

Executive Summary

In this master's thesis, I examined 50 key performance indicators to establish, whether those are valuable independent factors, and to which group they belong to. Therefore, I formulated the following research questions: *Are the key performance indicators factors? To which factor group do they belong, or are they independent factors?*

I expected to find at least five to six larger subgroups within the 50 factors investigated. These subgroups were preliminarily titled: value, quality, profitability, growth, and momentum. Furthermore, I defined some individual factors. The aim of this master's thesis was to sort and sort out factor the zoo with the use of various statistical methods. The result of this research should enable investors to make better decisions when selecting factors for factor portfolios.

I examined a selection of 50 from more than 150 potential factors presented in the literature. For the purpose of sorting out the factor zoo, I present various statistics and statistical analyses on the long-short factor portfolios. Based on these analyses, I provide evidence to show which of those factors truly are factors, whether they are valuable, and which factor group they belong to.

The research in this master's thesis can be used to create single factor and multi-factor portfolios. The correlation matrix, hierarchical clustering analysis, and principal component analysis sort out the factor zoo and provide an overview of the relationships between different factors and groups of factors. This will enable an investor to carefully select one or more factors to create a factor strategy. What the research can not provide is a realistic prediction of the amount that one could actually earn with a factor. Unlike in science, I usually find that long-only applications or long-short portfolios with only a small short-leg are being implemented in practice. With the results of five back tests, I provide a glimpse on what can be done with this research and show that one can earn a premium with single factor and multifactor long-only portfolios.

In 1964, Sharpe (1964) introduced the capital asset pricing model. This is the origin of the first equity factor, named beta by Fama and MacBeth (1973). Over the past 58 years, more than 150 factors have been discovered and presented by various researchers. The result is a proverbial factor zoo as presented and investigated by Feng, Giglio, and Xiu (2020).

I applied four statistical methods to sort out the factor zoo. The methods applied include a Student's t-test, a correlation matrix, a hierarchical cluster analysis and a principal component analysis. I calculated monthly long-short and zero-cost portfolio returns and used these for the analyses. With the application of the Student's t-test, I show whether the factor returns are significantly different from zero. For a clear view on the relationships between the factors, I conducted the other three analyses. The hierarchical clustering analysis provided an overview over the entire investigated factor zoo. The principal component analysis was done to, check my

previous assumptions and verify the results of the hierarchical clustering analysis. The correlation matrix was used to confirm the various results.

I found several subgroups of equity risk premium factors. Furthermore, I found that some factors do not belong where I expected them to belong and that some factors are not of value to investors. Table 6 provides an overview of the subgroups of the researched factors created on the basis of the hierarchical clustering analysis. Most of these factors were found in groups that correspond to the economic rationale. Various principal component analyses for these can be found in Figures 20 to 29 in the annex. Based on the analyses and the t-tests, some key performance indicators can be defined as factors and others not. In total, the hierarchical clustering analysis showed eleven subgroups, which I can confirm with a principal component analysis. An additional analysis on the quality minus junk factors showed the relationships within this larger factor group presented by Asness, Frazzini, and Pedersen (2019). For the period from April 1991 to the end of December 2021, it was found that the three free cash-flow based factors in particular were the most valuable. Moreover, growth-based factors are valuable, albeit in different ways. Furthermore, a small cap premium seems to be paid in the S&P 500. For investors, knowledge of the relationships between the different factors presented in the HCA and PCA can be of great benefit in portfolio construction. Thus, not only was the factor zoo sorted out, but it was possible to create value for practice.