value stocks. The recurrence of the same stocks in value portfolios over the years was relatively low, with only 5 companies present in the best-performing value portfolio for over 50% of the sample period. This underscores the *k*-means algorithm's capability to pinpoint stocks with low valuation metrics across various years.

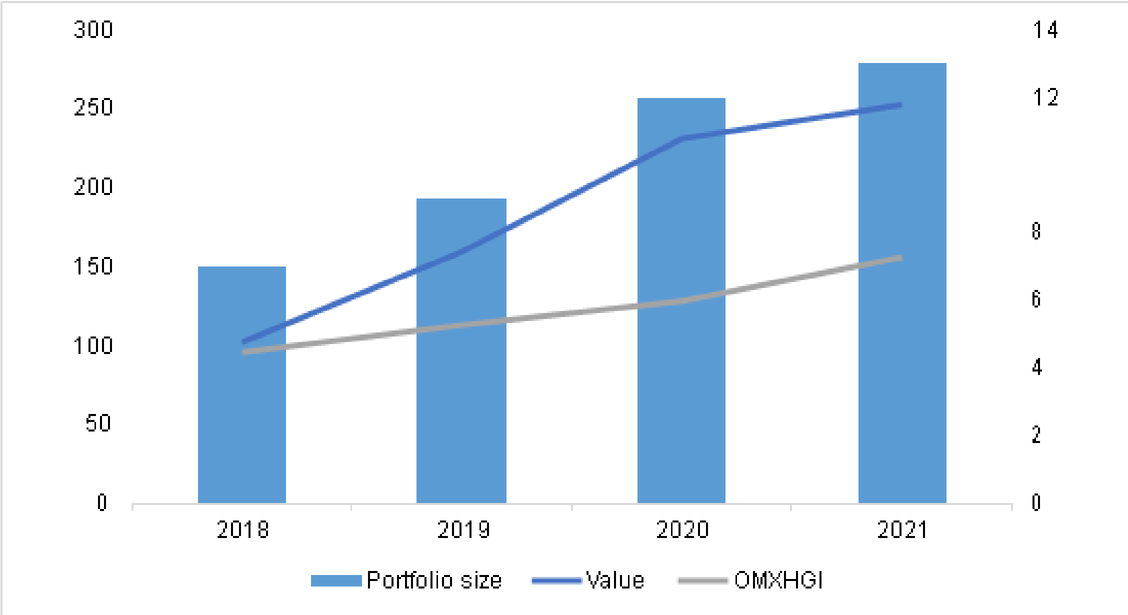

Figure 7. Value portfolio's return against OMXHGI over the 2018 to 2021 period

The results derived from this study indicate that the *k*-means algorithm effectively identified value stocks within the Finnish markets from 2005 to 2021. Moreover, these findings align with previous literature on value investing within the Finnish markets, as indicated by studies conducted by Leivo and Pätäri (2009, 2011), Leivo (2012b), Leivo et al. (2009), and Pätäri and Leivo (2009). Notably, this research sets itself apart by not relying on a single financial ratio to construct value portfolios. Instead, it employs a combination of P/E, P/CF, and P/B ratios. Additionally, the *k*-means algorithm determines the number of stocks within each value portfolio, differing from prior studies that often divide stocks into predetermined groups based on financial ratios, such as the approach taken by Fama and French (2006), who categorized stocks into five groups before forming value portfolios.

This distinction in methodology is crucial to consider when comparing and contrasting the outcomes of this research with those of prior studies. The utilization of multiple financial ratios in combination with an algorithm-based determination of portfolio size sets a different precedent in identifying value stocks. Therefore, while the results are consistent with the presence of a value anomaly in the Finnish markets, the methodology employed in this research may yield divergent or more refined outcomes when compared directly to studies adopting different approaches. Hence, understanding the nuances in methodology across studies is essential for a comprehensive interpretation and comparison of the findings.

Value portfolio's performance against the benchmark index during 2018-2021

It's apparent from the presented subchapter that a backtest of the value portfolio against the OMXHGI benchmark was conducted, focusing on the 2018-2021 period. The value portfolio was constructed utilizing 10 clusters, a number that had demonstrated superior risk-adjusted returns in the 2005-2017 period, as indicated in Figure 6. Figure 7 illustrates the returns and portfolio size of the value portfolio in comparison to the benchmark index. Notably, the size of the value portfolio fluctuated between 7 and 13 stocks during the 2018-2021 period. This variability in portfolio size reflects the *k*-means algorithm's capacity to identify stocks with favorable valuation metrics across different years. Across this period, the annualized returns of the value portfolio stood at 26.05%, surpassing the benchmark index's annualized returns of 11.74%. The most recurrently included stocks within the value portfolio during this period were Telia, Honkarakenne, Incap, Alandsbanken, and Oma Säästöpankki. Furthemore, Table 5 provides a detailed breakdown of the value portfolio's returns, Sharpe ratios, and Jensen's alphas for the years 2018-2021. It indicates that the years 2019 and 2020 were particularly significant for the value portfolio, showcasing robust Sharpe ratios and Jensen's alphas, contributing substantially to outperforming the

OMXHGI benchmark. Conversely, 2018 and 2021 exhibited relatively weaker risk-adjusted performance, suggesting a downturn in value stocks' performance during these periods, accompanied by increased return volatility.

The findings from the backtest analysis from 2018 to 2021 suggest that employing the *k*-means algorithm for constructing a value portfolio led to robust returns. During this period, the cumulative return using this approach was 152.47%, outperforming the OMXHGI's cumulative return of 55.90%. The average Sharpe ratio of 0.71 and average Jensen's alpha of 26.41% further substantiate the strong risk-adjusted performance.

However, a noteworthy revelation emerged regarding the choice of the value portfolio constructed from 10 clusters. Contrary to expectations, the portfolio built from 9 clusters outperformed the 10-cluster portfolio significantly. The value portfolio from 9 clusters generated a cumulative return of 267.32% during the 2018-2021 period, surpassing the 10-cluster portfolio by a considerable margin. Additionally, the 9-cluster portfolio exhibited better risk-adjusted performance, boasting a higher Sharpe ratio (1.15) and Jensen's alpha (40.88%). These findings underline the importance of optimizing the number of clusters in the *k*-means algorithm for constructing the value portfolio.

The methodology employed, which involved filtering out companies with negative earnings or cash flow over the previous 12 months, coupled with the utilization of the *k*-means algorithm, has proven to be highly effective. This approach has enabled the construction of portfolios comprising high-quality value companies, as evidenced by the substantial returns. Moreover, the research's methodology differed notably from certain academic literature, particularly in the filtering process used when constructing the value portfolio. By excluding companies with negative earnings or cash flow, the research reduced the risk of including financially distressed firms in the value portfolio. Additionally, the use of three financial ratios (P/E, P/B, and P/CF) as input for the *k*-means algorithm facilitated the identification of undervalued stocks across multiple valuation metrics, enhancing the algorithm's ability to pick genuinely undervalued stocks.

The performance achieved by this strategy in the Finnish markets over the 2005-2021 period was exceptional. The best-performing value portfolio delivered an annualized return of 30.27%, showcasing strong performance. With a Sharpe ratio of 0.94 and Jensen's alpha of 35.17%, this strategy consistently produced significant alpha, surpassing the returns from previous studies in the same market. Notably, the exclusion of companies with negative earnings or cash flow in the previous 12 months notably enhanced the value strategy's returns compared to prior research. These results collectively emphasize the effectiveness of employing the *k*-means algorithm and incorporating stringent filtering criteria for constructing high-performing value portfolios in the Finnish markets.

Conclusions, suggestionsand limitations

The analysis conducted in this research study showcases the presence of value anomalies within different markets and time periods, despite some inconsistencies in findings. The approach employed in this study consistently produced risk-adjusted excess returns, surpassing the benchmark index and indicating the effective identification and inclusion of attractively priced value stocks in the Finnish markets from 2005 to 2021. The frequent turnover of stocks within the value portfolio suggests that the approach adeptly recognized and incorporated stocks with favorable valuations. This dynamic strategy contributed to robust returns, supporting the idea of mean reversion in valuation metrics as proposed in existing literature, as evidenced by the frequent portfolio turnover. Moreover, the strategy's ease of implementation and automation through the *k*-means algorithm enhances its accessibility and manageability. However, it's important to note that actual returns might decrease when considering trading costs and taxes. Nonetheless, achieving an average annual return of 30.27% stands as a remarkable feat, signaling a significant performance gap between the value strategy and the OMXHGI benchmark index.

Further exploration in this field is crucial. Implementing this strategy in different markets, particularly in larger and more diverse ones like the U.S., could provide invaluable insights into its adaptability and performance variations across various market landscapes. Additionally, investigating the inclusion of different or supplementary financial ratios beyond those considered in this study might offer new perspectives because potentially enhances overall performance.

Competing Interests

The author(s) declare that there are no competing interests relevant to the content of this article.

References

Affonso, F., Magela, T., Dias, R., & Pinto, A. L. (2021). Financial Times Series Forecasting of Clustered Stocks. *Mobile Networks and Applications, 26*. https://doi.org/10.1007/s11036-020-01647-8

Al-Augby, S., Majewski, S., Majewska, A., & Nermend, K. (2014). A Comparison Of *k*-means And Fuzzy C-Means Clustering Methods For A Sample Of Gulf Cooperation Council Stock Markets. *Folia Oeconomica Stetinensia, 14*(2).

Athanassakos, G. (2011). The Performance, Pervasiveness and Determinants of Value Premium in Different US Exchanges: 1985-2006. *The Journal of Investment Management, 9*. https://ssrn.com/abstract=1916202

Basu, S. (1977). Investment Performance of Common Stocks In Relation To Their Price-earnings Ratios: A Test of The Efficient Market Hyporesearch. *The Journal of Finance, 32*(3). https://doi.org/10.2307/2326304

Bini, B. S., & Mathew, T. (2016). Clustering and Regression Techniques for Stock Prediction. Procedia Technol, 24, 1248-1255.

Bodie, Z., Kane, A., & Marcus, A. J. (2014). Investments (10th Edition). *McGraw-Hill*.

Cheong, D., Kim, Y. M., Byun, H. W., Oh, K. J., & Kim, T. Y. (2017). Using genetic algorithm to support clustering-based portfolio optimization by investor information. *Applied Soft Computing, 61*.

El, M., Guennoun, Z., & Hamza, F. (2012). Stocks Portfolio Optimization Using Classification and Genetic Algorithms. *Applied Mathematical Sciences, 6*(94). https://doi.org/10.1016/j.asoc.2017.08.042

Fama, E.F., & French, K.R. (2006). The value premium and the CAPM. *Journal of Finance, 61*(5). https://doi.org/10.1111/j.1540-6261.2006.01054.x

Graham, B., & Dodd, D. (1934). Security Analysis: The Classic (1st ed.). *McGraw Hill Professional*.

Hanson, D. (2013). ESG Investing in Graham & Doddsville. *Journal of Applied Corporate Finance, 25*(3). https://doi.org/10.1111/jacf.12024

Jensen, G.R., Johnson, R.R., & Mercer, J.M. (1997). New evidence on size and price-to-book effects in stock returns. Financial Analysts Journal, 53(6), 34–42.

Jensen, M.C. (1967). The Performance of Mutual Funds in the Period 1945-1964. The Journal of Finance, 23(2), 389–416.

Leivo, T. (2012a). Pricing Anomalies in the Finnish Stock Market. *Lappeenranta University of Technology, Lappeenranta*.

Leivo, T., & Pätäri, E. (2009). The Impact of Holding Period Length on Value Portfolio Performance in the Finnish Stock Markets. *Journal of Money, Investment and Banking, 8*(8). https://urn.fi/URN:NBN:fi-fe202101151879

Leivo, T.H. (2012b). Combining value and momentum indicators in varying stock market conditions: The Finnish evidence. *Review of Accounting and Finance, 11*(4). https://doi.org/10.1108/14757701211279187

Leivo, T.H., & Pätäri, E.J. (2011). Enhancement of Value Portfolio Performance Using Momentum and the Long-short strategy: The Finnish Evidence. *Journal of Asset Management, 11*(6). https://doi.org/10.1057/jam.2009.38

Leivo, T.H., Pätäri, E.J., & Kilpiä, I.J.J. (2009). Value enhancement using composite measures: The Finnish evidence. *International Research Journal of Finance and Economics, 33*.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance, 7*(1).

Miller, M., & Prondzinski, D. (2020). Value Style Investing Versus Growth Style Investing: Evidence from the 2002-2019 Business Cycle. *Journal of Accounting and Finance, 20*(1). https://doi.org/10.33423/jaf.v20i1.2748

Mirkin, B. (1996). Mathematical Classification and Clustering. *Journal of the Operational Research Society, 48*. https://doi.org/10.1007/978-1-4613-0457-9

Muangprathub, J., Intarasit, A., Boongasame, L., & Phaphoom, N. (2020). Portfolio Risk and Return with a New Simple Moving Average of Price Change.

Nanda, S.R., Mahanty, B., & Tiwari, M.K. (2010). Clustering Indian stock market data for portfolio management. *Expert System with Applications, 37*(12). https://doi.org/10.1016/j.eswa.2010.06.026

Nicholson, S.F. (1960). Price-to-earnings ratio. *Financial Analysts Journal, 16*(4). https://doi.org/10.2469/faj.v16.n4.43

Pätäri, E., & Leivo, T. (2009). Performance of the Value Strategies in the Finnish Stock Markets. *Journal of Money, Investment and Banking*. https://urn.fi/URN:NBN:fi-fe202101151879

Sharpe, W.F. (1964). Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk. *The Journal of Finance, 19*(3). https://doi.org/10.1111/j.1540-6261.1964.tb02865.x

Sharpe, W.F. (1966). Mutual Fund Performance. *The Journal of Business, 39*(1). https://doi.org/10.1086/294846

Zuhroh, I., Rofik, M., & Echchabi, A. (2021). Banking stock price movement and macroeconomic indicators: k-means clustering approach. *Cogent Business and Management, 8*(1). https://doi.org/10.1080/23311975.2021.1980247

This page intentionally left blank