ECONOMY AND FINANCE • VOL. 13 • ISS. 1 • MARCH 2026 • DOI: 10.33908/EF.2026.1.2

A HYBRID APPROACH TO INVESTMENT FUND ANALYSIS

Panel modeling and cluster analysis in the Hungarian market

László Vancsura – Katalin Plántek – Anett Parádi-Dolgos[1]

ABSTRACT

Our study aims to examine the performance and cost structure of Hungarian investment funds from 2017 to 2023. Using panel regression models, we analyzed the determinants of annual yield and the Total Expense Ratio (TER), with a particular focus on yield volatility, fund size, past performance, and asset class classification. Our results show that the level of risk, the current size of net asset value, and costs are significantly related to fund performance. Furthermore, our cluster analysis revealed that Hungarian funds can be divided into three distinct groups: conservative, stable-yield, and high-risk but potentially high-return funds. The findings emphasize that for investors, identifying costs and the risk profile is crucial for conscious portfolio construction. This study contributes to increasing financial literacy by showing how funds' internal characteristics and market strategies affect performance. From a methodological perspective, the research combines classical econometric panel modeling with elements of machine learning, offering a new approach to investment fund analysis. The results are relevant not only for domestic investors and fund managers but also for the international literature and regulatory environment, as they shed light on the heterogeneity of the market structure and the diversity of investment strategies.

JEL codes: G10, G23, M14

Keywords: investment funds, ESG, TER, panel regression, k-means clustering, financial awareness

1 *László Vancsura* Assistant research fellow, Department of Agricultural Logistics, Trade and Marketing; Institute of Agricultural and Food Economics; Hungarian University of Agriculture and Life Sciences; Kaposvár. E-mail: vancsura.laszlo@uni-mate.hu.
Katalin Plántek PhD-student, Doctoral School of Economics and Regional Sciences, Hungarian University of Agriculture and Life Sciences, Kaposvár Campus. E-mail: plantek.katalin@phd.uni-mate.hu.
Anett Parádi-Dolgos associate professor, Department of Investment, Finance and Accounting, Institute of Rural Development and Sustainable Economy, Hungarian University of Agriculture and Life Sciences, Kaposvár. E-mail: paradi-dolgos.anett.katalin@uni-mate.hu.

INTRODUCTION

Investment funds have become an integral part of both household and institutional investors' portfolios worldwide. These instruments enable diversified asset allocation and provide the benefits of professional asset management as well as access to international and sector-specific markets. The popularity of investment funds has increased particularly in low-interest-rate environments, where the pursuit of returns has placed greater emphasis on structured, cost-efficient, and transparent financial products (Célérier–Vallée, 2015). However, selecting an investment fund is a complex decision influenced by multiple factors, including past performance, risk level, cost structure, asset class, and the fund's geographic and thematic focus. In recent years, sustainability considerations – particularly Environmental, Social, and Governance (ESG) criteria – have emerged as an additional dimension. Their integration is gaining importance not only for ethical reasons but also for economic considerations (Kiymaz, 2019; Goncalves et al., 2021). Increasingly, research and investors recognize that funds incorporating ESG aspects do not necessarily underperform; under certain conditions, they can provide competitive returns with lower risk (Gupta–Chaudhary, 2023; Friede et al., 2015). At the same time, investors' financial literacy – the ability to interpret and weigh the risks and expected outcomes of various financial products – plays a crucial role in long-term financial well-being (Choowan et al., 2025). The criteria investors use in their decision-making are closely connected to their financial literacy. For example, it influences whether they prefer funds with lower TER and more sustainable long-term performance or are willing to pay higher fees in pursuit of potentially higher returns (Li et al., 2020; Atkinson et al., 2015). Both domestic and international evidence shows that costs significantly affect returns (Wermers, 2000; Levine, 2023), yet many investors remain insufficiently sensitive to this factor. The present study aims to provide a comprehensive analysis of the performance and cost structure of investment funds operating in Hungary between 2017 and 2023. The research focuses particularly on the factors influencing annual returns (Annual_yield) and the total expense ratio (TER). We examine return volatility, fund size, fund age, asset class type, as well as the impact of ESG and geographic classifications. Furthermore, the study seeks to explore how these attributes relate to different levels of financial literacy, including whether a cost-conscious investor selects different types of funds than a return-seeking but risk tolerant investor. This analysis contributes to a better understanding of the factors driving investor decisions and the relationships between product characteristics, performance, and cost structure. The findings can provide valuable guidance not only to investors for informed decision-making but also to regulators and fund managers in developing a transparent and competitive product offering.

1.1 Determinants of Investment Fund Performance

Interest in investment fund returns and their determinants spans from classical efficiency theories (e.g., Jensen, 1968; Sharpe, 1966) to modern behavioral finance. The volatility of returns, as a fundamental measure of risk, is closely associated with achieved performance (Fama–French, 1993). Recent studies indicate that funds with higher volatility, particularly in certain asset classes such as equity or commodity funds, may generate higher average returns (Baker et al., 2020; Zhang–Kuo, 2021). At the same time, the relationship between NAV (Net Asset Value), as a proxy for fund size, and performance is complex. Some research suggests that larger funds benefit from economies of scale and can operate with lower transaction costs (Adams et al., 2022; Chen et al., 2004), while others link excessive size to reduced flexibility and return potential (Al Shaekh et al., 2024). Fund age, reflecting managerial experience or operational stability, may also be relevant in predicting performance, although empirical results are mixed. For instance, Chen et al. (2004) found that more mature funds produce more stable returns but adapt less rapidly to market changes.

Singh and Tandon (2021) examined 81 Indian investment funds in the 2013–2019 period to identify which fund characteristics influence performance. The variables included net asset value (NAV), portfolio turnover ratio (PTR), fund size (AUM), total expense ratio (TER), and fund age (Age). Relationships between fund characteristics and performance were estimated using fixed-effects panel regression. In the first analysis, gross return was used as a performance measure, revealing significant associations with NAV, PTR, AUM and Age. In the second analysis, risk-adjusted performance metrics (the Treynor ratio and Jensen's alpha) were employed, concluding that NAV and TER significantly affected fund performance. Junaeni (2022) investigated the effects of cash flow (CF), fund size (AUM), total expense ratio (TER), and portfolio turnover (PTR) on fund performance using a random-effects panel regression on data for 30 Indonesian funds from 2015 to 2017. Only cash flow exhibited a significant positive effect, while all other variables, including TER, were not significant. Babbar and Sehgal (2018) analyzed 237 open-ended Indian equity funds from 2007 to 2013 to determine which characteristics affect fund performance. The variables of interest were fund size (AUM) and its growth, total expense ratio (TER), portfolio turnover (PTR), net asset value (NAV), and fund age (Age). Using fixed effects panel regression, they found that fund age positively affected performance, whereas AUM, its growth, and NAV showed negative relationships. TER and PTR were not significantly associated with performance. Filip (2018) examined 152 Polish investment funds from 2002 to 2015, including fund size (AUM), total expense ratio (TER), and fund age (Age) as explanatory variables in a panel data analysis. He found that AUM and TER

positively influenced fund performance, while fund age had a negative effect. He concluded that the efficiency of domestic equity fund management is less dependent on size and age than on cost ratio, in contrast to foreign equity funds, where size and age are more decisive factors. Inês Sá et al. (2024) analyzed the performance of Portuguese investment funds investing in domestic and foreign equities over the 2005–2022 period using a robust six-factor model. They employed a fixed effects panel regression to measure the impact of fund characteristics on performance, incorporating fund age (Age), family size, total expense ratio (TER), fund size (AUM), and capital inflows (flows). Most domestic equity funds showed neutral results, whereas foreign equity funds underperformed. Fund age negatively affected performance, whereas TER had no significant association.

1.2 The Role of Costs in Investment Fund Performance

The total expense ratio of investment funds is one of the most important, yet often underestimated, factors in investor decision-making. Over the long term, costs can significantly reduce net returns, particularly in passive or low-turnover funds (Livingston, 2019). Empirical research consistently shows that funds with a high TER do not necessarily perform better; in fact, higher costs often predict weaker performance (Mansor et al., 2015; Sofi–Yahya, 2019; ESMA, 2019). At the same time, in actively managed funds that follow complex strategies, a higher TER may be justified if it generates genuine added value (Sheng et al., 2019). Investor sensitivity to costs is, however, not uniform. Evidence suggests that investors with lower financial literacy pay less attention to TER, whereas more financially aware participants actively seek alternatives with lower fees (Jiang et al., 2020; Sholl–Fontes, 2022).

1.3 Financial Literacy and Investment Decisions

The importance of financial literacy is increasingly recognized in the context of investment decision-making. The OECD (2020) defines financial literacy as a combination of skills and knowledge that enables investors to make informed and responsible decisions in managing their finances. Numerous empirical studies document that financially literate individuals are more likely to compare fund costs, evaluate past returns, and avoid funds with unjustifiably high fees (Hastings et al., 2013; Lusardi et al., 2017; Lusardi–Mitchell, 2014). The impact of financial literacy is particularly important for investment funds, where the complexity of products and fee structures can be difficult for less experienced investors to understand (Nilsson et al., 2024). Some studies attempt to capture the level

of financial literacy using specific indicators, for example whether investors prefer funds with low TER, long-term stable returns, or lower volatility (Choi et al., 2010). These behaviors can be considered indicators of conscious investor behavior, which can even be aggregated into indices (Müller–Weber, 2010; Grinblatt et al., 2016).

1.4 ESG and Sustainability Considerations

The rise of ESG (Environmental, Social, Governance) funds has introduced a new dimension to research on investment funds. Several studies in recent years have examined the performance and cost structure of funds with ESG characteristics. The findings suggest that these funds often operate with higher fees, yet increasing evidence indicates that they can offer competitive or even lower-risk performance (ESMA, 2022; Baker et al., 2022). Among investors, the role of ESG ratings becomes particularly pronounced for those who consciously seek socially responsible and sustainable alternatives – further underscoring the importance of financial literacy (Deka et al., 2023). This aspect is especially relevant for the present study, as the presence of ESG factors reflects not only market dynamics but also behavioral patterns. Both domestic and international literature confirm that the performance and costs of investment funds are influenced by multiple factors, of which volatility, fund size, age, and geographic and asset class classification are the most significant. The behavioral and structural aspects of financial literacy are manifested in cost sensitivity, asset selection, and preferences for long-term outcomes. The aim of this study is to quantitatively examine these factors and to explore the patterns observable in the relationships between return, risk, cost, and informed choices using the example of Hungarian investment funds.

2. MATERIALS AND METHODS

2.1 Regression model, Data, Variables and descriptive statistics

In our study, we examined the investment funds managed by Hungarian fund managers between 2017 and 2023 using a panel regression approach. The data were downloaded from the publicly available databases of the Hungarian National Bank and the Association of Hungarian Investment Fund and Asset Management Companies (BAMOSZ). For the modeling, we used the following variables:

- Annual_yield: the annual return of the investment fund expressed as a percentage.

- Yield_from_start: the cumulative return of the investment fund expressed as a percentage, calculated from the fund's inception until the end of the examined year.
- Previous_yield: the return achieved in the previous year, expressed as a percentage.
- Annual_std: the standard deviation of the returns in the current year.
- NAV_in_mrdFt_actual_year: the net asset value at the end of the year, expressed in billion HUF.
- Age_in_months: the age of the fund measured in months since inception.
- TER: the total expense ratio of the investment fund in the given year.
- ESG: a dummy variable representing the ESG rating.
- Emerging_market_dummy: a variable indicating whether the equities in the fund are geographically located in emerging markets.
- Developed_markets_dummy: a variable indicating whether the equities in the fund are geographically located in developed markets.
- The funds were categorized according to their investment policy into the following groups: Absolute_yield, Commodity, Real_estate, Bond, Money_market, Equity and Others.

To test for heteroskedasticity in the panel models, we applied the Breusch–Pagan and the White tests. For most models, the test results indicated the presence of heteroskedasticity; therefore, we had to employ robust estimation methods. We also examined multicollinearity, which was not present in any of the models, so we did not need to modify the originally designed combination of dependent and independent variables. However, extreme return values required treatment (*Figure 1*), and we restricted our sample to observations with annual returns between –50% and +50%. For the analysis, we used Stata version 17.

The descriptive statistics presented (*Table 1*) refer to the sample of 1,967 observations and provide an overview of the characteristics of the investment funds along several dimensions. The most important variables include annual return (Annual_yield), return volatility (Annual_std), net asset value (NAV) in the current and previous year, past returns, fund age, and the total expense ratio (TER).

Table 1
Descriptive statistics of the investment funds included in the study for the period between 2017 and 2023

Variable	Obs.	Mean	Std.	Min	Max
Annual_yield	1967	4.772	11.160	–34.130	49.300
Annual_std	1967	6.009	4.678	0.140	21.68
NAV_in_mrdFt_actual_year	1967	14.957	46.789	0.002	659.108
NAV_in_mrdFt_prev_year	1967	12.954	39.849	0.000	590.264
Previous_yield	1967	3.392	9.996	–34.130	49.300
Yield_from_start	1967	3.904	3.815	–7.880	37.910
Age_in_months	1967	147.785	73.266	24	360
TER	1967	1.583	0.704	0.039	4.533
ESG	154	–	–	–	–
Absolute_yield	658	–	–	–	–
Commodity	42	–	–	–	–
Real_estate	49	–	–	–	–
Bond	336	–	–	–	–
Money_market	28	–	–	–	–
Equity	525	–	–	–	–
Others	329	–	–	–	–

Source: authors' elaboration based on STATA17 output

The average annual return (Annual_yield) was 4.772%, but its standard deviation (11.16%) as well as minimum (–34.13%) and maximum values (49.3%) indicate an extremely wide dispersion. This shows that while most funds achieved positive returns, performance was highly variable: some funds suffered substantial losses, while others realized extremely high gains over the sample period. The negative minimum and exceptionally high maximum highlight the diversity of risks and the impact of potential extreme market movements. The annual standard deviation (Annual_std), as a key indicator of return volatility, averaged 6.009%, with a relatively high dispersion (4.678). The lowest value (0.14) likely corresponds to a very stable fund, potentially a money market or bond fund, whereas the highest value (21.68) points to funds carrying substantial risk, such as equity or commodity funds.

The average net asset value (NAV) was 14.957 billion HUF in the current year and 12.954 billion HUF in the previous year. The minimum NAV values were close to zero, while the maximums (659 billion HUF in the current year, 590 billion HUF

in the previous year) reflect the large asset holdings of a few dominant players in the industry. Such wide dispersion suggests that the sample is highly heterogeneous, including both smaller, lesser-known funds and large, dominant market participants. The average past returns (Previous_yield and Yield_from_start) were 3.392% and 3.904%, respectively. The average age of the funds (Age_in_months) was 147 months, or about 12 years, which indicates relatively mature funds. However, the minimum value (24 months) shows that newly established funds were also included in the sample, while the maximum (360 months, i.e., 30 years) illustrates that several funds have a long track record. This diversity allows for an examination of the effects of long-term performance and fund maturity. The average TER was 1.583%, which corresponds to a medium level by international standards. The minimum was extremely low (0.039%), while the maximum (4.533%) is notably high, likely characterizing a fund employing complex or alternative investment strategies. The standard deviation of 0.704 signals considerable heterogeneity in costs, which may play an important role in investment decision-making. Overall, the descriptive statistics demonstrate that the Hungarian investment fund market is highly diverse in terms of returns, risks, size, and costs. The data reveal that the market simultaneously includes stable, low cost, low volatility funds, as well as riskier products that may offer higher returns but come with higher expenses. This diversity provides investors with a broad set of opportunities, while also underlining the need for careful analysis when making investment decisions.

2.2 Principal Component Analysis

Principal Component Analysis (PCA) is one of the most widely applied dimensionality reduction techniques in statistics and machine learning. Its primary objective is to project high-dimensional data into lower-dimensional spaces through linear transformation while minimizing information loss. A key advantage of PCA lies in its capacity for visualization, which facilitates the identification of trends, patterns, and outliers. The foundations of the method were established by Pearson (1901) and later developed by Hotelling (1933). Today, PCA is employed across numerous scientific disciplines, including medicine, environmental science, and economics. Oprea (2022) applied PCA to investigate the dynamics of Romanian government bond yield curves, demonstrating that 97% of yield curve variations could be explained by three components. In their stock market analysis, Yang et al., (2017) showed that even components with the smallest eigenvalues may contain relevant financial information. Zoričić et al., (2020) utilized PCA for the optimization of fundamentally weighted portfolios, constructing a minimum-variance portfolio.

2.3 Cluster Analysis

The k-means cluster analysis is one of the most widely applied methods in data mining. As part of the procedure, the original dataset is partitioned into a predefined number of clusters, denoted by k. The most critical element of the method is the initial choice of k, as this determines how many groups the algorithm will divide the data into. Choosing too low or too high a value may cause problems (Ahmed et al., 2020) and distort the results. During the analysis, each cluster is assigned a centroid, and each data point is allocated to the nearest one. After the initial clustering, the algorithm recalculates a new centroid for each cluster based on the mean of the assigned data points. These steps are then repeated iteratively until the centroids no longer change significantly.

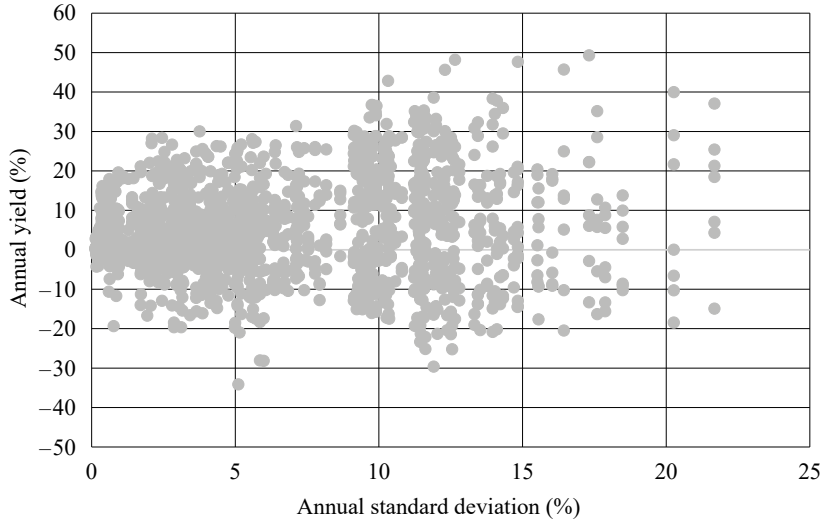
There are several approaches for selecting the optimal value of k, such as the elbow method, the silhouette coefficient, the gap statistic, and the canopy method (Yuan–Yang, 2019). In our analysis, we applied the elbow method and the silhouette coefficient. The elbow method is primarily a visual approach: it involves plotting the within-cluster sum of squares (the sum of squared distances between data points and their centroid) against different values of k. On the resulting graph, one should look for a breakpoint, referred to as the "elbow," which indicates that further increases in k no longer yield significant improvements in the fit. This value of k is then chosen for the analysis. However, identifying the "elbow" is not always straightforward. By contrast, the silhouette coefficient method – originally proposed by Rousseeuw (1987) – provides a more objective approach. In this method, a silhouette coefficient is calculated for each data point based on two measures: cohesion (the similarity of a data point to its own cluster) and separation (its dissimilarity to other clusters). Using these values, the silhouette coefficient is computed according to a specified formula, with values ranging from –1 to 1. The closer the value is to 1, the better the clustering quality. In practice, it is advisable to combine these methods: the elbow method offers a useful starting point, while the silhouette coefficient refines the clustering solution. K-means clustering can also be applied in portfolio selection and portfolio optimization processes (Bolos et al., 2025; Sukono et al., 2024; Tenkam et al., 2022).

3 RESULTS

3.1 Exploratory Data Analysis

Figure 1 illustrates the relationship between annual return and annual standard deviation (risk) at the level of individual investment funds. In general, a positive correlation can be observed between risk and return: funds with higher volatility (greater risk) typically display a wider range of returns, including higher positive outcomes, but also carry a greater likelihood of negative returns. Conversely, funds with lower volatility tend to exhibit returns within a narrower range, implying lower return potential but also a reduced probability of losses. The figure visually reinforces the well-known risk–return trade-off in finance, whereby achieving higher potential returns requires accepting greater uncertainty (risk).

Figure 1
Relationship between annual volatility and returns of the investment funds included in the analysis over the period 2017–2023

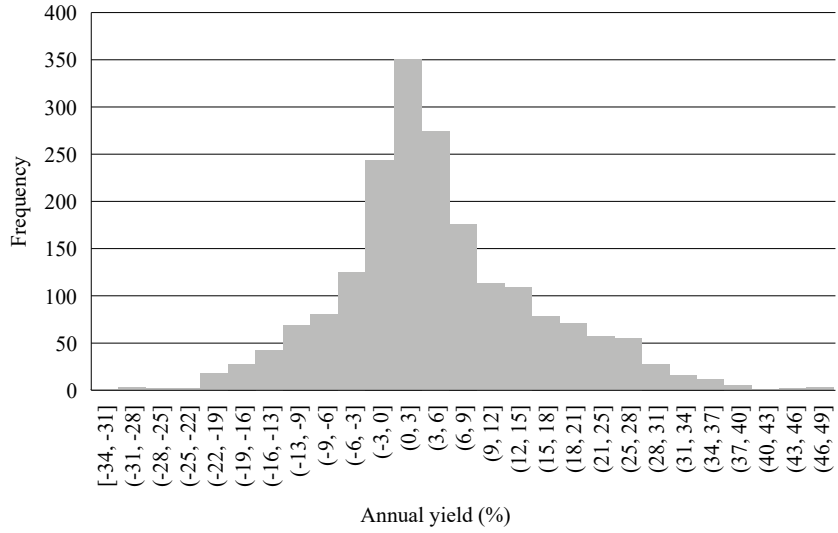


Source: authors' elaboration

The histogram of annual returns (*Figure 2*) illustrates that the distribution of returns among the examined funds is predominantly concentrated around values close to zero, indicating a high frequency of returns near the median. The distribution exhibits a slight right skew (positive skewness), pointing to the presence of funds with higher positive returns that extend the distribution in the positive

direction. While the majority of returns fall within the range of –20% to 20%, several funds display exceptionally high or low returns, which may reflect extreme market events or the outcomes of specific investment strategies. This distribution highlights the variability of investment fund returns and the potential for extreme outcomes.

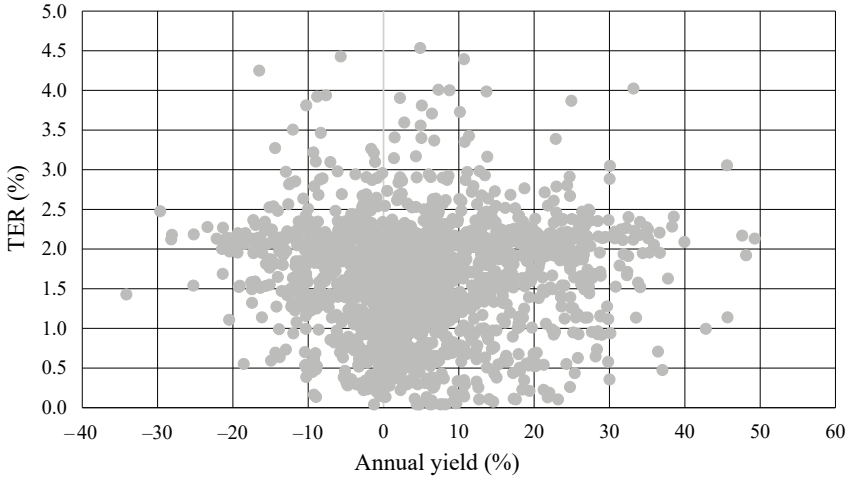
Figure 2
Distribution of annual returns among the examined investment funds (histogram)



Source: authors' elaboration

Figure 3 illustrates the relationship between the total expense ratio and annual returns for individual investment funds. Visually, no strong linear relationship is apparent between TER and annual returns; the data points appear relatively scattered across the plot. This suggests that higher costs do not necessarily correspond to higher returns, and vice versa. Although a large number of funds with lower TER values are concentrated in the lower portion of the figure, their returns span a wide range, similar to funds with higher TER. Based on this figure alone, TER does not appear to be a reliable predictor of annual return; however, expenses remain an important factor influencing net returns.

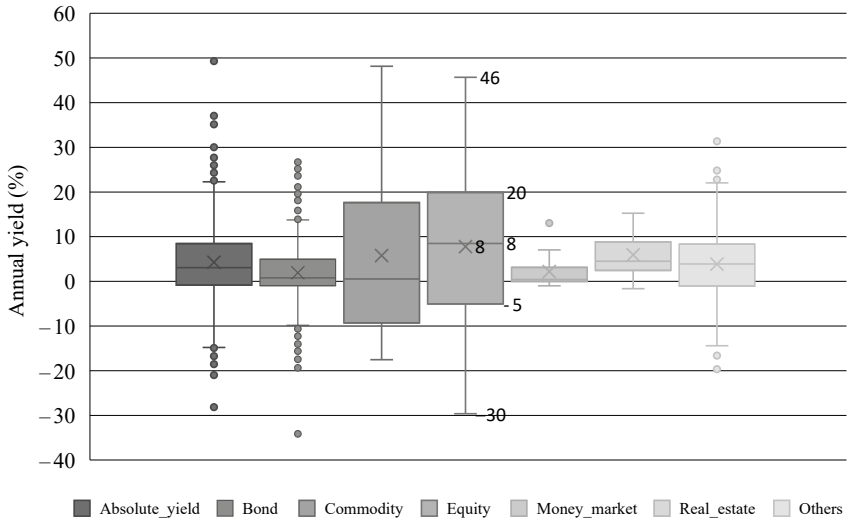
Figure 3
Relationship between annual returns and total expense ratio (TER) among the examined investment funds



Source: authors' elaboration

The boxplot analysis (*Figure 4*) of annual returns across different asset classes reveals clearly distinguishable return distributions and risk profiles. Equity funds exhibit the highest median annual returns, accompanied by a wide interquartile range and numerous outliers, indicating substantial upside potential alongside notable volatility. Bond funds generally show lower median returns and narrower interquartile ranges compared with equities, reflecting their typically lower risk profile, although outliers indicate the potential for significant gains and losses. Money market funds display the narrowest distribution, with low median returns and minimal dispersion, consistent with their primary objectives of capital preservation and low risk. Absolute yield funds demonstrate a wide distribution with a median close to zero, highlighting their diverse strategies and variable performance. Real estate funds present relatively stable positive median returns with moderate variability. Commodity funds show highly dispersed returns across a broad range, reflecting their inherent volatility and sensitivity to market fluctuations. Finally, the "Others" category exhibits a wide distribution, encompassing a variety of strategies and corresponding return patterns. These observed differences in return distributions underscore the distinct risk–return characteristics inherent in different asset classes, providing empirical support for asset allocation decisions in portfolio construction.

Figure 4
Distribution of annual returns across asset classes (boxplot)



Source: authors' elaboration

3.2 Panel OLS Analysis

The three presented panel OLS models (*Table 2*) examine the determinants of annual returns (Annual_yield) for investment funds. The objective of these models is to explore how returns relate to fund characteristics, risk profiles, size, past performance, costs, market focus, and asset class allocation. The three specifications progressively incorporate a broader set of explanatory variables, with R^2 values indicating a modest but consistent increase in explained variance (0.1957 → 0.2281), reflecting the models' consistency and stability. In all models, there is a highly significant and positive relationship between annual return and annual volatility (Annual_std), aligned with classical financial theory (e.g., CAPM): funds with higher volatility offer higher returns, implying that investors expect compensation for taking on greater risk. In Model 3, however, this relationship weakens and loses significance, suggesting that additional control variables (e.g., asset class types) partially account for the effect of volatility. The net asset value for the current year (NAV_in_mrdFt_actual_year) consistently exhibits a positive effect on returns, whereas the previous year's NAV (NAV_in_mrdFt_prev_year) has a negative impact. This supports the mean reversion hypothesis, which predicts that the effects of prior-year capital flows or asset movements may manifest in the

opposite direction in the subsequent year. The negative and significant coefficient of previous returns (Previous_yield) further indicates temporal dynamics: high past performance may be associated with declining future returns, in line with the concept of performance normalization. Conversely, the cumulative return from inception (Yield_from_start) shows a positive and significant effect, pointing to a potential sustained competitive advantage for long-term successful funds. Fund age (Age_in_months) becomes significant only in Model 3, and even there only weakly, implying that maturity alone is not a decisive determinant of returns but may be important when combined with certain market or category effects. Of particular note is the effect of the total expense ratio (TER): it is not significant in the first two models but has a strong negative effect in Model 3. This suggests that, when controlling for asset classes, high-cost funds – especially those investing in more passive instruments – tend to deliver lower net performance, aligning with cost-sensitive investor behavior. Regarding regional focus, the dummy variable for emerging market funds exhibits a significant negative effect in Model 2, whereas the developed market focus is not significant. This may reflect global risk distribution and the more stable performance of developed markets. Finally, the asset class dummy variables in Model 3 reveal a distinct pattern. Commodity and equity funds show the highest positive returns, while bond and money market funds have a significant negative impact. These findings are consistent with the historical return profiles of different asset classes: riskier, growth-oriented assets typically provide higher long-term returns. Absolute yield and real estate funds also exhibit positive effects, albeit to a lesser extent.

Table 2
Panel OLS regression results on the determinants of annual fund returns (Annual_yield) between 2017 and 2023

Variables	(1) Annual_yield	(2) Annual_yield	(3) Annual_yield
Annual_std	0.492*** (0.074)	0.497*** (0.071)	0.128 (0.121)
NAV_in_mrdFt_actual_year	0.086*** (0.018)	0.091*** (0.018)	0.090*** (0.018)
NAV_in_mrdFt_prev_year	–0.092*** (0.023)	–0.093*** (0.024)	–0.103*** (0.023)
Previous_yield	–0.377*** (0.034)	–0.384*** (0.033)	–0.404*** (0.034)
Yield_from_start	0.705*** (0.213)	0.730*** (0.206)	0.793*** (0.228)
Age_in_months	0.006 (0.004)	0.008 (0.005)	0.010* (0.005)

TER	−0.046 (0.337)	−0.082 (0.347)	−0.882** (0.372)
ESG	0.508 (0.862)	−0.276 (0.815)	−0.056 (0.838)
Emerging_market_dummy		−1.974*** (0.522)	
Developed_markets_dummy		1.201 (0.784)	
Absolute_yield			0.630 (0.499)
Commodity			8.105*** (2.109)
Real_estate			3.076*** (1.039)
Bond			−4.302*** (0.760)
Money_market			−2.218*** (0.786)
Equity			3.797*** (1.088)
Constant	−0.605 (0.639)	−0.212 (0.663)	1.523** (0.742)
Observations	1967	1967	1967
R^2	0.1957	0.2051	0.2281

Robust standard errors for OLS are reported in parenthesis.
***p< 0.01; **p < 0.05; *p < 0.1.
Source: authors' elaboration

In summary, annual returns are significantly influenced by multiple factors, including risk level, fund size, past performance, asset class, and cost structure. The models provide a solid foundation for understanding investment fund performance; however, based on the R^2 values, a substantial portion of returns may still be explained by other unobserved or time-varying factors. Therefore, further analyses should consider incorporating additional variables, such as macroeconomic indicators, management efficiency, or performance relative to benchmarks.

The three presented regression models (*Table 3*) explore the determinants of the total expense ratio of investment funds. The explanatory variables included in the models cover multiple dimensions: return and risk indicators, fund size, past performance, ESG rating, and asset class classification. The three specifications progressively expand the set of explanatory variables, with R^2 values increasing from 0.2123 to 0.2984, indicating improved model fit. Across the models, TER exhibits a significant and positive relationship with fund volatility (Annual_std), suggesting that higher-risk (more volatile) funds generally operate with higher costs. This

is consistent with market experience, as actively managed funds employing complex strategies tend to charge higher fees to investors. Annual returns (Annual_yield), however, are not significant predictors of TER in any model, implying the absence of a direct relationship between returns and fee levels; higher-fee funds do not necessarily deliver higher returns. The NAV variables show contrasting effects: current-year NAV is always positive and significant, whereas prior-year NAV exhibits a weak, negative, and nonsignificant effect. This pattern suggests that costs are more closely tied to the current asset base than to past holdings. The positive coefficient also implies that larger funds – despite potential economies of scale – do not necessarily operate with lower TER, possibly due to marketing or operational expenses. Regarding past performance variables, Previous_yield is only marginally significant at the 10% level in Model 3, while cumulative returns from inception (Yield_from_start) show a significant negative relationship with TER, suggesting that longer-term successful funds may sustain lower costs or attract greater investor confidence, reducing the need for high fee levels. ESG ratings display a significant positive effect across models, supporting the industry observation that ESG funds often carry higher fees, likely due to additional screening, evaluation, and sustainability requirements. In Model 3, this effect is slightly weaker, but remains significant, indicating that the relationship may partially depend on asset class. The regionally focused dummy variables in Model 2 show mixed results. Funds targeting developed markets exhibit slightly lower TER with weak significance, while funds focused on emerging markets have no significant effect, suggesting that regional focus alone does not fully explain cost differences. Model 3 further disaggregates by asset class, revealing their clear impact on TER. Absolute return, commodity, real estate, and equity funds operate with significantly higher TER, whereas money market funds have substantially lower costs. This aligns with market practice, where simpler, passively managed, or short-term liquidity-focused funds are generally cheaper, while complex, actively managed strategies (e.g., commodity or real estate) command higher fees. Bond fund costs do not differ significantly from the baseline category.

Table 3
Panel OLS regression results on the determinants of total expense ratio (TER) between 2017 and 2023

Variables	(1) TER	(2) TER	(3) TER
Annual_yield	−0.000	−0.001	−0.001
	(0.001)	(0.001)	(0.001)
Annual_std	0.070***	0.071***	0.049***
	(0.008)	(0.009)	(0.017)
NAV_in_mrdFt_actual_year	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)
NAV_in_mrdFt_prev_year	−0.000	−0.000	−0.000
	(0.001)	(0.001)	(0.001)
Previous_yield	−0.001	−0.001	−0.002*
	(0.001)	(0.001)	(0.001)
Yield_from_start	−0.020***	−0.019***	−0.018**
	(0.007)	(0.007)	(0.008)
Age_in_months	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
ESG	0.237**	0.252***	0.170*
	(0.095)	(0.091)	(0.095)
Emerging_market_dummy		−0.100	
		(0.073)	
Developed_markets_dummy		−0.183*	
		(0.098)	
Absolute_yield			0.348***
			(0.098)
Commodity			0.516**
			(0.242)
Real_estate			0.678***
			(0.184)
Bond			−0.104
			(0.097)
Money_market			−0.535***
			(0.106)
Equity			0.366**
			(0.177)
Constant	1.173***	1.202***	1.057***
	(0.076)	(0.077)	(0.116)
Observations	1967	1967	1967
R^2	0.2123	0.2201	0.2984

Robust standard errors for OLS are reported in parenthesis.
***p< 0.01; **p < 0.05; *p < 0.1.
Source: authors' elaboration

In summary, the presented models indicate that the variation in TER is primarily explained by the fund's risk profile, current size, long-term performance, sustainability characteristics, and asset class classification. Return, fund age, and past asset values are individually less relevant. The detailed breakdown provided by Model 3 is particularly useful for understanding which types of funds are associated with higher cost levels, offering investors valuable insights for assessing the price–value relationship. However, the models only partially explain the variance in TER, suggesting that incorporating additional factors (such as benchmarks, passive versus active management, marketing expenditure, or management fee structures) could improve estimation accuracy.

3.3 Cluster Analysis

The principal component analysis (PCA) extracted two principal components, which explain a substantial portion of the variance in the examined financial characteristics (TER, Annual_std, Annual_yield) (Table 4). The first principal component (PC1) exhibits strong positive loadings on TER (0.6565) and Annual_std (0.6912), indicating that this component primarily captures information related to fund costs and risk. Higher PC1 values correspond to higher costs and volatility. In contrast, the second principal component (PC2) has a pronounced loading on Annual_yield (0.9374), while the contributions of TER (-0.3365) and Annual_std (-0.0900) are smaller and in the opposite direction. This suggests that PC2 predominantly represents the annual return of the funds, with higher PC2 values associated with higher returns. In summary, PC1 can be interpreted as the risk/cost dimension and PC2 as the return dimension, facilitating the understanding of the multidimensional characteristics of investment funds within a lower dimensional space.

Table 4
Principal component (PCA) loadings on investment funds' cost, risk, and return characteristics

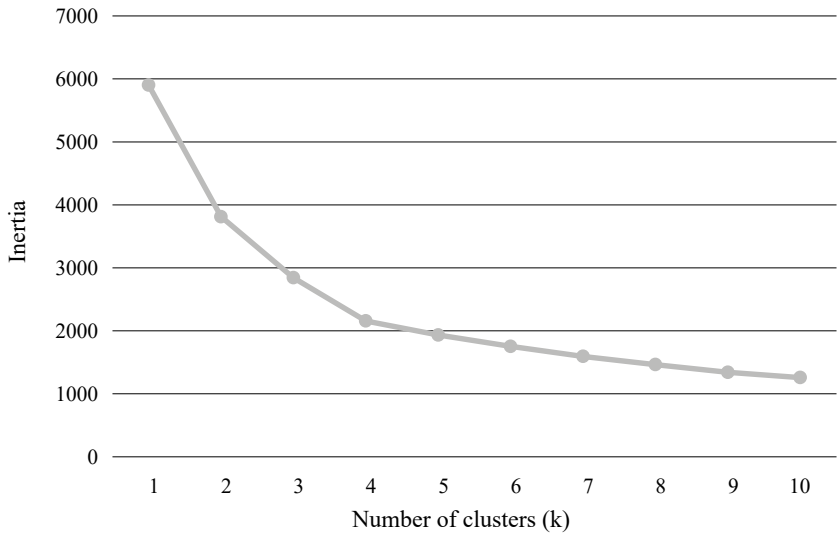
PC	TER	Annual_std	Annual_yield
PC1	0.6569	0.6924	0.2985
PC2	–0.3374	–0.0840	0.9376

Source: authors' elaboration

Figure 5 illustrates the results of the elbow method applied to k-means clustering, showing the inertia values corresponding to different numbers of clusters (k). Inertia represents the sum of within cluster deviations, which decreases con-

tinuously as the number of clusters increases. Based on the observed values, the most pronounced "elbow" occurs around 3-4 clusters, where further increases in the number of clusters yield only marginal improvement. This suggests that the optimal number of clusters in this case is likely 3 or 4, providing an appropriate balance between simplicity and clustering performance.

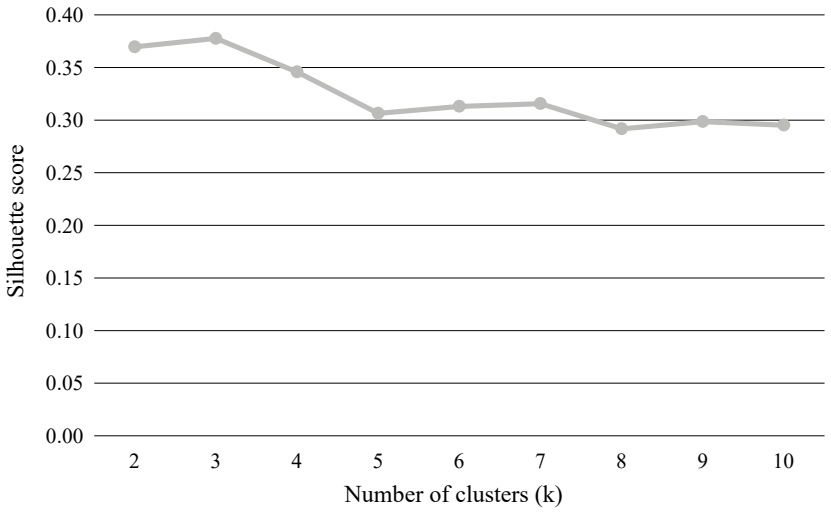
Figure 5
Elbow method for determining the optimal number of clusters



Source: authors' elaboration

Figure 6 presents the silhouette scores corresponding to different numbers of clusters, which measure both the separation between clusters and the internal cohesion within clusters. Higher values indicate better defined, more distinct clusters. The results show that the highest score occurs at three clusters (0.378), slightly higher than the two-cluster solution. As the number of clusters increases further, the silhouette scores gradually decline, indicating that additional clustering may reduce overall clustering quality. Therefore, the optimal number of clusters is identified as three.

Figure 6
Silhouette method for determining the optimal number of clusters



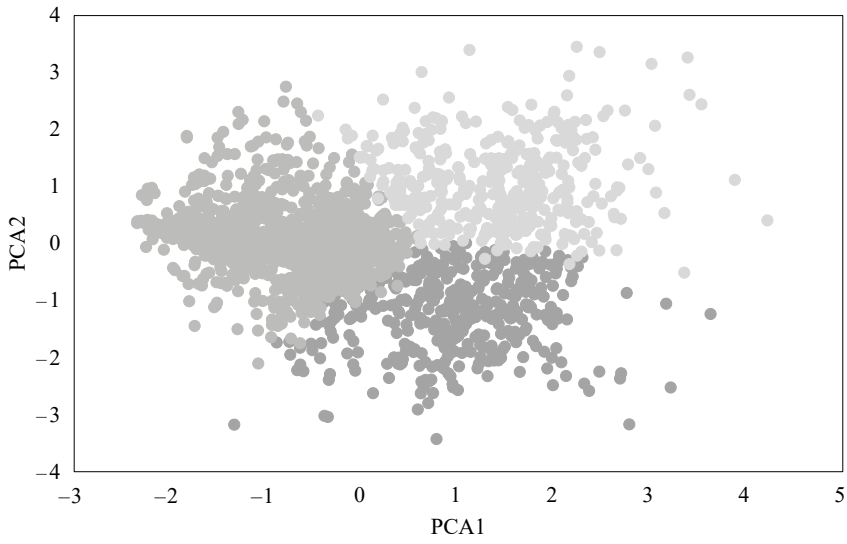
Source: authors' elaboration

To determine the optimal number of clusters, both the elbow and the silhouette methods were applied. The elbow curve indicates a noticeable inflection point around the three- to four-cluster solutions, reflecting a substantial improvement in clustering efficiency compared with lower cluster counts. Meanwhile, the silhouette score reaches its maximum at the three-cluster configuration, supporting significant separation between clusters and strong internal cohesion. The empirical results consistently suggest that a three-cluster solution is optimal for the dataset under study, robustly confirmed by both methodological approaches.

The cluster analysis (*Figure 7* and *Table 5*) identified three clearly distinguishable investment fund profiles. The first cluster ("Expensive and Loss-Making") comprises funds with the highest cost ratios (average 2.18%) and substantial risk (10.5% annual standard deviation), while simultaneously exhibiting negative average returns (–5.4%). For these funds, the high fees were not accompanied by adequate performance, resulting in a loss of value for investors. Extreme negative returns (minimum value: –34.1%) suggest that these funds are particularly sensitive to market shocks and suboptimal management strategies. The second cluster ("Low-Cost and Stable") is characterized by low costs (1.25%) and moderate risk (2.9%), while providing an average positive return of 2.9%. The low volatility and stable performance indicate that this group embodies a conservative, cost-efficient investment strategy. Although the returns are moderate, the risk–return

profile is particularly favorable for risk–averse investors. This is further supported by maximum returns reaching up to 30%, demonstrating that even with restrained risk taking, above average performance is still attainable. The third cluster ("High–Risk but Profitable") exhibits a medium cost level (1.96%), high risk (10.3%), and outstanding average returns (20.5%). While the mean returns are very attractive, the high standard deviation (7.8%) highlights the considerable uncertainty of outcomes. Some funds in this group achieved exceptionally high returns (up to 49.3%), likely resulting from successful market positioning and active risk taking. However, the extreme variability also implies that these funds are more suitable for speculative investors willing to assume higher risk in pursuit of maximum returns. Overall, the results demonstrate marked differences in cost, risk, and return among investment funds, allowing for the clear identification of fund groups matching distinct investor preferences.

Figure 7
Clustering of investment funds based on cost, risk, and return characteristics



Source: authors' elaboration

Table 5
Cluster analysis summary statistics

Cluster	TER				Annual_std				Annual_yield			
	mean	std	min	max	mean	std	min	max	mean	std	min	max
Expensive and Loss-Making	2.1768	0.5690	0.5916	4.5329	10.5324	3.8541	0.6300	21.6800	–5.4216	8.2033	–34.1300	10.7000
Cheap and Stable	1.2476	0.5901	0.0389	2.9031	2.9320	2.0301	0.1400	10.8000	2.9466	6.4267	–19.3600	30.0200
Risky but Profitable	1.9552	0.5273	0.2267	4.0221	10.3465	3.6802	2.0200	21.6800	20.5484	7.8116	7.0800	49.3000

Source: authors' elaboration

The results of the cluster analysis (*Table 6*) thus identified three distinctly different investment fund profiles, clearly reflecting the trade-offs between cost, risk, and return. The "Expensive and Losing" group's negative performance exemplifies that high costs alone do not guarantee higher returns, especially when shortcomings in risk management are also present. By contrast, the "Cheap and Stable" cluster shows that a combination of low costs and prudent risk taking can yield moderate but reliable performance. These two groups stand in sharp contrast: one results in losses for investors, while the other provides predictable long-term capital preservation. The third cluster, "Risky but Profitable," represents the classic relationship between high risk and high return, which is typically associated with more speculative investment strategies. Here, the return potential is significantly higher, but the risk arising from volatility is also substantially elevated, making it acceptable only for risk-tolerant investors. Overall, the comparison of the three clusters highlights that the investment fund market can be segmented according to different investor preferences: some seek stability and low costs, others prioritize outstanding returns achievable at higher risk, while the avoidable category serves as a warning regarding the dangers of inefficient fund structures.

Table 6
Summary of characteristics of the three investment fund clusters

Cluster Name	Key Characteristics	Typical Investor Profile
Expensive and Loss-Making	High TER (avg. 2.18%) High risk (avg. 10.53%) Negative average return (avg. –5.42%)	Risky, costly funds that on average result in a loss. Generally, not recommended for investors.
Cheap and Stable	Low TER (avg. 1.25%) Moderate risk (avg. 2.93%) Positive but moderate return (avg. 2.95%)	Ideal for conservative, risk averse investors. Provides stable performance at low costs.
Risky but Profitable	Medium TER (avg. 1.96%) High risk (avg. 10.35%) Outstanding return (avg. 20.55%, max. 49.3%)	Recommended for speculative, return-maximizing investors willing to take high risk for high potential gains.

Source: authors' elaboration

4 DISCUSSION

The results of the analysis offer several important insights into the performance, cost structure, and role of financial literacy in the Hungarian investment fund market. Based on the models and cluster analyses, it is evident that the evolution of annual yield (Annual_yield) and the total expense ratio (TER) is not determined by a single factor but rather by the complex interplay of multiple factors. This finding corroborates previous international research, which indicates that fund performance is influenced by a combination of volatility, size, fund age, and asset class classification (Fama–French, 1993; Singh–Tandon, 2021; Junaeni, 2022; Filip, 2018). The results highlight the critical role of cost structure in long-term returns. High TER values do not guarantee higher returns; indeed, the cluster analysis shows, as exemplified by the "Expensive and Unprofitable" group, how excessive fees are often associated with weaker performance. This aligns with international evidence reported by ESMA (2019) and Mansor et al., (2015), indicating that costs can significantly reduce net returns over time, especially in actively managed funds that fail to create sufficient added value. In contrast, the "Cheap and Stable" cluster demonstrates that a combination of low costs and moderate risk can provide reliable and sustainable returns, representing an ideal choice for cost–conscious and risk–averse investors. The risk–return relationship becomes particularly apparent in the "Risky but Profitable" cluster. This group assumes high risk while offering outstanding average returns, illustrating the classic high–

risk, high–return trade–off (Baker et al., 2020; Zhang–Kuo, 2021). However, the significant fluctuations in volatility signal that such funds are more suitable for speculative investors willing to accept higher risk for the potential of higher returns. This finding underscores the importance of financial literacy: investors capable of weighing costs, risk, and returns can more effectively select the funds best suited to their objectives (Hastings et al., 2013; Lusardi–Mitchell, 2014).

ESG considerations also played a significant role in the analysis. The assessment of funds with sustainability features confirmed hypotheses proposed in the literature (Friede et al., 2015; Gupta–Chaudhary, 2023; ESMA, 2022), showing that ESG integration does not necessarily reduce fund performance and can, in certain cases, provide competitive returns at lower risk. Positive investor attitudes toward ESG funds (particularly among highly financially literate individuals) indicate that sustainability and cost sensitivity are not mutually exclusive but can be integrated as part of the decision-making process (Deka et al., 2023; Choowan et al., 2025).

An interesting observation is that, in the domestic market, fund size (NAV/AUM) and age do not always exhibit a significant relationship with returns, which can be partially attributed to local market characteristics. Research by Filip (2018) and Sá et al., (2024) shows a similar pattern: domestic equity funds' performance is less dependent on fund size and age, while these factors are more determinant for foreign equity investments. This suggests that investors should consider not only fund characteristics but also the market context when making investment decisions.

Methodologically, the cluster analysis provided valuable insight into the heterogeneity of the Hungarian fund market. The three well-defined groups ("Expensive and Unprofitable," "Cheap and Stable," and "Risky but Profitable") clearly illustrate that the investment fund market can be segmented along multiple dimensions. This segmentation enables investors to select funds that align with their risk profile, cost sensitivity, and return expectations, demonstrating the practical value of financial literacy (Müller–Weber, 2010; Grinblatt et al., 2016).

Overall, the study confirms the coherent body of evidence in the literature: investment fund performance and cost structure are determined by complex factors, with risk, return, costs, and sustainability considerations interacting closely. In the context of the Hungarian market, the analysis highlights that financial literacy and cost sensitivity significantly influence investment decisions, and that ESG integration is becoming an increasingly important dimension in investment decision-making.

5 CONCLUSIONS

The results of our analysis examining the performance of investment funds clearly emphasize that a conscious, information-based approach is essential for financial decision-making. Our panel regression analyses of Hungarian investment funds for the period between 2017 and 2023 support the practical necessity of financial literacy in multiple respects. First, the strong relationship between returns and risks reinforces the propositions of classical financial theories (e.g., CAPM), which suggest that higher-volatility funds generally offer higher returns. At the same time, our models also clearly showed that this relationship is not automatic and does not remain significant in all cases, particularly when asset class characteristics are taken into account. This phenomenon offers a warning that simple decision-making based solely on past performance and risk may not lead to optimal outcomes. Instead, deeper understanding and conscious risk management are required. The examination of the total expense ratio offers particularly important insights from the perspective of financial literacy. While TER does not directly affect returns in every model, in more detailed specifications (including asset classes), its negative impact becomes clear: higher costs typically result in lower net returns. This finding highlights the importance of investor cost sensitivity and suggests that careful evaluation of fee structures is at least as critical as return expectations in conscious fund selection. Our models also indicate that prior fund performance, particularly the previous year's return, often exhibits a negative relationship with future returns. This mean reversion effect suggests that chasing short-term successes may not pay off. In fact, selecting overvalued funds can entail return risk. Therefore, past returns can only serve as supplementary information, not as the sole basis for decision-making.

Asset class differences are particularly noteworthy. Commodity and equity funds generally offer higher returns but are accompanied by higher volatility and often higher TER. By contrast, bond and money market funds may provide a safer alternative, albeit with lower return potential. This spectrum allows investors to select funds aligned with their financial goals and risk tolerance but only if they possess the necessary financial knowledge and awareness. Based on these findings, it can be concluded that financial literacy (the integrated consideration of returns, risks, costs, asset classes, and past performance) is a fundamental prerequisite for making well-informed and successful investment decisions.

In addition to the panel regression results, the cluster analysis provided important complementary insights. Based on standardized values of returns, volatility, TER, and fund size, three clusters were identified. The first cluster typically included smaller, low-volatility funds that provide moderate returns at low costs. This group represents a relevant option for conservative investor profiles. The sec-

ond cluster mainly consisted of medium-sized funds with higher costs but stable returns, where investor confidence and longer-term performance may play a role. The third cluster included riskier funds with high volatility, exceptional return potential, and higher costs, appealing primarily to speculative, risk-tolerant investors. This clustering result underscores the strong heterogeneity of the Hungarian investment fund market and its clear targeting of distinct investor segments. A practical implication is that enhancing financial literacy helps investors more easily identify which cluster of funds aligns best with their risk profiles and return expectations.

The investment fund market exhibits substantial diversity, which, while offering a wide choice, also increases the risk of poor decisions. Therefore, a conscious, data-driven approach is not only desirable but indispensable for achieving sustainable long-term financial success.

Abbreviations

TER	Total Expense Ratio
ESG	Environmental, Social, Governance
PCA	Principal Component Analysis
CAPM	Capital Asset Pricing Model
ESMA	European Securities and Markets Authority

REFERENCES

Adams, J. – Hayunga, D. – Mansi, S. (2022): Index fund trading costs are inversely related to fund and family size. *Journal of banking & finance*, 140, 106527. https://doi.org/10.1016/j.jbankfin.2022.106527.

Ahmed, M. – Seraj, R. – Islam, S. M. S. (2020): The k-means algorithm: A comprehensive survey and performance evaluation. *Electronics*, 9(8), 1295. https://doi.org/10.3390/electronics9081295.

Atkinson, A. et al., (2015): "Financial Education for Long-term Savings and Investments: Review of Research and Literature", OECD Working Papers on Finance, Insurance and Private Pensions, (39), *OECD Publishing*, Paris. https://doi.org/10.1787/5jrtgzfl6g9w-en.

Babbar, S. – Sehgal, S. (2018): Mutual fund characteristics and investment performance in India. *Management and Labour Studies*, 43(1–2), 1–30. https://doi.org/10.1177/0258042X1774518.

Baker, M. – Egan, M. L. –Sarkar, S. K. (2022): How do investors value ESG? (No. W30708). *National Bureau of Economic Research*. https://doi.org/10.3386/w30708.

Boloș, M. I. – Rusu, Ș. – Leordeanu, M. – Sabău-Popa, C. D. – Perțicaș, D. C. – Crișan, M. I. (2025): K-Means Clustering for Portfolio Optimization: Symmetry in Risk–Return Tradeoff, Liquidity, Profitability, and Solvency. *Symmetry*, 17(6), 847. https://doi.org/10.3390/sym17060847.

Célérier, C. M. – Vallée, B. (2015): The motives for financial complexity: An empirical investigation. *SSRN*, (2289890). https://doi.org/10.5167/uzh-107107.

Chen, J. – Hong, H. – Huang, M. – Kubik, J. D. (2004): Does fund size erode mutual fund performance? The role of liquidity and organization. *American Economic Review*, 94(5), 1276–1302. https://doi.org/10.1257/0002828043052277.

Choi, J. J. – Laibson, D. – Madrian, B. C. (2010): Why does the law of one price fail? An experiment on index mutual funds. *The Review of Financial Studies*, 23(4), 1405–1432. https://doi.org/10.1093/rfs/hhp097.

Choowan, P. – Daovisan, H. – Suwanwong, C. (2024): Effects of financial literacy and financial behavior on financial well-being: Meta-analytical review of experimental studies. *International Journal of Financial Studies*, 13(1), 1. https://doi.org/10.3390/ijfs13010001.

Deka, J. – Sharma, M. – Agarwal, N. – Tiwari, K. (2023): Linking ESG-Investing Consciousness Behavioral Biases and Risk-Perception: Scale Validation with Specifics of Indian Retail Investors. *European Journal of Business Science and Technology*, 9(1), 70–91. https://doi.org/10.11118/ejobsat.2023.004.

European Securities and Markets Authority {ESMA} (2022): The drivers of the costs and performance of ESG funds (*ESMA document* number: 50–165–2146).

European Securities and Markets Authority {ESMA} (2019): Net performance of active and passive equity UCITS (TRV Report on Trends, Risks and Vulnerabilities No. 2, 2019). *ESMA*.

Fama, E. F. – French, K. R. (1993): Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1), 3–56. https://doi.org/10.1016/0304-405X(93)90023-5.

Fikri, S. M. – Yahya, M. H. (2019): The fund characteristics, fees, and expenses structure between conventional and Islamic mutual fund. *Asian Academy of Management Journal of Accounting & Finance*, 15(1), 157–190. https://doi.org/10.21315/aamjaf2019.15.1.7.

Filip, D. (2018): The impact of fund attributes on performance: Empirical evidence for Polish equity funds. *Zbornik radova Ekonomskog fakulteta u Rijeci*: časopis za ekonomsku teoriju i praksu, 36(2), 465–488. https://doi.org/10.18045/zbefri.2018.2.465.

Friede, G. – Busch, T. – Bassen, A. (2015): ESG and financial performance: aggregated evidence from more than 2000 empirical studies. *Journal of sustainable finance & investment*, 5(4), 210–233. https://doi.org/10.1080/20430795.2015.1118917.

Gonçalves, T. – Pimentel, D. – Gaio, C. (2021): Risk and performance of European green and conventional funds. *Sustainability*, 13(8), 4226. https://doi.org/10.3390/su13084226.

Grinblatt, M. – Ikäheimo, S. – Keloharju, M. – Knüpfer, S. (2016): IQ and mutual fund choice. *Management Science*, 62(4), 924–944. https://doi.org/10.1287/mnsc.2015.2166.

Gupta, H. – Chaudhary, R. (2023): An analysis of volatility and risk-adjusted returns of ESG indices in developed and emerging economies. *Risks*, 11(10), 182. https://doi.org/10.3390/risks11100182.

Hastings, J. S. – Madrian, B. C. – Skimmyhorn, W. L. (2013): Financial literacy, financial education, and economic outcomes. *Annual Review of Economics*, 5(1), 347–373. https://doi.org/10.1146/annurev-economics-082312-125807.

Hotelling, H. (1933): Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6), 417. https://doi.org/10.1037/h0071325.

Jiang, J. – Liao, L. – Wang, Z. – Xiang, H. (2020): Financial literacy and retail investors' financial welfare: Evidence from mutual fund investment outcomes in China. *Pacific-Basin Finance Journal*, 59, 101242. https://doi.org/10.1016/j.pacfin.2019.101242.

Junaeni, I. (2022): The effect of fund cash flow, fund size, expense ratio and turnover ratio on mutual fund performance. International Journal of Science, *Technology & Management*, 3(3), 660–672. https://doi.org/10.46729/ijstm.v3i3.529.

Kiymaz, H. (2019): Factors influencing SRI fund performance. *Journal of Capital Markets Studies*, 3(1), 68–81. https://doi.org/10.1108/JCMS-04-2019-0016.

Levine, J. (2021): Investment AUM Fee Costs: Evaluating a Simple Formula. arXiv preprint arXiv:2107.00837. https://doi.org/10.48550/arXiv.2107.00837.

Li, J. – Li, Q. – Wei, X. (2020): Financial literacy, household portfolio choice and investment return. *Pacific-Basin Finance Journal*, 62, 101370. https://doi.org/10.1016/j.pacfin.2020.101370.

Livingston, M. – Yao, P. – Zhou, L. (2019): The volatility of mutual fund performance. *Journal of Economics and Business*, 104, 105835. https://doi.org/10.1016/j.jeconbus.2019.02.001.

Lusardi, A. – Mitchell, O. S. (2014): The economic importance of financial literacy: Theory and evidence. *American Economic Journal: Journal of Economic Literature*, 52(1), 5–44. https://doi.org/10.1257/jel.52.1.5.

Lusardi, A. – Michaud, P. C. – Mitchell, O. S. (2017): Optimal financial knowledge and wealth inequality. *Journal of political Economy*, 125(2), 431–477. https://doi.org/10.1086/690950.

Mansor, F. – Bhatti, M. I. – Ariff, M. (2015): New evidence on the impact of fees on mutual fund performance of two types of funds. *Journal of International Financial Markets, Institutions and Money*, 35, 102–115. https://doi.org/10.1016/j.intfin.2014.12.009.

Müller, S. – Weber, M. (2010): Financial literacy and mutual fund investments: who buys actively managed funds?. *Schmalenbach Business Review*, 62(2), 126–153. https://doi.org/10.1007/BF03396802.

Nilsson, J. – Carlsson Hauff, J. – Carlander, A. (2024): How consumers evaluate costs in complex services: the case of mutual fund fees. *European Journal of Marketing*, 58(13), 56–80. https://doi.org/10.1108/EJM-06-2021-0466.

OECD (2020): OECD/INFE 2020 International Survey of Adult Financial Literacy. Paris: *OECD Publishing*.

Oprea, A. (2022): The use of principal component analysis (PCA) in building yield curve scenarios and identifying relative-value trading opportunities on the Romanian government bond market. *Journal of Risk and Financial Management*, 15(6), 247. https://doi.org/10.3390/jrfm15060247.

Pearson, K. (1901): LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin philosophical magazine and journal of science*, 2(11), 559–572. https://doi.org/10.1080/14786440109462720.

Rousseeuw, P. J. (1987): Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. and *Journal of computational applied mathematics*, 20, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7.

Sá, M. I. – Leite, P. – Correia, M. C. (2024): Can mutual fund characteristics predict future performance? Evidence from Portugal. *Studies in Economics and Finance*, 41(5), 1106–1118. https://doi.org/10.1108/SEF-07-2023-0441.

Scholl, B. – Fontes, A. (2022): Mutual fund knowledge assessment for policy and decision problems. *Financial Services Review*, 30(1), 31–56. https://doi.org/10.61190/fsr.v30i1.3470.

Shaekh, A. A. – Rabby, S. M. – Hossain, S. A. (2024): Impact of Fund Size on Performance and Risk of Closed-End Mutual Funds: An Empirical Analysis in the Context of Bangladesh. *SSRN* 4878186. http://dx.doi.org/10.2139/ssrn.4878186.

Sheng, J. – Simutin, M. – Zhang, T. (2023): Cheaper is not better: on the 'superior' performance of high-fee mutual funds. *The review of asset pricing studies*, 13(2), 375–404. https://doi.org/10.1093/rapstu/raac019.

Singh, A. B. – Tandon, P. (2022): Association between fund attributes and fund's performance: a panel data approach. *Bench-marking: An International Journal*, 29(1), 285–304. https://doi.org/10.1108/BIJ-10-2020-0545.

Sukono, Rosadi, D. – Maruddani, D. A. I. – Ibrahim, R. A. – Johansyah, M. D. (2024): Mechanisms of stock selection and its capital weighing in the portfolio design based on the MACD-K-means-Mean-VaR model. *Mathematics*, 12(2), 174. https://doi.org/10.3390/math12020174.

Tenkam, H. M. – MBA, J. C. – Mwambi, S. M. (2022): Optimization and diversification of cryptocurrency portfolios: a composite copula-based approach. *Applied Sciences*, 12(13), 6408. https://doi.org/10.3390/app12136408.

Wermers, R. (2000): Mutual fund performance: An empirical decomposition into stock-picking talent, style, transactions costs, and expenses. Review of Finance, 55(4), 1655–1695. https://doi.org/10.1093/rof/rfw074.

Yang, L. – Rea, W. – Rea, A. (2017): Financial insights from the last few components of a stock market PCA. *International Journal of Financial Studies*, 5(3), 15. https://doi.org/10.3390/ijfs5030015.

Yuan, C. – Yang, H. (2019): Research on K-value selection method of K-means clustering algorithm. J, 2(2), 226–235. https://doi.org/10.3390/j2020016.

Zoričić, D. – Dolinar, D. – Golubić, Z. L. (2020): Factor-based optimization of a fundamentally-weighted portfolio in the illiquid and undeveloped stock market. *Journal of risk and financial management*, 13(12), 302. https://doi.org/10.3390/jrfm13120302.