DOI: 10.1111/eufm.12333

ORIGINAL ARTICLE

EUROPEAN FINANCIAL MANAGEMENT WILEY

Automated investment management: Comparing the design and performance of international robo-managers

Nils Helms | **Reinhold Hölscher** | **Matthias Nelde**

Department of Business Studies and Economics, Technical University Kaiserslautern, Kaiserslautern, Germany

Correspondence
Nils Helms, Erwin-Schrödinger-Straße 52, 67663 Kaiserslautern, Germany.
Email: helms@wiwi.uni-kl.de

Abstract

Robo-managers offer automated asset management; however, their overall performance is highly debated. We analyze 15 robo-managers from Germany, the United States and the United Kingdom by conducting a comprehensive qualitative and quantitative study. The qualitative comparison shows considerable differences between the various robo-managers, not only across but also within countries. The quantitative evaluation utilizes different measures to evaluate the performance of the robo-manager sample. Our results indicate that each country has one particularly favourable robo-manager. Furthermore, we find that the costs and characteristics of rebalancing measures have only a small effect on performance.

KEYWORDS

automated investment management, digitalization, performance, robo-advisor, robo-manager, qualitative and quantitative study

JEL CLASSIFICATION

G2, G11, G29

We thank John A. Doukas, the editor, and an anonymous referee of European Financial Management as the paper has enormously benefited from their comments.

1 | INTRODUCTION AND LITERATURE REVIEW

The question of how to successfully invest in financial markets is a recurring and highly debated topic for consumers, firms and academics alike. In light of the digitalization taking place in the financial industry, new technologies to support and improve investment decisions have emerged (Aziz et al., 2021). Robo-advisors are at the forefront of those trends, promising successful investment decisions by machines rather than human advisors. These companies rely on algorithmic trading strategies that can be made automatically, without hardly any human decisions in the process. Although each robo-advisor has its own platform and strategies, a few commonalities concerning their procedures are observable. First and foremost, robo-advisors ask potential investors several standardized questions regarding their investment history, experience with specific instruments and asset classes, investment horizon and risk preferences. The use of algorithms allows the robo-advisor to generate an optimized portfolio structure based on the answers to the questions. The portfolio offered includes different proportions of equities, bonds, real estate and commodities depending on the investor's risk preferences. Concerning the investment horizon, robo-advisors usually adjust the portfolio weights on an ongoing basis, because of the natural drift of the asset structure from the originally assigned weights over time. The underlying idea is that this standardized and automated advice can be offered at a reasonable price while delivering good financial performance.

However, the usefulness of robo-advisors is discussed controversially. On the one hand, the suppliers and some industry experts welcome the idea of a simple, transparent and inexpensive way to gain financial profits, especially in times of low interest rates. On the other hand, investor demand remains below supplier expectations (Hock, 2020). The objective of this paper is to examine robo-advisors from different countries from a qualitative and quantitative perspective based on distinct, well-defined criteria.

The current state of research on robo-advisory can be summarized as follows: even though robo-advisors have been around for over a decade, academic research is still scarce. Apart from a large number of rather popular scientific and journalistic articles, few significant publications address the legal (Fein, 2015; Ji, 2017), regulatory (Baker & Dellaert, 2018; Grischuk, 2017) and financial issues. Henceforth, we will analyze the current state of research, focusing on the financial sector. The majority of the articles in the financial area either focus on the optimal design patterns of robo-advisor services or aim to study changes in customer behaviour, portfolio composition or financial performance through robo-advisor use.

The work of Jung et al. (2018) falls into the first category, investigating the optimal design of a robo-advisor for risk-averse investors. Glaser et al. (2019) analyze how a robo-advisor should be designed to be optimally adapted to customer needs. Further, Rossi and Utkus (2020) examine investors' needs and wants in terms of financial advice, clarifying why individual investors choose to hire financial advisors in the first place and the implications for the design of robo-advisers.

In the second category, for example, is the research of Gulden (2019), who examines the factors that drive the acceptance and use of robo-advisors. Scheurle (2017) considers how an investor's portfolio composition changes through the use of a robo-advisor. Further, D'Acunto et al. (2019) analyze how the characteristics and portfolios of robo-advisor users differ from those of nonusers. Additionally, Rossi and Utkus (2021) study how former self-directed investors change their portfolios after using a hybrid robo-advisor service and which investors benefit the most. Belanche et al. (2019) investigate the characteristics that determine if a person is willing to use a robo-advisor. D'Hondt et al. (2019), in turn, investigate which users would benefit the most from a robo-advisor. Cheng et al. (2019) analyze the role of trust in the acceptance of robo-advisor services.

Only a very small number of studies conducted a comprehensive comparison of certain features of robo-advisor services. Tertilt and Scholz (2019) focus on the determination of an investor's risk profile. They examine the depth and quality of the risk-profiling tools used by robo-advisors. Nelde (2019), in turn, compares the investment behaviours of private investors, the market portfolio and robo-advisors in Germany. Nelde's (2019) research aims to assess the scale of rationality in the overall decision-making process based on a comprehensive evaluation of the German robo-advisor market. Puhle (2019) carries out a comparison of German robo-advisors with regard to realized performance. However, the study does not rely on observed asset structures; instead, the asset allocation is estimated with regression analysis. Additionally, the performance figures are not independently collected or calculated but, rather, copied from the external finance platform Brokervergleich.de, whose reliability remains questionable. In a recent study, D'Acunto and Rossi (2021) propose a taxonomy of robo-advisors based on a literature review, market observations and clearly defined dimensions, enabling the segmentation of business models according to their degree of personalization, involvement, discretion and human interaction.

The literature clearly shows that the robo-advisor industry lacks global qualitative and quantitative performance analysis. We are not aware of any scientific article that carries out a comprehensive survey of international robo-advisors. Furthermore, we are also elaborating on the decision-making principles applied by robo-advisors in regard to portfolio construction and composition, as well as the differences observed between the individual providers. These observations are linked to whether and how the investment performance of robo-advisors can be compared. If comparability is ensured, it should be possible to study and identify whose performance in a real market setting is superior to that of competing robo-advisors.

The remainder of this paper is structured as follows. In Section 2, we first develop a catalogue of the criteria that serves as a basis for our qualitative analysis. We then compare multiple robo-advisors from three different countries. Substantial differences in robo-advisor characteristics are further analyzed. On the basis of the findings of this qualitative analysis, Section 3 presents the methodological foundation for the quantitative analysis, using econometrics (bootstrapping) to establish quantitative comparability. The limitations of the analysis are also discussed. On the basis of these considerations, we carry out a comprehensive performance analysis of the robo-advisors in Section 4. The effects of rebalancing and fees and their relation to the overall results are explicitly addressed. Lastly, Section 5 presents the key results.

2 | QUALITATIVE ANALYSIS OF THE ROBO-MANAGER PLATFORMS

The starting point of this analysis is the assessment of several robo-advisor platforms, focusing on the evaluation of qualitative characteristics. First, we will explain the difference between robo-advisors and robo-managers, as well as the selection criteria of our study sample. Second, we will develop the qualitative characteristics and, third, analyze them in the context of the selected companies.

2.1 | Taxonomic robo-advisory differentiation and sample selection

In this section, we explain the selection process for the robo-advisors in our sample. The main selection criterion of this study is based on the segmentation of robo-advisors according to their business model and investment methodology. We employ the recently developed robo-advisor taxonomy of D'Acunto and Rossi (2021). These authors propose a categorization of robo-advisors based on the degree of personalization, discretion, involvement and human interaction in the asset management process. Robo-advisor business models can range from highly automated investment platforms with no human involvement to fairly manual systems that require a high degree of human interaction. Human involvements should be emphasized in the differentiation between traditional robo-advisors, which require the investor to approve every trading decision and so-called robo-managers, which trade automatically on behalf of the investor. Robo-managers leave no discretion to the investor because the investment advice cannot be overruled. Robo-advisors, on the contrary, allow for manual changes to the algorithmic investment advice.

The following analysis will focus solely on robo-advisors with the highest degree of automation (robo-managers), which also have the lowest degree of human interaction (exclusion of hybrid robo-advisors). Our sample, therefore, includes only robo-managers for passive investors and not robo-managers meant for active trading. Consequently, we will use the term *robo-manager* for our selected subset of the overall robo-advisor market.

Our selection can be explained as follows. The specified subset of robo-managers shows the starkest contrast with traditional human advisors.[1] Demonstrating the current state of and public benefit from robo-managers for long-term passive investment offers great insight for demographic groups unaccustomed to self-determined investment decisions. This group of novice investors with little to no prior capital market exposure is especially vulnerable to unfavourable and maladjusted investment advice. Additionally, robo-managers do not allow for traditional face-to-face interaction with a human adviser, requiring the investor to have great trust in algorithms. Without investment experience, novice investors must rely heavily on the trustworthiness of the robo-managers and the profitability of the investment algorithms and rebalancing methods. Given the lack of pertinent research, this study aims to shed light on the overall quality of such highly automated robo-managers for long-term investing and the degree of comparability of the methods and results in different countries.[2]

The quality of our specified robo-managers sample is verified by an econometric model. Such an evaluation of the performance of different robo-managers requires the unambiguous definition of the assets and portfolio weights selected. Hybrid robo-advisors such as Vanguard PAS offer prospective investors only a portfolio suggestion after an initial conversation with a human advisor. This interaction and the dependency of the resulting portfolio cannot be replicated in our econometric study. Therefore, we exclude hybrid robo-advisors from our sample, even though we are aware that this approach eliminates the two largest US robo-advisors, Merrill Edge Guided Investing and Vanguard PAS, from the analysis. Both of these

[1]What investors need in terms of investment advice and whether this can be met by our robo-manager sample is not covered in this paper. For further details, see Rossi and Utkus (2020).

[2]It is likely that the robo-managers we selected appeal only to certain groups of customers, but this issue will not be explored further. D'Acunto and Rossi (2021) note, for example, that robo-managers such as Wealthfront and Betterment allow little investor discretion, although Betterment recently launched a more flexible investment product.

robo-advisors rely to some degree on human advisors in the asset management process, either in the collection of investor characteristics or in the creation of the asset structure.[3]

The second selection criterion is the number of assets under management (AUM) as a proxy for the size of the robo-manager. The general idea is to evaluate the largest and most important robo-managers of selected countries. As information on the financial key performance indicators of robo-managers is not widely or easily available, it was difficult to obtain the AUM of specific robo-managers. We only found on an aggregated basis, the results of the statistics platform Statista. However, Statista publishes only the AUM of selected robo-managers worldwide. Using these figures as a first approximation of the market situation, we used various search platforms to identify additional robo-managers with comparably large AUM.

We selected three markets to ensure representativeness. The US market was selected for being an established market, where robo-managers have been around for approximately a decade. The US market is often prescribed as the nucleus of robo-managers. Apart from the US market, the United Kingdom is also included in our analysis, because of its advanced developmental stage. The German robo-manager market is the third market. Although the German market is not as established as the US or UK market, it has experienced enormous growth in recent years and its importance has, therefore, increased significantly. We are not aware of a study that compares these three markets both qualitatively and quantitatively.

As a third criterion, certain basic requirements had to be met, such as the availability of the data on the portfolio construction at the securities level and how these securities are weighted. If a company did not meet the minimum requirements regarding available information, it was excluded from the study.

Table 1 depicts the five selected robo-managers by country and their AUM. Although we made the final selection of robo-manager platforms in the first quarter of 2019, the data in Table 1 are from February 2020.

2.2 | Criteria for the qualitative analysis of the platforms

To characterize the 15 robo-managers selected, we first developed, collected and analyzed the qualitative criteria before the quantitative measurement of performance. The aim of the qualitative analysis is to use the predefined criteria to gain a better understanding of the heterogeneity of the robo-managers' approach to portfolio selection and subsequent portfolio management. Depending on the heterogeneity of these results, different adjustments will be necessary for the quantitative analysis for comparability. We employ a structuring approach that differentiates between qualitative criteria which can be explained by differences between countries and criteria that identify differences between the robo-managers, but which are independent of the country.

We analyze the different robo-managers based on central financial selection criteria that determine the investment amount, the costs, how the risk level of the investors is determined,

[3]For a detailed explanation of the differences between fully automated and hybrid robo-advisors, see D'Acunto and Rossi (2021), whose differentiation we use in this paper. However, it should be noted that some of the robo-managers we selected do not completely exclude human intervention. For example, some report that the algorithmic decisions are monitored by teams of experts. This monitoring, however, does not refer to individual investment decisions, but to the process as a whole. Therefore, the robo-managers we selected all largely avoid human intervention and interaction, which differentiates them from hybrid robo-advisors.

TABLE 1 AUM of selected robo-manager platforms, by country

This table reports the assets under management (AUM), measured in billion €, of selected robo-manager platforms grouped by home country. The AUM-data are from February 2020, whereas the selection of the five robo-manager platforms per country was already made in the first quarter of 2019.

Germany		United Kingdom		United States of America	
	Total amount in billion €		Total amount in billion €		Total amount in billion €
Scalable	2.20	Wealthsimple	3.52	Schwab	43.00
Cominvest	0.60	Nutmeg	2.00	Betterment	18.00
Liqid	0.57	Scalable	1.74	Wealthfront	26.00
Quirion	0.35	Moneyfarm	0.87	Axos Invest	0.14
Growney	0.10	Wealthify	0.17	SigFig	1.74
Volume in billion €	3.82	Volume in billion €	8.30	Volume in billion €	88.88

how the risk is measured, how the portfolio risk is continuously rebalanced and how the robo-manager constructs the investment portfolio. These features are set on a company-by-company basis and can, therefore, be determined individually by the respective robo-manager. We must consider, however, the extent to which these criteria and their concrete characteristics actually depend primarily on the individual business model of the company or whether systematic differences can also be explained at the country level, for example, by differing market conditions. This differentiation should be particularly evident in the case of the robo-manager Scalable,[4] as this provider offers its services in both Germany and the United Kingdom.

Unlike the above-mentioned features, the regulatory requirements that the individual providers in the three countries must fulfil are country-specific and represent a central condition in investment advice.[5] However, in addition to country-specific requirements, there can also be regulatory differences in the services offered within each country.

First, we briefly characterize each criterion. Second, we present the specific characteristics of the respective robo-managers in Section 2.3. Third, in Section 2.3, we evaluate whether the characteristics observed are country-specific or whether the design is individual and independent of the country and, for example, attributable to the robo-manager's business model.

Specifically, we assessed robo-managers based on the following criteria:

- Regulatory requirements
- Minimum investment amount required by the robo-manager
- Fees for using the robo-manager

[4]In February 2021, Scalable announced that it was ending its retail offering in the UK but that the robo-manager would continue services through cooperation with Barclays.

[5]In addition to the criteria we consider, there could be further country-specific differences resulting from the customer market and the different expectations of these customer groups; however, these are not the focus of this paper. For the basic scope of robo-advisors, which is beyond that of our highly automated robo-advisors, see D'Acunto and Rossi (2021).

- Number of risk classes in which investors are classified
- Available asset classes
- Risk measure for risk quantification
- Detailed description of the portfolio model
- Type of rebalancing required to maintain the investor's risk level

First, we will give a brief overview of the current discussion on the regulatory requirements of robo-managers. The emergence of digital investment advice has raised a debate on how such providers need to be regulated, irrespective of the concrete country specifics. This debate includes a discussion of whether companies that provide algorithmic investment recommendations can meet the requirements for existing investment advice and what modifications are needed.

Bayón and Vega (2018) explain that a complete regulatory framework for the fintech industry is still pending and thus the scope and depth of robo-manager regulation have not yet been finalized and should, therefore, be discussed. Baker and Dellaert (2018) similarly consider this issue and identify questions that regulators must answer in the context of fully comprehensive regulation. The main issues are competence, honesty, customer information, appropriate investment product selection architecture and reliable information technology infrastructure that the robo-managers should provide. It is important to ensure that the regulation is not overly strict and still allows innovation. Baker and Dellaert (2018) conclude that a variety of interdisciplinary efforts from the fields of law, economics and computer science are still needed to adequately regulate robo-managers. It would also be useful for such regulation to distinguish between the extent to which the robo-managers' decisions are fully automated and the extent to which they are human-based. Even if the robo-managers we study are highly automated, their asset management and trading systems must be developed and the algorithms implemented, resulting, in any case, in human influence that can limit the objectivity of the decisions. The implementation of these basic regulatory requirements in the individual countries will be discussed in more detail in Section 2.3.

The minimum investment amount represents the quantity capital investors must invest in a robo-manager to use its automated asset management. This criterion is set by the individual robo-manager, depending on the business model. If the investment amount is very low, new customer groups can emerge that were previously denied access to professional wealth management because the minimum investment amount was too high. We do not expect a country-specific or regulatory requirement for the minimum investment amount.

The fees for using the robo-manager platform are a criterion that shows, on the one hand, whether the automated asset management is equally priced by all providers and whether it is more cost-efficient compared to human asset management.[6] On the other hand, the costs will also be used in the quantitative analysis, especially for measuring the robo-manager's performance, because a distinction must be made between performance before and after taking fees into account. If, for example, a platform demands higher fees, it only makes sense from an economic perspective that the robo-manager generates a comparatively higher return for the investor. With regard to the fees for using the robo-manager platform, it is quite conceivable

[6]For example, Foerster et al. (2017) determine that, in a sample of Canadian households, the average fee for financial advisory amounts to 2.52%. In a traditional financial advisory, fees for financial advice are specified at around 1%–2%; however, fees around 1% are almost exclusive to large investment amounts. Private investors with relatively small investments regularly pay fees in the area of 2%. See, for example, Stanek (2020).

that different fee levels exist in the individual countries, for example, due to differences in the intensity of competition. From a regulatory point of view, the costs must be transparent in all countries.

With the help of questionnaires, the investors are divided into risk classes and each risk class represents an individual risk propensity that has an influence on the portfolio composition. The number of different risk classes provides an indication of the robo-manager's ability to classify varied types of investors in terms of risk. This, in return, will help the robo-manager in constructing an adequate portfolio with the specific risk–return combination preferred by the investor. Furthermore, we will use this feature to analyze the performance variations between different risk classes and robo-managers. The regulatory framework in all three countries requires that the investor's risk appetite be considered in making investment decisions. In Section 2.3, we examine whether this is carried out differently for the 15 robo-managers on a country-specific basis.

Depending on the risk class, the investor is assigned to a specific predefined portfolio composition. The criterion regarding available asset classes is used to evaluate the different kinds of securities included in the portfolio. Further, this criterion provides insight into the different portfolios of the respective robo-manager. We expect the business model to be the main driver of the differences between the portfolio asset classes. There should be no country-specific restrictions regarding the asset classes for the three countries.

A risk measure is required to quantify the risk within each risk class. This risk measure is, therefore, important in the allocation of a portfolio to a risk class when the investor's initial portfolio is first defined, as well as in continuously reassessing whether the portfolio's risk level still matches the investor's risk preference. From a scientific point of view, various risk measures can be applied. If the robo-managers use different risk measures, the quantitative analysis should determine whether there is a correlation between the risk measure used and the resulting performance. The analysis will show whether the selected robo-managers use different risk measures. We expect that differences can be explained by the robo-manager's business model and the associated valuation framework, as well as the understanding of risk, but not by country-specific characteristics.

The detailed description of the portfolio model applied by the robo-manager is a particularly interesting criterion because it allows drawing conclusions on the risk–return capabilities of certain models. We expect different approaches and it will, therefore, be interesting to see the extent to which the approaches are based on classical capital market theory or whether behavioural economic approaches, which have been increasingly discussed in the last years, will also be considered. However, these options exist in all three countries, which is why we expect no country-specific characteristics in this context.

To remain at the risk level previously defined by the investor, all robo-managers have a rebalancing function. The specific type and threshold of the rebalancing method are detailed in the qualitative analysis. Country-specific differences are less expected, as there are no direct regulatory requirements in this regard. Rather, again depending on the robo-manager, different business models are expected to cause the observed differences. The effects of rebalancing on financial performance are also investigated in the quantitative analysis.

2.3 | Results of the qualitative analysis

The results of the qualitative analysis are presented in Tables 2–4, which are structurally identical and differ by country. The leftmost column lists the individual criteria from Section 2.2. In the following, we report basic findings that are especially noteworthy.

TABLE 2 Qualitative criteria of German robo-managers

This table reports five German robo-managers sorted by the qualitative criteria that are presented in Section 2.2. These criteria include the minimum investment amount that is required by the robo-manager, the fees for using the robo-manager, the number of risk classes in which investors are classified, the available asset classes, the risk measure which is used for risk quantification, a detailed description of the portfolio model, the type of rebalancing to maintain the investor's risk level and the regulatory requirements. The data are from December 2019. [a]All the criteria in this table are based on information from white papers, company websites, interviews, or email correspondence with the robo-managers.

Qualitative criteria[a]	**Germany**				
Robo-manager	Scalable	Cominvest	Liqid	Quirion	Growney
Minimum investment amount required by the robo-manager	€10,000	€3,000	€100,000	€5,000	€1
Fees for using the robo-manager	0.75% p.a. of the investment volume plus 0.19% of the product costs	0.95% p.a. of the investment volume plus 0.07%–1.49% of the product costs	Depending on investment volume, for example, for €5,000,000, min. 0.25% p.a. of the investment volume, 0.16% of the product costs; max. 0.6% p.a. of the investment volume, 0.66% of the product costs. For an investment sum of €100,000, min. 0.5% p.a. of the investment volume, 0.16% of the product costs; max. 0.9% p.a. of the investment volume, 0.5% of the product costs	Up to €10,000, 0% p.a. of the investment volume, 0.21% other product costs. After that, 0.04%–0.48% p.a., depending on the investment volume, plus 0.21% of the product costs. Total costs: min. 0.21% p.a. (€5000) and 0.69% p.a. (€5,000,000) in the basic package	Below €10,000, 0.99% p.a. of the investment volume; below €50,000, 0.69% p.a.; starting from €50,000, 0.39% p.a.; max. 0.27% of the product costs

TABLE 2 (Continued)

Qualitative criteria[a]	Germany				
No. of risk classes in which investors are classified	23	5	10	10	5
Available asset classes	Bonds, stocks, real estate, cash	Bonds, stocks	Bonds, stocks, commodities	Bonds, stocks, cash	Bonds, stocks
Risk measure for risk quantification	Value at risk (VaR), expected shortfall and maximum drawdown (various risk measures used to compensate for disadvantages)	Volatility	Risk preference model according to prospect theory	Stock quota (refusal of risk measures such as the VaR)	Not specified
Detailed description of the portfolio model	Individual customer optimization model with a specific optimization formula (enhancement of a Markowitz model), with an explanation of the model	Dynamic portfolio management using a Matlab model (with no further explanation of the model)	Dynamic capital market model in combination with aspects of behavioural economics (no further explanation of the model)	Fama–French 3-factor model	Investments in a world portfolio are being made (with no further explanation of the model)

TABLE 2 (Continued)

Qualitative criteria[a]	Germany				
Type of rebalancing to maintain the investor's risk level	Continuous monitoring, reoptimization if a deviation from the customer's risk goal is detected (balancing via a 1-year VaR limit)	Daily supervision of investor portfolios regarding composition, allowed volatility and risk minimization; rebalancing at a loss rate of 10%; reassessing the asset classes for suitability and weighting every 16 weeks	Rebalancing, if a defined threshold is exceeded, to adjust the risk profile. Definition of the thresholds for every asset class and subclass published by the provider. The higher the risk class, the greater the rebalancing tendency	Scheduled rebalancing once every year to reset portfolios to their planned allocation. If the stock or bond quota exceeds or falls below the planned quota by 10%, the portfolio is balanced unscheduled	Once every year, the asset classes are reset to their original weightings; regular inspections and adjustments if necessary.
Regulatory requirements	Authorized and regulated by BaFin as a financial portfolio manager	Product of Commerzbank AG, which is admitted as a CRR credit institute under European Central Bank supervision and admitted as a financial portfolio manager	Authorized and regulated by BaFin as a financial portfolio manager	Authorized and regulated by BaFin as a financial portfolio manager	Authorized and regulated by BaFin as a financial portfolio manager

TABLE 3 Qualitative criteria of UK robo-managers

This table reports five UK robo-managers sorted by the qualitative criteria that are presented in Section 2.2. These criteria include the minimum investment amount that is required by the robo-manager, the fees for using the robo-manager, the number of risk classes in which investors are classified, the available asset classes, the risk measure which is used for risk quantification, a detailed description of the portfolio model, the type of rebalancing to maintain the investor's risk level and the regulatory requirements. The data are from December 2019. Abbreviation: FCA, Financial Conduct Authority. [a]All the criteria in this table are based on information from white papers, company websites, interviews and email correspondence with the robo-managers. [b]With differences regarding fees and asset structure compared with Scalable Germany.

Qualitative criteria[a]	**United Kingdom**				
Robo-manager	Nutmeg	Moneyfarm	Wealthsimple	Wealthify	Scalable[b]
Minimum investment amount required by the robo-manager	£500	£500	£0	£0	£10,000
Fees for using the robo-manager	In the base package up to £100,000: 0.45% p.a. of the investment volume; after that, 0.25% plus 0.17% of the product costs and 0.06% of the average. market spread	Up to £20,000, 0.7% p.a. of the investment volume, starting from £20,000 and gradually lowering to 0.43% p.a. plus 0.2% of the product costs and 0.09% of the market spread	Below £100,000, 0.7% p.a. of the investment volume; at £100,000 and above, 0.5% p.a. plus 0.18% of the product costs	Fees scaled according to the investment volume: up to £15,000, 0.7% p.a.; up to £50,000, 0.6% p.a.; up to £100,000, 0.5% p.a.; investment volume exceeding £100,000, 0.4% p.a. plus transaction costs, 0.07% and a 0.22% financing charge	0.75% p.a. of the investment volume and 0.16% of the product costs
No. of risk classes in which investors are classified	5	7	9	5	23

TABLE 3 (Continued)

Qualitative criteria[a]	United Kingdom				
Available asset classes	Bonds, stocks, cash	Bonds, stocks, real estate	Bonds, stocks, cash	Bonds, stocks, commodities, real estate, private equity, cash	Bonds, stocks, real estate, cash
Risk measure for risk quantification	Not specified, volatility used in some examples	Volatility and correlation	Not specified, partially in explanations using volatility	Not specified	VaR, estimated shortfall and maximum drawdown (use of various measures to compensate for disadvantages)
Detailed description of the applied portfolio model	No explicit model description, only reference to a diversified multiasset portfolio	Proprietary method that examines the efficient frontier, analogous to Markowitz's portfolio theory	Use of modern portfolio theory (Markowitz), explained in relation to diversification, but no further information on the process is provided	No explicit description of the method. Noted that the portfolio is constructed through algorithms and financial experts	Individual customer optimization model with a specific optimization formula (enhancement of the Markowitz model), with an explanation of the model
Type of rebalancing to maintain the investor's risk level	Weekly portfolio testing, comparison of the current and target weights for every exchange-traded fund; rebalancing if the total sum of the deviations exceeds 6%	Rebalancing is not specified. A blog entry states that rebalancing occurs infrequently and only through the investment committee	Rebalancing on the next business day if the target allocation deviates by more than 20% from the actual allocation	At least once every 3 months, increasing the frequency if necessary	Continuous supervision, rebalancing in case of a negative deviation from the customer's target risk (managed via the 1-year VaR limit)
Regulatory requirements	Authorized and regulated by the FCA as an investment adviser	Authorized and regulated by the FCA as an investment adviser	Authorized and regulated by the FCA as an investment adviser	Authorized and regulated by the FCA as an investment adviser	Authorized and regulated by the FCA as an investment adviser

TABLE 4 Qualitative criteria of US robo-managers

This table reports five US robo-managers sorted by the qualitative criteria that are presented in Section 2.2. These criteria include the minimum investment amount that is required by the robo-manager, the fees for using the robo-manager, the number of risk classes in which investors are classified, the available asset classes, the risk measure which is used for risk quantification, a detailed description of the portfolio model, the type of rebalancing to maintain the investor's risk level and the regulatory requirements. The data are from December 2019. Abbreviation: SEC, US Securities and Exchange Commission. [a]All the criteria in this table are based on information from white papers, company websites, interviews, or email correspondence with the robo-managers.

Qualitative criteria[a]	**United States**				
Robo-manager	SigFig	Schwab Intelligent Portfolios	Betterment	Wealthfront	Axos Invest (formerly WiseBanyan)
Minimum investment amount required by the robo-manager	\$2,000	\$5,000	\$0	\$500	\$1
Fees for using the robo-manager	Up to \$10,000, no yearly fee; starting from \$10,000, 0.25% p.a. of the investment volume plus 0.07%–0.15% of the product costs.	No advisory fees and no commission	Up to \$2 million, 0.25% p.a. of the investment volume; starting from \$2 million, 0.15% p.a.	0.25% p.a. of the investment volume plus between 0.07% and 0.16% of the product costs	No fees due to investment philosophy
No. of risk classes in which investors are classified	20	12	11	20	87 (18 used in study)
Available asset classes	Bonds, stocks	Bonds, stocks, commodities, cash	Bonds, stocks	Bonds, stocks, real estate	Bonds, stocks, real estate

TABLE 4 (Continued)

Qualitative criteria[a]	**United States**				
Risk measure for risk quantification	Volatility	Use of standard deviation in examples, no further specification	Description of deviation, but no further data on the risk measure	Not specified, exemplified by the use of volatility	Not specified, but volatility is mentioned in connection with the description of the investment
Detailed description of the applied portfolio model	Not specified, evidence of modern portfolio theory and Markowitz and Sharpe models (mean-variance optimization)	Portfolio theory and capital asset pricing model (CAPM) with modifications based on new insights (no specification of the additions)	Black–Litterman global portfolio optimization (no further explanation of the model)	Multifactorial Wealthfront Capital Markets Model based on the CAPM and Black–Litterman model	Modern portfolio theory with modified mean-variance optimization (no further description or explanation of the model)
Type of rebalancing to maintain the investor's risk level	Rebalancing in the case of a deviation of several percentage points from target weights (no explicit quantification), continuous updates of rebalancing models	Rebalancing if the portfolio deviates from the targeted allocation; no further data, solely an example in the company blog, where the general procedure is explained	Daily rebalancing if the portfolio deviates from its planned allocation by more than 3%; differentiation between cash flow, sell/buy and allocation change rebalancing	Threshold-based instead of time-based rebalancing to gain profits from market movements	Rebalancing in the case of a deviation from the target allocation by more than 5%; rebalancing also takes place if inflows and outflows are made by the investor to comply with a target allocation
Regulatory requirements	Authorized and regulated by the SEC as a registered investment adviser	Authorized and regulated by the SEC as a registered investment adviser	Authorized and regulated by the SEC as a registered investment adviser	Authorized and regulated by the SEC as a registered investment adviser	Authorized and regulated by the SEC as a registered investment adviser

We first give an overview of the regulatory requirements in the three countries covered in our study. Tables 2–4 list who is responsible for the regulation of the respective robo-managers and what the company is allowed to offer in each country. The basic regulatory requirements cover customer information about the robo-manager and its services, the information the provider must obtain from the customer, how to ensure the appropriateness of the advice and how to consider special concerns relevant to the provision of automated advice. Therefore, information on some of the criteria presented in Section 2.2 is mandated from a regulatory perspective.

In Germany, the basis for national legislation is based on the European Markets in Financial Instruments Directive II (MiFID II) (European Securities and Markets Authority, 2018). Accordingly, financial products should be designed to meet customers' needs. Information on customers' financial situation and risk-bearing capacity, risk tolerance and investment objectives and needs should be made available. Furthermore, MiFID II serves to expand and harmonize the requirements for the qualification of employees and its requirements have been implemented in national laws. In Germany, the Federal Financial Supervisory Authority (BaFin) is responsible for ensuring compliance with the requirements. The legislator has formulated clear requirements regarding the knowledge, experience and reliability of financial portfolio managers. All the robo-managers we analyze are licensed as financial portfolio managers by BaFin. Unlike investment advisors, financial portfolio managers make independent investment decisions for clients and have the authority to dispose of funds and securities in the clients' accounts (BaFin, 2017). In contrast to simple robo-advisors, the highly automated robo-managers we analyze are not limited to a one-time investment recommendation but are characterized by a certain regularity and continuity in decision making. The implementation of investment decisions is not the responsibility of the customer but is handled by the fiduciary. Further details are set out in the Notice on the granting of authorization to provide financial services pursuant to section 32 (1) of the German Banking Act.

The US robo-managers we analyze are regulated by the US Securities and Exchange Commission (SEC) and operate as registered investment advisors. As such, they fall under the Investment Advisers Act of 1940, with the SEC issuing guidance in 2017 on "electronic investment advice" that clarifies how robo-advisors are covered in the Investment Advisers Act of 1940 (Lazaro, 2019; Monaco et al. 2017; US SEC, 2017).[7] This guidance states that robo-managers must provide clients with information about their business model (e.g., information about their algorithms), describe the advisory services they offer and explain how this information is provided. The guidelines also explicitly address whether the fiduciary duty to act in the customer's best interests[8] can be guaranteed by the questionnaires completed by the customers. Furthermore, the disclosures of the robo-managers must sufficiently highlight inherent risks and the special role of the algorithms. An appropriate compliance programme and policies must also be implemented that consider the unique nature of the business model (Monaco et al., 2017; US SEC, 2017). Even if the Investment Advisers Act of 1940 is criticized in part as being structurally weak (Harvey & Thel, 2019; Ji, 2017), no comparably differentiated guideline could be identified in Germany. However, as early as 2015, the European Securities and Market

[7]In addition, the Financial Industry Regulatory Authority (FINRA, 2016) issued the Report on Digital Investment Advice, which identifies various issues related to digital investment advice, including robo-advisors. However, the report does not create any new legal requirements or change any existing regulatory obligations.

[8]Ji (2017) discusses whether robo-advisors are good fiduciaries.

Authority (ESMA) published a first paper at the European level that addresses the need for regulation tailored to robo-advisors (ESMA, 2016).[9]

The UK robo-managers we analyze are authorized by the Financial Conduct Authority (FCA) for various financial services, including advising and managing investments. Furthermore, until Brexit, the United Kingdom was subject to MiFID II, as Germany is, which resulted in a certain analogy to Germany in terms of regulatory requirements. In addition, the FCA (2017a) issued guidance in 2017 on streamlined advice services, including automated robo-advice services. In this context, the FCA emphasizes that streamlined advice services are possible within the regulatory framework, on the condition that these focus on the customer's needs and are in line with FCA rules. The FCA has also declared that robo-advisors have the potential to solve problems, particularly for customers to whom traditional investment advice is not available, thus making an important contribution to promoting competition in the UK financial advice market. However, the FCA (2017b) also mentions that robo-advisors create new types of risks for investors, but that well-designed models have great potential to reduce compliance risk. Overall, UK regulation is focused on openness to innovation, as discussed by Baker and Dellaert (2018) and not on preventing robo-advisors from establishing themselves in the market.

Generally speaking, the regulatory environment in the different countries has developed differently over time. However, we could not find any regulatory requirements regarding the characteristics of individual robo-advisors in the different countries that could explain their differences. If automated asset managers become even more important in the future and specific regulations are issued for them, this could change. Confirmed by Scalable's statement, we assume that it is not the regulatory environment that is responsible for differences in a few criteria between Scalable Germany and Scalable UK, which will be shown below. Scalable explained on enquiry that the differences between its German and UK offerings are due to the different markets and the selection of exchange-traded funds (ETFs), as well as the different taxation of investors in Germany and the United Kingdom. Differences due to a different regulatory approach were, therefore, not mentioned.

Next, we look at the minimum investment amount that is required to open an account. This amount varies greatly between robo-managers, ranging from €0 to €100,000, which shows that robo-managers presumably want to address different customer groups. In Germany, Liqid stands out with an investment amount of €100,000, contradicting the basic idea that robo-managers provide small investors with access to professional asset management. Across countries, the investment amounts are usually much lower, with five providers (one German, two UK and two US) even requiring no minimum amount or a symbolic €1. Overall, no country-specific differences are observed.

There are also noticeable differences between robo-managers in terms of fees. Two providers from the United States (Schwab Intelligent Portfolios and Axos Invest) charge no fee. Axos Invest even explicitly declares this to be one of its core investment philosophies.[10] All other robo-managers charge fees for asset management. Asset management is usually charged as a percentage consisting of a basic fee for use of the robo-manager platform and a percentage for the product costs, which varies depending on the investment product. Some providers also scale the basic fee according to the amount invested. It is worth noting that there are considerable

[9]Further information is available in the 2016 and 2018 ESMA reports.
[10]The terms *costs* and *fees* are used synonymously in this paper for the sake of simplicity.

differences regarding the fees, with the basic fee ranging from 0% to 0.99% and product costs ranging from 0% to 1.49% according to the robo-managers' self-disclosure. We also note that US robo-managers have significantly lower fee levels than the robo-managers from the other two countries.

This lower fee level can be explained, on the one hand, by economies of scale, as AUM in the United States are significantly higher than in the other two countries and, on the contrary, by differences in the intensity of competition and providers' fee policies. From a regulatory perspective, there are no reasons for the lower fees in the United States. Comparing the UK and Germany, we note that Scalable charges the same basic fees in both countries, except the product costs differ slightly. The robo-managers with the highest fees (Cominvest, Growney) are based in Germany. Overall, qualitatively, the robo-managers show significantly different fee levels, but no conclusion can be drawn as to whether higher costs are justified by better performance. This question can only be analyzed quantitatively.

The number of risk classes varies from five to 23, a large range among robo-managers. As described above, the number of risk classes indicates a robo-manager's ability to classify different types of investors in terms of risk. This allows the robo-manager to match the risk–return combination of the portfolio with the investor's preferences. In this context, it is also interesting to know what type of asset classes are included in the portfolio. For some providers, the portfolio consists only of equities and bonds, whereas other providers explicitly include commodities, real estate, private equity and cash positions.

To determine whether investors with similar risk profiles are assigned comparable portfolios across robo-managers, the respective portfolio compositions for low-, medium- and high-risk classes are depicted in Table 5. This approach is based on the assumption that all investors with low-, medium- and high-risk aversion, respectively, are assigned to the lowest, medium and highest risk levels, based on each robo-manager's range of risk classes. This approach is intuitive, at least as far as the highest and lowest risk levels are concerned. Regarding the medium risk attitude, the differentiation becomes easier the more risk classes are available. The risk class row in Table 5 shows the precise number of portfolios per robo-manager we used in the mapping from risk classes to risk levels. For the Scalable Germany robo-manager, the portfolio composition for risk class 1 is the lowest level, risk class 12 is the medium level and risk class 23 is the highest risk level. The same applies to the other 14 robo-managers.

If, for example, the lowest risk class is selected, the robo-manager Scalable Germany invests 77.91% in bonds and 2.95% in real estate, with the remaining 19.14% held in cash, whereas the UK robo-manager Wealthsimple invests 26.98% in equities, 72.03% in bonds and 0.99% in cash. A large range of different portfolio allocations is also noted for the medium- and high-risk classes. In the high-risk class, one provider (Scalable Germany) still invests 43.13% in bonds, only 41.84% in equities, 13.86% in real estate and 1.17% in cash. Other providers (Cominvest, Growney, Betterment) invest 100% of the investment amount in equities in the high-risk class. Overall, portfolio composition analysis shows that the robo-managers in this study use very different asset structures for comparable risk levels.

In summary, the number of asset classes clearly differs between the various providers and differences can also be seen in the composition of the investment universe. This finding illustrates the individuality and heterogeneity of the robo-managers. For regulatory purposes, all countries require an assessment of investor risk, but the number of questions and the specific securities derived from the assessment are not specified to such an extent that country-specific differences can be identified. Even if, for example, the composition of the portfolios differs slightly between Scalable Germany and

TABLE 5 Portfolio compositions of the low-, medium- and high-risk classes in different countries

This table reports portfolio compositions of the low-, medium- and high-risk classes in different countries with data from March 2019. Panel (a) shows the results for German robo-managers, panel (b) details UK robo-managers and panel (c) outlines US robo-managers. Each panel is arranged in the same manner: We examine the risk classes low, medium and high. For each risk class the five robo-managers from Table 1 are listed. The risk level row shows the precise number of portfolios per robo-manager we used in the mapping from risk classes to risk level. The corresponding portfolio composition is reported as the percentage of bonds, stocks, real estate, commodities and cash in the portfolio. [a]AI, Axos Invest; B, Betterment; C, Cominvest; G, Growney; L, Liqid; M, Moneyfarm; N, Nutmeg; Q, Quirion; SF, SigFig; S-Ger, Scalable Germany; SIP, Schwab Intelligent Portfolios; S-UK, Scalable UK; WS, Wealthsimple.

	(a) Germany														
	Risk class														
	Low					**Medium**					**High**				
Robo-mgr.[a]	S-Ger	C	L	Q	G	S-Ger	C	L	Q	G	S-Ger	C	L	Q	G
Risk level	1	1	1	1	1	12	3	5	6	3	23	5	10	11	5
Bonds	77.91%	90.00%	90.00%	99.60%	80.00%	58.77%	50.00%	50.00%	49.60%	50.00%	43.13%	0.00%	0.00%	0.00%	0.00%
Stocks	0.00%	10.00%	5.00%	0.00%	20.00%	29.52%	50.00%	45.00%	50.00%	50.00%	41.84%	100.00-%	95.00%	99.60%	100.00%
Real estate	2.95%	-	-	-	-	10.67%	-	-	-	-	13.86%	-	-	-	-
Commodities	-	-	5.00%	-	-	-	-	5.00%	-	-	-	-	5.00%	-	-
Cash	19.14%	-	-	0.40%	-	1.04%	-	-	0.40%	-	1.17%	-	-	0.40%	-
	(b) United Kingdom														
	Risk class														
	Low					**Medium**					**High**				
Robo-mgr.[a]	N	M	WS	W	S-UK	N	M	WS	W	S-UK	N	M	WS	W	S-UK
Risk level	1	1	1	1	1	3	3	2	3	12	5	6	3	5	23
Bonds	82.70%	97.00%	72.03%	89.00%	69.86%	42.30%	66.00%	52.10%	48.00%	63.18%	0.00%	24.00%	10.51%	8.00%	46.70%
Stocks	17.00%	0.00%	26.98%	8.00%	4.78%	57.40%	32.00%	46.91%	44.00%	25.87%	99.7%	74.00%	88.50%	78.00%	40.32%
Real estate	-	3.00%	-	1.00%	2.78%	-	2.00%	-	3.00%	10.03%	-	2.00%	-	5.00%	11.62%
Commodities	-	-	-	0.00%	-	-	-	-	2.00%	-	-	-	-	4.00%	-

TABLE 5 (Continued)

	(b) United Kingdom														
	Risk class														
	Low					**Medium**					**High**				
Private Equity	-	-	-	0.00%	-	-	-	-	1.00%	-	-	-	-	3.00%	-
Cash	0.30%	-	0.99%	2.00%	22.58%	0.30%	-	0.99%	2.00%	1.59%	0.30%	-	0.99%	2.00%	1.36%
	(c) United States														
	Risk class														
	Low					**Medium**					**High**				
Robo-mgr.[a]	SF	SIP	B	WF	AI	SF	SIP	B	WF	AI	SF	SIP	B	WF	AI
Risk level	1	1	1	1	1	10	6	6	10	9	20	12	11	20	18
Bonds	74.00%	70.60%	100.00%	74.00%	87.30%	40.00%	32.00%	50.00%	34.00%	49.90%	10.00%	0.00%	0.00%	10.00%	8.40%
Stocks	26.00%	0.00%	0.00%	21.00%	9.20%	60.00%	54.00%	50.00%	57.00%	45.50%	90.00%	94.00%	100.00%	74.00%	90.80%
Real estate	-	-	-	5.00%	3.60%	-	-	-	9.00%	4.70%	-	-	-	16.00%	0.60%
Commodities	-	0.00%	-	-	-	-	2.00%	-	-	-	-	0.00%	-	-	-
Cash	-	29.40%	-	-	-	-	12.00%	-	-	-	-	6.00%	-	-	-

Scalable UK, this is due to the different securities available in Germany and the UK (discussed in Section 2.2) and not due to regulatory requirements.

The next step is to determine the risk measure used by the robo-managers. This parameter is not explicitly outlined by some providers, but it can be derived indirectly from descriptions and sample calculations. Volatility is often used as a measure of risk; however, its use has several disadvantages.[11] Therefore, some providers choose a different measure. For example, Scalable combines different risk measures to derive a reliable risk statement. The robo-manager Quirion rejects all classic risk measures and regards only the equity ratio as a reliable risk measure. Overall, those two providers handle the topic of risk measures explicitly and present detailed reasons for their selection, whereas the other providers make no further comments on risk measurement. In our view, the robo-managers in all three countries (except for Scalable) should be more transparent about how they measure risk. The regulatory requirements in their current form are not specific and binding enough to motivate robo-managers to disclose standardized and reproducible risk information.

The robo-managers also provide varying degrees of detail on the models used to structure and manage portfolios. Many providers refer to portfolio theory and related efficiency considerations. Scalable Germany, for example, describes its optimization model and even lists the optimization formula, enabling the prospective investor with the necessary expertise to understand the portfolio design. Other providers describe the basic theories and models they use for portfolio structuring and portfolio management but do not state the optimization formula explicitly. Furthermore, some providers detail the behavioural economic theories considered, thus representing a view that differs from classical capital market theory. Three providers (Nutmeg, Wealthify, Growney) do not report on the basic model used for portfolio design and only point out that investments are made in a portfolio. For detailed descriptions of the applied portfolio model, the providers predominantly use models based on portfolio theory, but some modify these in different ways and the behavioural economic aspects that have become more prominent in recent years have thus also found their way into some models. Moreover, the robo-managers' explanations vary greatly in regard to the design and optimization of the portfolio model. No country-specific differences can be identified, the data are heterogeneously distributed across all three countries and there are no pertinent regulatory requirements. In our opinion, the selected procedures of the robo-managers are due to their business models.

For investors to remain in their initially defined risk classes, all robo-managers carry out rebalancing; however, the specific form of rebalancing varies from provider to provider. Some robo-managers monitor the portfolio composition on an ongoing basis and rebalance it if a certain threshold value is exceeded. Others rebalance once a year, often stating that manual adjustments are performed at the asset manager's discretion. The applied threshold values are seldom explicitly stated. When published, the thresholds range between 3% and 20%, reflecting the heterogeneity of the providers in regard to the rebalancing criterion. The heterogeneity is equally distributed across all three countries. From a regulatory perspective, there are currently no concrete indications in these countries of the design of such rebalancing. From our point of view, robo-managers should generally be more transparent in how they achieve continuous alignment with the investor's risk appetite.

If we summarize our selected criteria again, we can see clear differences between the robo-managers. The criterion with the highest potential to explain differences across countries is regulation. Although the robo-manager characteristics vary across the sample, in our view,

[11]For an extensive illustration of the disadvantages, see, for example, Keppler (1990).

these differences are not attributable to regulation, but, rather, to the robo-manager's business model and its positioning in the market. The heterogeneity in the individual criteria observed across countries can be confusing for customers; therefore, greater transparency from the providers is desirable. Ultimately, of course, the customer can always reach out to the robo-manager to ask for more information. If the idea of a highly automated robo-manager is also applied to customer support, chatbots could answer such investor questions, as already in practice. Greater transparency and consistency could also be achieved through more specific regulatory requirements. Against this backdrop, we believe there is a need for further specification of the digital investment advice phenomenon from a regulatory perspective, particularly for highly automated robo-managers. Once these specifications have been finalized, the regulation could become more important for the specific design of individual robo-managers and lead to greater differences between them across countries.

3 | METHODOLOGICAL FRAMEWORK OF THE QUANTITATIVE ANALYSIS

The qualitative analysis clearly shows substantial differences between the examined robo-managers in all categories. However, the qualitative analysis does not allow for a ranking of the robo-managers. This section, therefore, uses econometrics to develop a method that enables a comparable performance measurement. The limitations of the analysis are detailed at the end of the section.

3.1 | Generating input data from bootstrapping simulations

To determine if a robo-manager's investment decisions are profitable, a uniform and comprehensive database must be generated. This database enables us to derive the characteristic return distributions of the model portfolios, using the necessary information on portfolio structures, costs and rebalancing methods already gathered in Section 2. Nevertheless, the simple approach of using historical price data is inherently flawed, because it only accounts for one specific market outcome. The advantages of robo-managers in deviating market situations (what-if scenarios; see Kleijnen, 2012) cannot be depicted with such a methodology (Pfiffelmann et al., 2016). Hypothetical market scenarios based on historical price patterns are needed to overcome this shortcoming (Hambuckers & Heuchenne, 2016). The simulation and forecasting methods used for this purpose examine how hypothetical path scenarios can be derived from past return series (Mansini et al., 2015). Numerous repetitions of the simulation procedure create a larger testing sample, which can then be used for significance tests (Tsay, 2010).

The methodological objective of simulating return series is to approximate and thus replicate the characteristics of the underlying historical time series as precisely as possible. Time series analysis is conducted to analyze and interpret historical return series with the intent of discovering certain regularities and patterns (Tsay, 2010). The structures found are then reproduced in an overall statistical model as the realization of a particular formal process with random characteristics (Möller, 2003). Most scholars differentiate two main approaches: parametric and nonparametric simulation methods. Parametric approaches consider past returns as drawing from a probability distribution that can be described by a function. Additional hypothetical return series can, therefore, be generated by determining and parameterizing a distribution function (Jobson & Korkie, 1981;

Memmel, 2003). Nonparametric approaches, on the contrary, rely exclusively on historically observed rates of returns when simulating return paths.

The following econometric analysis uses the nonparametric method of bootstrapping. In an urn model, historical returns are randomly drawn with replacement (Davison & Hinkley, 1997). New hypothetical return series are generated by a well-defined joining operation of the draws (Hall, 1992). Various procedures use different methods to determine lengths and structures of the sampled returns to represent the time series in the best possible way. With the intent to accommodate the concept of weakly stationary return time series, we use block bootstrapping in the form of stationary bootstrapping (Cogneau & Zakamouline, 2013). The basic idea is to consider the dependencies between neighbouring observations by using temporally related blocks of returns from the historical time series (Annaert et al., 2009). In stationary bootstrapping, the block lengths are considered a variable, whereby the length of a block of returns is determined from a geometric distribution (Politis & Romano, 1994). The parametrization of the geometric distribution accounts for the serial structure of the underlying return series characteristics. It is a measure of the average block length of the bootstrapping simulation. There is no unambiguous optimal rule for considering the serial structures in a historical time series (Politis & White, 2004). Many studies use the optimization method according to Politis and White (2004), especially as corrected by Patton et al. (2009). We also utilize this approach to determine title-specific block lengths.

In addition, we must define the method for calculating the rate of returns and aggregating returns over time. In step one, logarithmic returns are determined from a historical time series of the daily price data of a financial instrument. The prices are adjusted for the effects of dividends and splits to obtain standardized data. In step two, the daily returns are aggregated into monthly returns. This aggregation is a common procedure in econometric studies of financial stocks, as it provides a good trade-off between large volumes of data and the reduction of random influences. We subsequently simulate 20,000 return time series for each instrument, with a length of 20 years. In total, $20{,}000 \times 20$ years $= 400{,}000$ years or $20{,}000 \times 20$ years $\times 12$ months $= 4{,}800{,}000$ months, are simulated for each financial security. In our opinion, the selected values represent a reasonable trade-off between a high volume of data and technical feasibility.

The bootstrapping was performed without considering dependencies between different instruments. On the one hand, this procedure can be practically justified, as there is no procedure for stationary bootstrapping considering correlations with other financial securities.[12] On the other hand, the procedure can also be justified theoretically, as the simulated time series represent purely fictitious states in the sense of a what-if analysis. In such a case, it is not necessary to account for correlations, as the aim is not to forecast market outcomes. Instead, the focus is on generating an extensive database of hypothetical returns of the financial securities under consideration. Each return path generated by stationary bootstrapping represents a possible market scenario, but it should not be considered as a forecast. The connection with other financial instruments can be neglected due to the high number of simulation runs.

Finally, the monthly logarithmic returns generated are the basis for the subsequent econometric evaluation of robo-manager performance.

[12]In studies that rely on traditional bootstrapping with single return realizations, correlations can be mapped by using simultaneous return realizations and drawing “en bloc” (James & Yang, 2010). For an example, see Zeisberger et al. (2007). This procedure cannot be implemented in stationary bootstrapping due to the different optimal block lengths of the individual financial securities.

3.2 | Performance-based comparison

This study aims to integrate as many robo-manager characteristics as possible. In addition to the asset structures utilized, fees and rebalancing are also integrated. These features, in conjunction with the bootstrapping of returns, lead to a large number of characteristic return series. Performance measures (PM) condense this information into an overall evaluation of financial benefit. Figure 1 shows the necessary steps to calculate the measures.

This study can be roughly split into three phases: (1) data retrieval and processing, (2) the generation of representative robo-manager return series and (3) a review of the data and calculation of performance measure differences (ΔPM).

The objective of the first phase is to retrieve historical time series information and to prepare the data for further analysis. In the first step, raw historical price data for the entire investment universe are extracted from market data providers. The investment universe is represented by the investment instruments utilized by the robo-managers. Most robo-managers invest in a wide range of ETF products. We use Yahoo Finance (https://finance.yahoo.com/) as the source for ETF prices for the US robo-managers, while we use Ariva (http://www.ariva.de) for the German and UK robo-managers. This selection is based on the availability and

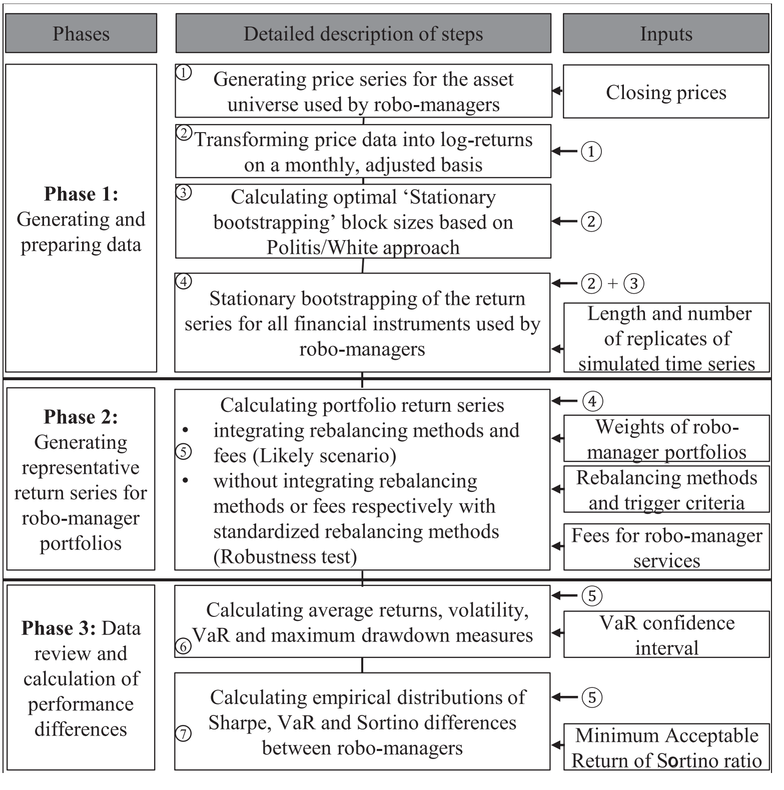

FIGURE 1 Performance analysis steps This figure presents the performance analysis steps (total of seven steps). These steps can be grouped into three phases. Phase 1 consists of generating and preparing the data, phase 2 deals with generating representative return series for robo-manager portfolios and phase 3 covers data review and the calculation of performance differences

consistency of the price history. In detail, we import the daily closing prices in the respective national currencies of the robo-managers, using all the price data since the launch of the financial instrument until 30 June 2019.

In a second step, we transform the price data into logarithmic returns[13] and reduce the frequency from daily to monthly returns. In addition, the prices are adjusted for the effects of dividends and splits, as explained in Section 3.1. The results are the securities' monthly adjusted logarithmic returns.

In a third step, we determine the optimal stationary bootstrapping block length for each financial security, based on the data from the second step. As explained in Section 3.1, we use the procedure of Politis and White (2004).

In the fourth step, we simulate the return series by utilizing stationary bootstrapping. Monthly logarithmic returns are simulated analogously to the input data. In addition, the length of the simulated time series and the number of simulations must be defined. As shown in Section 3.1, these values are set at 20 years and 20,000 simulations, for a total of $20{,}000 \times 20 = 400{,}000$ years simulated per financial security. The results of the stationary bootstrapping are saved for each financial title as an interim result, thus completing the data preparation of the first stage of the process. A total of 20,000 simulated return series are ultimately available for a period of 20 years for each financial security.

In phase two, we generate representative return series of the sample portfolios of the robo-managers. This is accomplished by closely replicating the sample portfolios of the robo-managers. This stage also considers the differences in portfolio structuring identified in Section 2. In a fifth step, we calculate the portfolio returns from the simulated return series of the individual financial securities, taking into account the asset weights. To aggregate the returns in this step, we use simple returns to calculate the portfolio returns as a weighted average of the portfolio components temporarily transformed into discrete returns. We also include the rebalancing methods and trigger criteria of the robo-managers, insofar as these are detailed by the companies. The weights of financial securities that have changed over time are thus reset to their initial values when a certain criterion is triggered.

At this point, some restrictions must be stated. For simplification, an absolute threshold criterion with a trigger value of 10% is used for companies with insufficient information. Use of an absolute threshold criterion is one of the most common procedures in this regard. The trigger value of 10%, in turn, is in the upper range of the empirically observed values, to avoid overoptimization. Some robo-managers only publish the procedure, but not the trigger value. In those cases, the value is derived from plausible statements from the robo-manager, such as calculation examples and hypothetical method descriptions. A further restriction concerns the risk-based rebalancing used by Scalable. This procedure could not be implemented within the scope of the study. Despite the comprehensive white paper by Scalable Capital (2016), it is not possible to reproduce the rebalancing process exactly. Analogous to the procedure for missing specifications, an absolute threshold criterion with a trigger value of 10% is, therefore, assumed for Scalable. For robo-managers with mixed strategies, such as annual rebalancing in connection with a threshold criterion, only the stronger criterion of the threshold rebalancing is implemented. Table 6 presents the rebalancing methods used in the study.

[13]Generally, logarithmic returns are preferred in this study, as they offer advantages in temporal aggregation and more plausible assumptions for the calculation of financial ratios. If the use of discrete returns is more advantageous, such as in the calculation of portfolio returns, the logarithmic returns are temporarily transformed into discrete returns.

TABLE 6 Rebalancing methods

This table reports the rebalancing methods. The robo-advisors are grouped by home countries. For companies with insufficient information, an absolute threshold criterion with a trigger value of 10% is assumed. "Yearly" corresponds to annual rebalancing.

Scalable (Germany)	Threshold: 10%	Nutmeg	Drift: 5%	Schwab Intelligent	Threshold: 5%
Cominvest	Threshold: 10%	Moneyfarm	Threshold: 10%	Betterment	Drift: 3%
Liqid	Threshold: 5%	Wealth-simple	Threshold: 20%	Wealthfront	Threshold: 5%
Quirion	Threshold: 10%	Wealthify	Quarterly	SigFig	Threshold: 5%
Growney	Yearly	Scalable (UK)	Threshold: 10%	Axos Invest	Threshold: 5%

Along with the rebalancing, the fees of the robo-managers are also integrated into the return distribution. As no specific investment amount is assumed for the analysis, the highest management fees are always used for the smallest investment amounts for each robo-manager. We further assume that the costs shown as an annual rate are broken down on a monthly basis as a 12th of the annual rate. The product costs are already included in the initial price data from the financial data providers. Besides the standard scenario in which the observable robo-manager behaviour is reproduced as closely as possible, the return series without rebalancing and fees are also determined for a robustness test.

In the third and final phase, we calculate several descriptive statistics of the generated portfolio returns and subsequently derive a distribution of the performance differences. In step six, the following portfolio statistics are calculated:

- $\text{MD} = -\text{minimum}\left(\text{vector}\left(\frac{\prod_i\left(1+r_{i,\text{monthly}}^{\text{simple}}\right)}{\text{cumulative maxima}\prod_i\left(1+r_{i,\text{monthly}}^{\text{simple}}\right)}-1\right)\right).$
- Simple annual return, $r_{\text{annual}} = \exp\left(\frac{1}{20}\sum_{\text{i}=1}^{240} r_{i,\text{monthly}}^{\log}\right) - 1.$
- Annual volatility, $\sigma_{\text{annual}} = \exp\left(\sigma\left(r_{i,\text{monthly}}^{\log}\right)\sqrt{12}\right) - 1.$
- Annual VaR, $\text{VaR}_{\text{annual}}^{95\%} = \left(5\% - \text{quantile}\left(r_{i,\text{monthly}}^{\text{simple}}\right) + 1\right)^{12} - 1.$
- Maximum drawdown, $\text{MD} = -\text{minimum}\left(\text{vector}\left(\frac{\prod_i\left(1+r_{i,\text{monthly}}^{\text{simple}}\right)}{\text{cumulativemaxima}\prod_i\left(1+\text{r}_{\text{i,monthly}}^{\text{simple}}\right)}-1\right)\right)$ (sharpest percentage decline in the rate of return within a drawdown period)

The portfolio statistics must be interpreted as simple returns. Because of the use of simulation data, the variables are actually distributions of the statistical measure, rather than single values. For subsequent interpretation, these distributions are often condensed into a single value by calculating the mean.

Finally, in step seven, we use the simulated return series to determine the empirical distribution of performance differences. PM are a proven instrument in theory and practice for evaluating the benefits and success of an investment decision. A PM is usually constructed by dividing the return of a financial instrument by a risk measure. On one hand, the PM expresses the absolute advantage of an investment by exceeding or falling below a certain reference point. On the other hand, PM also make it possible to compare several investment alternatives. The performance difference between two alternative investments expresses the degree of advantageousness (Hölscher & Nelde, 2016). Due to their low conceptual restrictions and high practical relevance, a large number of different PM have been developed in the past (Lückoff, 2011). Our intention in this study is not to use as many PM as possible, but, rather, to select measures that allow for a clear and comprehensive interpretation. The decisive characteristic is the type of risk perception (Behr et al., 2008). The Sharpe, VaR and Sortino ratios are, therefore, selected from the vast number of PM available.

The Sharpe ratio (SR) is a very widely used PM (Ferson, 2013). The ratio divides the excess return, that is, the difference between the portfolio return r_p and the risk-free interest rate r_f, by the portfolio risk, measured by the standard deviation σ_P:

$$\mathrm{SR}_P = \frac{r_P - r_f}{\sigma_P}.$$

This key calculation determines the premium per unit of standard deviation (Sharpe, 1994). The larger this ratio, the more successful the investment strategy. For simplification and in consideration of the current interest rate, we assume a risk-free interest rate r_f of 0%. The SR is based on classical portfolio theory according to Markowitz. As the standard deviation is used as a measure of risk, the SR is classified as a symmetrical PM.

In addition, there are performance ratios that use asymmetric risk measures, which means that only negative deviations from a certain target point represent risky situations. In our analysis, the VaR ratio (VR_p) is used as an asymmetrical performance indicator. This ratio determines the excess return $r_p - r_f$ in relation to the VaR of a certain confidence level CL (Dowd, 2000):

$$\mathrm{VR}_P = \frac{r_P - r_f}{\mathrm{VaR}_{\mathrm{CL}}}.$$

Analogous to the SR, the premium is determined per unit of VaR, whereby the greater the ratio, the more successful the investment strategy. As in descriptive statistics Section 4.1, we set CL to 95% and use a risk-free interest rate r_f of 0%. Our use of the VR is substantiated by the high prevalence of VaR in the valuation of financial investments and of nonnormally distributed yield patterns (Chen & Chiang, 2016; Linsmeier & Pearson, 2000).

The Sortino ratio (SO) is used as the second asymmetrical PM. This risk measure is based on the standard deviation, whereby only returns below a certain target yield r_z are taken into account. The risk measure used is typically referred to as a second-order lower partial moment, or, in some cases, downside variance. In addition, the target return r_z is used as a benchmark for the excess return, instead of the risk-free return r_f, resulting in the following formula for SO (Lückoff, 2011):

$$SO_P = \frac{r_P - r_z}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}[\max(r_z - r_t; 0)]^2}}.$$

When comparing several portfolios, a higher value indicates a better performance. We employ the SO in this study because of its frequent use in asymmetric PM. SO somewhat resembles the SR. However, it should be noted that the risk measure used is based on very few extreme values compared to the normal standard deviation σ (Bacon, 2013). Thus, the Sortino measure is highly susceptible to outliers but is smoothed in the following investigation by the high number of repetitions of the bootstrapping procedure. The calculations assume a value of 0% for the target return r_z.

By calculating the performance for all the simulated portfolio returns, we create a distribution of the PM for all the robo-manager portfolios. These distributions are used to derive performance differences and to determine their significance.

ΔPM generally serve to indicate the superiority or inferiority of one investment strategy over another. They are calculated by the simple difference

$$\Delta PM = \mathrm{PM}(\text{investment strategy 1}) - \mathrm{PM}(\text{investment strategy 2}).$$

A positive performance difference indicates that strategy 1 is preferable to strategy 2. With negative values, the reverse is true. As the results of the stationary bootstrapping do not account for correlation, the simulated PM can be combined with each other as desired. This procedure leads to a total of $20{,}000 \times 20{,}000 = 400{,}000{,}000$ different performance differences when comparing two investment strategies. To facilitate the interpretation, we use the mean of the PM distribution. An investment strategy represents a certain sample portfolio of a specific robo-manager. Theoretically, we can compare all the sample portfolios of all the robo-managers against each other. However, for the sake of clarity, only the portfolios of the lowest, medium and highest risk levels per robo-manager are compared, following the reasoning of Section 2.3. This comparison is carried out for each of the three PM specified.

In addition to the performance difference ΔPM, the significance $\propto_s$ of the mean value is determined. The distribution of the performance differences is used. In essence, the aim is to determine whether the mean value deviates significantly from zero (Vinod & Morey, 2000). To accomplish this, we search for the probability of the zero value being outside the confidence interval in the series of simulated performance differences, ordered from small to large. The resulting value $\propto_s$ then determines the significance of the mean deviation (Nelde, 2019).

3.3 | Limitations of the quantitative analysis

Before we start analyzing the results, we want to point out the most important limitations of our approach. The discussion is divided into the limitations of the general econometric methodology and the limitations of the approach selected to interpret the results.

First and foremost, the econometric methodology does not consider potential dynamic adjustments to the portfolio weights by the robo-managers in response to market changes. The target allocations of the portfolio weights are, therefore, considered constant over time. This conflicts to a certain extent with the statement of most robo-managers, which declare that they react to market conditions, especially market turbulence, by adjusting the portfolio weights. This potential intervention to dynamically adjust portfolio weights or overwrite algorithmic portfolio changes that can limit losses could have an impact on the performance and ranking of the robo-managers we survey. However, these adjustments cannot be considered in this study, because the adjustment algorithms are not published. Some robo-managers also reserve the right to intervene manually in the event of unexpected market turbulence. These adjustments cannot be reproduced algorithmically. Even a purely historical analysis of portfolio changes is insufficient, as the dynamic portfolio weights are unknown for the bootstrapped return series. The target weights of the sample portfolios are, therefore, considered constant.

A further limitation concerns the reproduction of the rebalancing. To interpret the results appropriately, one must always take into account that, in the standard scenario, the rebalancing actually used by the robo-manager might not be integrated into the investigation. We were only able to reproduce the rebalancing methods of 10 of the 15 companies examined. Assumptions must be made for the remaining five companies—Scalable (Germany), Cominvest, Scalable (UK), Moneyfarm and SigFig—due to lack of information or implementation problems, thereby limiting the validity of the analysis.

The integration of fees is another limitation. In our implementation, the highest fees are always incurred by the smallest investment amounts per robo-manager. Depending on the specific investment amount, performance differences can, therefore, arise that were not taken into account in this study. However, the effect is not considered to be particularly large.

In addition to the criteria developed in Section 2, other factors can influence performance. Within the scope of the investigation, the individual effects of the respective instruments of the individual robo-managers must be considered. Although the securities of the individual robo-managers are taken into account, the risk–return effects attributable to the instruments are not considered.

In terms of the approach to interpret the results, one can question the general comparability of the lowest-, medium- and highest-risk classes of the model portfolios. Our interpretation is based on the assumption that investors with low, medium and high levels of risk aversion are also classified by each robo-manager into the lowest, medium and highest risk levels, respectively, from the available range of risk classes. However, the results of the research contradict this basic idea to a certain extent. The portfolio characteristics in the risk classes considered differ greatly between robo-managers. These differences are particularly drastic at the risk levels measured in terms of standard deviation for the companies Scalable and Wealthsimple. The risk of the Wealthsimple portfolio in the lowest risk class is greater than that of Scalable's portfolio in the highest risk class (4.05% vs. 3.41%, respectively). Furthermore, the risk levels of almost all robo-managers increase with rising risk classes. Only for Wealthsimple does the risk between the lowest and the medium risk class decrease slightly (from 3.99% to 4.05%). These results are based on the empirical character of the study and the specific approach of the individual robo-managers in the construction of the model portfolios. In this sense, it is questionable whether a specific investor is actually assigned to comparable risk classes for all robo-managers.

The interpretation of the portfolio characteristics based on the mean across all risk levels reduces a multidimensional item to a one-dimensional number, which inherently removes the granularity of the observations. No conclusions can be drawn from the results for individual sample portfolios and investment strategies due to the high level of aggregation as an average value across all risk levels. Rather, trend statements can be made about the characteristics of the individual robo-managers and countries examined. However, the results do not reflect any specific investments the robo-managers actively offer their clients.

TABLE 7 Descriptive statistics of the robo-managers

This table reports the descriptive statistics of the 15 robo-managers from Table 1 grouped by home country and using all the price data since the launch of the financial instrument until 30 June 2019. The descriptive statistics are mean values of all risk-levels of a robo-manager. The return series of the portfolios corresponding to the different risk classes are simulated 20,000 times over 20 years by stationary bootstrapping. The descriptive statistics include the annual return, the annual volatility, the annual Value at Risk with a confidence level of 95% ($VaR_{95\%}$), the maximum drawdown (MD), the Sharpe ratio (SR), the VaR ratio (VR) and the Sortino ratio (SO). For each descriptive statistic, the country means is also reported in an additional row.

		r (%)	σ (%)	$VaR_{95\%}$ (%)	MD (%)	SR	VR	SO
Germany	Scalable (Germany)	5.6	3.0	−10.4	4.0	1.87	0.50	1.19
	Cominvest	5.9	6.1	−24.8	11.2	1.05	0.22	0.56
	Liqid	4.	4.2	−17.2	7.0	1.18	0.26	0.65
	Quirion	4.2	3.1	−12.8	4.4	1.37	0.31	0.76
	Growney	4.6	3.8	−15.7	6.4	1.24	0.28	0.69
	Mean	5.0	4.0	−16.2	6.6	1.34	0.31	0.77
United Kingdom	Nutmeg	6.0	3.3	−12.7	4.5	1.68	0.42	1.05
	Moneyfarm	4.0	5.1	−25.2	10.3	0.74	0.13	0.35
	Wealthsimple	4.7	4.7	−18.3	8.3	1.00	0.23	0.50
	Wealthify	6.3	3.3	−11.5	3.8	1.95	0.53	1.32
	Scalable (UK)	4.8	3.0	−11.6	4.4	1.54	0.39	0.95
	Mean	5.2	3.9	−15.9	6.3	1.38	0.34	0.83
United States	Schwab Intelligent	6.3	2.7	−8.9	2.5	2.35	0.69	1.69
	Betterment	4.9	4.3	−17.5	8.2	1.36	0.48	1.15
	Wealth Front	7.2	6.0	−23.2	10.5	1.26	0.29	0.70
	SigFig	5.9	5.5	−23.0	10.8	1.14	0.24	0.59
	Axos Invest	5.8	5.9	−24.1	12.1	1.16	0.26	0.64
	Mean	6.0	4.9	−19.3	8.8	1.45	0.39	0.95

4 | RESULTS OF THE ECONOMETRIC STUDY

4.1 | Descriptive statistics and interpretation of the key indicators

The starting point of the analysis is the individual evaluation of all the robo-managers. In a first step, only the mean values across all risk levels are considered. The averaging process is exemplified by the company Scalable. Scalable offers 23 risk levels and assigns each risk level to a specific portfolio. The return series of these portfolios are simulated 20,000 times over 20 years. Each simulation results in a specific characteristic of the statistic. Therefore, a mean value is calculated for all 20,000 simulation iterations, resulting in an overall mean value for each variable. In the following analysis, the mean value for all 23 risk levels and, therefore, all 23 sample portfolios, is represented in the mean value. Table 7 presents the descriptive statistics of the robo-managers, by country.

Among the German robo-managers, Table 7 shows that the annual return ranges from 4.2% for Quirion to 5.9% for Cominvest, with a mean of 5.0%. For the UK robo-managers, the spread between returns is slightly wider, ranging from 4.0% for Moneyfarm to 6.3% for Wealthify. The average value of 5.2% is similar to that of the German robo-managers. US robo-managers' returns range from 4.9% for Betterment to 7.2% for Wealthfront. Consequently, the spread is the same as for the UK robo-managers and the average return of 6.0% is well above that of the German and UK robo-managers.

The standard deviation σ for the German robo-managers has a significantly wider range of results compared to the returns, from 3.0% to 6.1%. Companies with higher returns often tend to have higher volatility. However, this correlation does not apply to Scalable (Germany), because this robo-manager has the lowest volatility despite having the second-highest return. The mean value, in turn, is 4.0%. UK robo-managers' range also starts at 3.0% but only extends to 5.1%. This result is also reflected in the lower average value of 3.9%, compared to that of the German robo-managers. The US robo-managers have the broadest range, from 2.7% to 6.0%. The average of 4.9% is significantly higher than those for the German and UK robo-managers. Thus, US robo-managers tend to generate riskier portfolio structures. With the lowest risk level of 2.7%, Schwab Intelligent can be identified as the outlier.

If the VaR risk measure is calculated at the 95% confidence level, there is large overlap in the interpretation of the results for the standard deviation. Despite the methodological differences, US robo-managers still offer the riskiest portfolios. Besides, it is worth mentioning that the German robo-managers are somewhat riskier than the UK robo-managers.

The maximum drawdown measures the largest percentage reduction in returns from a high point to a low point within a drawdown period. This drawdown period ends when the original high point is reached again and a new drawdown period thus starts at that point. The previous interpretations also hold for the maximum drawdown; however, the range of values is extremely large, from 2.5% for Schwab Intelligent to 12.1% for Axos Invest. Overall, the US robo-managers show the highest maximum drawdown, with an average of 8.8%, followed by the German robo-managers, with 6.6% and the UK robo-managers with 6.3%.

Subsequently, the PM defined are analyzed and reconciled with the robo-managers' advantages. We start with the SR, where German robo-managers range between 1.05 for Cominvest and 1.87 for Scalable, with a difference of 0.82 and an average across all values of 1.34. For the UK robo-managers, the difference between the best and worst results is larger, at 1.21, with 1.95 for Wealthify and 0.74 for Moneyfarm. However, the UK average value of 1.38 is slightly higher than the German value. The difference between the best and worst among the US robo-managers is 1.21, with 2.35 for Schwab Intelligent and 1.14 for SigFig, identical to this measure for UK robo-managers. The US mean value is the highest, at 1.45. In an aggregated view, the US robo-managers offer the most profitable investment decisions, followed closely by the UK robo-managers and then the German robo-managers.

In terms of the VR, the spread for the German robo-managers is 0.28, with a high of 0.50 for Scalable and a low of 0.22 for Cominvest, with a mean value of 0.31. The UK robo-manager result is higher, at 0.34, analogous to SR. The same applies to the spread of 0.40, with 0.53 for Wealthify and 0.13 for Moneyfarm. The US robo-managers, on the other hand, rank first in terms of both the spread of 0.45 (0.69 for Schwab Intelligent and 0.24 for SigFig) and the average value of 0.39. Consequently, the ranking of SR is also reflected in the VR.

Lastly, the PM based on SO is analyzed. The German robo-managers have a range of 0.63, resulting from the values of 1.19 for Scalable and 0.56 for Cominvest, with a mean of 0.77. For the UK robo-managers, the range is 0.97 (1.32 for Wealthify and 0.35 for Moneyfarm) and the

TABLE 8 Robo-manager rankings in a comparison of different performance measures
This table reports the robo-managers' rankings in comparisons of the different performance measures. These performance measures include the Sharpe ratio; the VaR ratio; and the Sortino ratio. The rankings were derived from the descriptive statistics presented in Table 7.

		Sharpe ratio rank	VaR ratio rank	Sortino ratio rank
Germany	Scalable (Germany)	3	3	3
	Cominvest	13	14	13
	Liqid	10	10	10
	Quirion	6	7	7
	Growney	9	9	9
United Kingdom	Nutmeg	4	5	5
	Moneyfarm	15	15	15
	Wealthsimple	14	13	14
	Wealthify	2	2	2
	Scalable (UK)	5	6	6
United States	Schwab Intelligent	1	1	1
	Betterment	7	4	4
	Wealth Front	8	8	8
	SigFig	12	12	12
	Axos Invest	11	11	11

mean is 0.83, both higher than the German values. The US values are again the highest, with a range of 1.1 (1.69 for Schwab Intelligent and 0.59 for SigFig) and an average value of 0.95. Thus, results comparable to those for SR are found for SO.

The comparison of the results for different PM also shows that the choice of PM has only a very small effect on the ranking of the robo-managers, despite the considerable conceptual differences. The correlations are illustrated in Table 8.

The comparison of rankings across the three PM clearly shows that they are subject to only very minor changes. The largest deviation is that of Betterment, with a rank shift of three places. The ranking changes for the other robo-managers involve either only one place or none. Furthermore, it is evident that the top three robo-managers are from three different regions. The robo-manager with the highest performance overall is clearly Schwab Intelligent from the US, followed by Wealthify from the United Kingdom and Scalable from Germany. Thus, the dominance of a certain region regarding performance cannot be established.

The presentation and analysis of the descriptive statistics allow for an initial assessment of the robo-managers' investment strategies at a highly aggregated level. The greatest weakness of the evaluation method selected is the averaging across all risk classes for each robo-manager. This approach does not allow for an analysis of individual portfolio structures or investment strategies. Therefore, the following section provides a comprehensive analysis of the lowest-, medium- and highest-risk classes for each robo-manager.

4.2 | Performance analysis by risk classes

As mentioned in Section 2, the robo-managers have different numbers of risk classes. To ensure comparability between the different robo-managers, only the marginal areas, that is, the highest and lowest risk classes and the medium risk class are considered average risk settings.[14]

In a first step, the results for SR are analyzed. Table 9 reports the performance differences between all robo-managers, with the values referring to the difference between investment strategies 1 and 2, or the difference between the column robo-manager and the row robo-manager. The first value (not equal to zero) in the upper left corner is −0.09, which means that, in the lowest risk level, the performance difference between Cominvest and Scalable (Germany) is −0.09. Cominvest is, therefore, less favourable than Scalable, or Scalable is more favourable than Cominvest. Furthermore, the significance of the results is presented, where * indicates $\alpha \leq 10\%$, ** indicates $\alpha \leq 1\%$ and *** indicates $\alpha \leq 0.1\%$. For clarity, the performance comparison parameters selected based on SR are also presented in Table 10; these are the return, the standard deviation and the number of positive performance differences for the Sharpe ratio (ΔSR) in comparison with the other robo-managers. Further, Table 10 reports the maximum and minimum performance differences and resulting spread.

At the lowest risk level, the robo-manager Betterment offers the most advantageous investment strategy. The performance differences with all the other robo-managers are positive and the significance of these results is consistently high. This positive outcome is due to an extremely low-risk strategy that is characterized by an average annual return of 0.81% and a volatility of only 0.37%. In second place is the company Axos Invest, which chooses the exact opposite path. This robo-manager has 13 positive performance differences (i.e., with all robo-managers except for Betterment), using a comparatively high-risk investment strategy, with an average annual return of 4.23% and a volatility of 2.24%. The third most successful company is Wealthify. Wealthify has 12 positive performance differences and also uses a comparatively high-risk investment strategy. The majority of the performance differences can be described as significant.

Comparison of the investment strategies of the lowest risk levels across all robo-managers also reveals very different risk–return exposures: They range from Betterment, with a return of 0.81% and volatility of 0.37%, to Wealthsimple, with a return of 3.13% and volatility of 4.05% and Wealthfront, with a return of 5.77% and volatility of 3.88%. The spread of risk exposures is $4.05\% - 0.37\% = 3.68\%$.

At the medium risk level, the robo-manager with the highest number of positive SR differences changes. The robo-manager Schwab Intelligent now has 14 positive performance differences. The investment strategy chosen thus leads to an average return of 6.56% with a volatility of 2.68%. The results are also highly significant. The company with the second most positive differences is Wealthify. This company has a similar strategy as Schwab, with an average return of 6.46% and a volatility of 3.04%. This is also true for Scalable (Germany), which has the third most positive differences. All differences are significant in the majority of cases. The range of volatilities in the medium-risk class again extends to 5.73% for SigFig and to 2.68% for Schwab Intelligent, for a spread of 3.05%, which is slightly lower than in the lower-risk class.

At the highest risk level, Nutmeg exhibits the highest number of positive performance differences. The investment strategy chosen leads to an average return of 10.33% with a

[14]In principle, it is conceivable that a higher number of risk classes would enable a more precise assessment of the investor's risk attitude. However, this consideration is not the focus of this article.

TABLE 9 Performance differences based on the Sharpe ratio (SR)

This table reports the performance differences based on the Sharpe ratio (SR) between all the robo-managers (categorized according to the lowest, medium and highest risk levels). These performance differences are calculated as the mean of 400,000,000 SR differences equal to SR(investment strategy 1) – SR(investment strategy 2). *, **, ***Statistical significance at the 10%, 5%, 1% levels, respectively.

Lowest risk level

ΔSR = SR(1) – SR(2)	Investment strategy 1														
Investment strategy 2	Scalable (Ger)	Cominvest	Liqid	Quirion	Growney	Nutmeg	Moneyfarm	Wealthsimple	Wealthify	Scalable (UK)	Schwab	Betterment	Wealthfront	SigFig	Axos Invest
Scalable (Ger)	0.00	-0.09	-0.23	-0.34	0.04	-0.79*	-0.91**	-0.5*	0.62*	-1.05***	0.58*	0.86*	0.23	0.32	0.63*
Cominvest		0.00	-0.15	-0.25	0.13	-0.7*	-0.82**	-0.41	0.71*	-0.96**	0.67*	0.94*	0.32	0.41	0.72*
Liqid			0.00	-0.10	0.28	-0.55*	-0.67**	-0.27	0.86**	-0.82**	0.82**	1.09**	0.46*	0.55*	0.87**
Quirion				0.00	0.38	-0.45*	-0.57*	-0.16	0.96**	-0.71*	0.92**	1.19**	0.57*	0.65*	0.97**
Growney					0.00	-0.83**	-0.95***	-0.54*	0.58*	-1.09***	0.54*	0.81*	0.19	0.28	0.59*
Nutmeg						0.00	-0.12	0.29	1.41***	-0.26	1.37***	1.64***	1.02**	1.11**	1.42***
Moneyfarm							0.00	0.41*	1.53***	-0.14	1.49***	1.76***	1.14***	1.23***	1.54***
Wealthsimple								0.00	1.12***	-0.55*	1.08***	1.36**	0.73*	0.82**	1.13***
Wealthify									0.00	-1.67***	-0.04	0.23	-0.39	-0.30	0.01
Scalable (UK)										0.00	1.63***	1.91***	1.28***	1.37***	1.68***
Schwab											0.00	0.27	-0.35	-0.27	0.05
Betterment												0.00	-0.63	-0.54	-0.22
Wealthfront													0.00	0.09	0.40
SigFig														0.00	0.31
Axos Invest															0.00

Medium risk level

Investment strategy 2	Scalable (Ger)	Cominvest	Liqid	Quirion	Growney	Nutmeg	Moneyfarm	Wealthsimple	Wealthify	Scalable (UK)	Schwab	Betterment	Wealthfront	SigFig	Axos Invest
Scalable (Ger)	0.00	-0.91**	-0.9**	-0.65*	-0.8*	-0.21	-1.22***	-0.94**	0.08	-0.19	0.41	-0.75*	-0.73*	-0.98**	-0.93**
Cominvest		0.00	0.01	0.26	0.11	0.7*	-0.31	-0.03	0.98**	0.72*	1.32***	0.15	0.17	-0.08	-0.02
Liqid			0.00	0.25	0.10	0.69*	-0.32	-0.04	0.98**	0.71*	1.31***	0.14	0.17	-0.08	-0.03
Quirion				0.00	-0.15	0.43	-0.58*	-0.29	0.72*	0.46*	1.05**	-0.11	-0.09	-0.34	-0.29
Growney					0.00	0.59*	-0.42	-0.14	0.88**	0.61*	1.21**	0.04	0.06	-0.19	-0.13
Nutmeg						0.00	-1.01**	-0.73*	0.29	0.02	0.62*	-0.54*	-0.52*	-0.77*	-0.72*
Moneyfarm							0.00	0.28	1.3***	1.04**	1.63***	0.47*	0.49*	0.24	0.29
Wealthsimple								0.00	1.02**	0.75*	1.35***	0.18	0.21	-0.04	0.01
Wealthify									0.00	-0.26	0.33	-0.83*	-0.81*	-1.06**	-1.01**
Scalable (UK)										0.00	0.6*	-0.57*	-0.55*	-0.8*	-0.74*
Schwab											0.00	-1.16***	-1.14**	-1.39***	-1.34***
Betterment												0.00	0.02	-0.23	-0.18
Wealthfront													0.00	-0.25	-0.20
SigFig														0.00	0.05
Axos Invest															0.00

Highest risk level

Investment strategy 2	Scalable (Ger)	Cominvest	Liqid	Quirion	Growney	Nutmeg	Moneyfarm	Wealthsimple	Wealthify	Scalable (UK)	Schwab	Betterment	Wealthfront	SigFig	Axos Invest
Scalable (Ger)	0.00	-1.36***	-1.06**	-0.8*	-1.01**	0.24	-1.29***	-1**	-0.53*	-0.14	-0.03	-1.26***	-1.27***	-1.2***	-1.5***
Cominvest		0.00	0.30	0.56*	0.35	1.59***	0.07	0.36	0.83*	1.22***	1.33***	0.10	0.09	0.16	-0.14
Liqid			0.00	0.26	0.05	1.29***	-0.23	0.06	0.52*	0.92**	1.03**	-0.20	-0.22	-0.15	-0.44
Quirion				0.00	-0.21	1.03**	-0.49*	-0.20	0.26	0.66*	0.77*	-0.46	-0.47*	-0.41	-0.7*
Growney					0.00	1.24***	-0.28	0.01	0.47*	0.87**	0.98**	-0.25	-0.26	-0.20	-0.49*
Nutmeg						0.00	-1.53***	-1.23**	-0.77*	-0.37	-0.26	-1.49***	-1.51***	-1.44***	-1.73***
Moneyfarm							0.00	0.30	0.76*	1.16***	1.26***	0.03	0.02	0.09	-0.21
Wealthsimple								0.00	0.46	0.86*	0.97**	-0.26	-0.28	-0.21	-0.5*
Wealthify									0.00	0.40	0.51*	-0.72*	-0.74*	-0.67*	-0.96**
Scalable (UK)										0.00	0.11	-1.12**	-1.14**	-1.07**	-1.36***
Schwab											0.00	-1.23***	-1.24***	-1.18***	-1.47***
Betterment												0.00	-0.01	0.05	-0.24
Wealthfront													0.00	0.07	-0.22
SigFig														0.00	-0.29
Axos Invest															0.00

TABLE 10 Comparison of the Sharpe ratio differences

This table reports the performance comparison parameters based on the Sharpe ratio for the 15 robo-managers listed in Table 1. These parameters include the return; the standard deviation, the number of positive performance differences for the Sharpe ratio (ΔSR),the maximum performance difference, the minimum performance difference and the resulting spread.

Risk level															
Low	Scalable (Ger)	Cominvest	Liqid	Quirion	Growney	Nutmeg	Moneyfarm	Wealthsimple	Wealthify	Scalable (UK)	Schwab	Betterment	Wealthfront	SigFig	Axos Invest
Average return	2.25%	4.31%	2.14%	0.87%	2.50%	1.10%	0.73%	3.13%	5.55%	0.53%	2.13%	0.81%	5.77%	4.67%	4.23%
Standard deviation	1.78%	3.66%	2.18%	0.93%	1.91%	2.27%	1.96%	4.05%	2.95%	2.36%	2.43%	0.37%	3.88%	2.96%	2.24%
# positive ΔSR	7	6	5	4	8	2	1	3	12	0	11	14	9	10	13
Max. difference	1.05	0.96	0.82	0.71	1.09	0.26	0.14	0.55	1.67	-0.14	1.63	1.91	1.28	1.37	1.68
Min. difference	-0.86	-0.94	-1.09	-1.19	-0.81	-1.64	-1.76	-1.36	-0.23	-1.91	-0.27	0.22	-0.63	-0.54	-0.22
Spread	1.91	1.91	1.91	1.91	1.91	1.91	1.91	1.91	1.91	1.76	1.91	1.68	1.91	1.91	1.91

Medium	Scalable (Ger)	Cominvest	Liqid	Quirion	Growney	Nutmeg	Moneyfarm	Wealthsimple	Wealthify	Scalable (UK)	Schwab	Betterment	Wealthfront	SigFig	Axos Invest
Average return	6.15%	6.06%	4.60%	4.36%	4.35%	6.18%	3.67%	4.44%	6.46%	5.70%	6.56%	5.70%	7.35%	6.11%	5.92%
Standard deviation	3.01%	5.29%	3.64%	3.10%	3.47%	3.37%	4.39%	3.99%	3.04%	3.06%	2.68%	4.40%	5.58%	5.73%	5.28%
# positive ΔSR	12	4	5	9	6	10	0	2	13	11	14	6	8	1	3
Max. difference	1.22	0.31	0.32	0.58	0.42	1.01	-0.24	0.28	1.30	1.04	1.63	0.47	0.49	0.24	0.29
Min. difference	-0.41	-1.32	-1.31	-1.05	-1.21	-0.62	-1.63	-1.35	-0.33	-0.60	0.33	-1.16	-1.14	-1.39	-1.34
Spread	1.63	1.63	1.63	1.63	1.63	1.63	1.39	1.63	1.63	1.63	1.30	1.63	1.63	1.63	1.63

High	Scalable (Ger)	Cominvest	Liqid	Quirion	Growney	Nutmeg	Moneyfarm	Wealthsimple	Wealthify	Scalable (UK)	Schwab	Betterment	Wealthfront	SigFig	Axos Invest
Average return	7.12%	7.55%	7.55%	6.96%	7.36%	10.33%	6.84%	6.52%	6.62%	6.85%	8.79%	7.11%	7.71%	6.85%	6.34%
Standard deviation	3.41%	9.97%	7.14%	5.36%	6.73%	4.46%	8.40%	5.96%	4.24%	3.50%	4.28%	8.38%	9.22%	7.60%	10.22%
# positive ΔSR	13	1	6	9	7	14	2	8	10	11	12	4	3	5	0
Max. difference	1.50	0.14	0.44	0.70	0.49	1.73	0.21	0.50	0.96	1.36	1.47	0.24	0.22	0.29	-0.14
Min. difference	-0.24	-1.59	-1.29	-1.03	-1.24	0.24	-1.53	-1.23	-0.77	-0.37	-0.26	-1.49	-1.51	-1.44	-1.73
Spread	1.73	1.73	1.73	1.73	1.73	1.50	1.73	1.73	1.73	1.73	1.73	1.73	1.73	1.73	1.59

standard deviation of 4.46%. The company Scalable (Germany) drops to second place concerning the number of positive performance differences, choosing a significantly lower-risk investment strategy (with a return of 7.12% and a volatility of 3.41%). In third place is Schwab Intelligent, with a return of 8.79% and a volatility of 4.28%. The majority of the performance differences can again be described as significant. The range of investment strategies implemented is extremely wide in terms of risk–return exposure, ranging from the lowest risk for Scalable (Germany), with a volatility of 3.41% and a return of 7.12%, to the highest risk for Axos Invest, with a volatility of 10.22% and a return of 6.34%. The range of volatility is, therefore, 6.81%.

Comparison across all the risk levels shows that only Schwab Intelligent and Wealthify have a double-digit number of positive performance differences. These companies thus manage to make better investment decisions across all risk classes in terms of the SR compared with the other companies. Furthermore, the robo-managers choose very different investment and trading strategies in all the risk classes in terms of risk–return exposure. A certain standardization and uniformity in the investment strategies selected between the largest robo-managers cannot, therefore, be observed yet. Investors should, therefore, be aware that the lowest-risk class, for example, has a completely different risk–return profile, depending on the robo-manager. The same statement applies to the medium- and highest-risk classes.

In the next step, the results are analyzed based on the PM of the VaR ratio depicted in Tables 11 and 12. Basically, the change in results compared to those for SR is only minor. At the lowest risk level, Betterment, Wealthify and Axos Invest still exhibit the highest numbers of positive performance differences. Only four robo-managers change in their number of positive performance differences. Overall, the significance of the results is very high. Analogous to SR, the resulting VaR range is wide, 17%, with 95% confidence, from −0.36% for Betterment to −17.36% for Wealthsimple.

At the medium risk level, Schwab Intelligent has the highest number of positive performance differences, followed by Wealthify and Scalable (Germany). Only four companies change the number of positive performance differences, compared to when SR is used. The risk exposure extends from −8.49% for Schwab Intelligent to −24.63% for SigFig, for a range of 16.14%. This value is in a similar range to that for lowest risk level.

At the highest risk level, Nutmeg followed by Scalable (Germany) and Schwab Intelligent, shows the highest numbers of positive performance differences. There is, therefore, no change compared with SR-based valuation. Overall, the ranking changes again only for four robo-managers. The VaR extends from −11.47% for Scalable (Germany) to −41.45% for Axos Invest and thus covers a range of 29.98%, a significant increase compared to the low- and medium-risk classes. In an overall assessment, only the companies Wealthify and Schwab Intelligent achieve a double-digit number of positive performance differences at all risk levels.

Finally, the results based on SO are analyzed. To sum up our findings, the results are quite similar to those for the SR and VR. A detailed analysis is, therefore, not necessary. The detailed results are presented in Tables A1 and A2.

The results show that only a few robo-managers consistently succeed in making better investment decisions than the other robo-managers across all risk classes. Schwab Intelligent and Wealthify stand out with their consistent double-digit numbers of positive performance differences. A single robo-manager that dominates all the others in terms of investment decisions does not exist in the research sample. Furthermore, we have shown that the concrete PM plays only a minor role in the results of the study, especially in terms of the number of positive performance differences. Table 13 illustrates this finding and shows the maximum

TABLE 11 Performance differences based on the VaR ratio

This table reports the performance differences based on the VaR ratio (VR) between all the robo-managers (categorized according to the lowest, medium and highest risk levels). These performance differences are calculated as the mean of 400,000,000 VR differences equal to VR(investment strategy 1) – VR(investment strategy 2). *, **, ***Statistical significance at the 10%, 5%, 1% levels, respectively.

Lowest risk level

ΔVR = VR(1) – VR(2)	Investment strategy 1														
Investment strategy 2	Scalable (Ger)	Cominvest	Liqid	Quirion	Growney	Nutmeg	Moneyfarm	Wealthsimple	Wealthify	Scalable (UK)	Schwab	Betterment	Wealthfront	SigFig	Axos Invest
Scalable (Ger)	0.00	-0.01	-0.02	-0.08	0.02	-0.19*	-0.21**	-0.12	0.25*	-0.24**	0.21*	2.11**	0.09	0.09	0.24*
Cominvest		0.00	-0.01	-0.07	0.03	-0.19*	-0.2**	-0.11	0.26*	-0.23**	0.22*	2.11***	0.10	0.10	0.24*
Liqid			0.00	-0.06	0.05	-0.17*	-0.19**	-0.09	0.27*	-0.22**	0.23*	2.13***	0.12	0.11	0.26*
Quirion				0.00	0.10	-0.11*	-0.13*	-0.04	0.33**	-0.16*	0.29**	2.18**	0.17*	0.17*	0.32**
Growney					0.00	-0.22**	-0.24***	-0.14*	0.22*	-0.26***	0.19*	2.08**	0.07	0.06	0.21*
Nutmeg						0.00	-0.02	0.08	0.44***	-0.05	0.4***	2.3***	0.29**	0.28**	0.43***
Moneyfarm							0.00	0.09*	0.46***	-0.03	0.42***	2.32***	0.31***	0.3***	0.45***
Wealthsimple								0.00	0.37***	-0.12*	0.33***	2.22***	0.21*	0.21*	0.35**
Wealthify									0.00	-0.49***	-0.04	1.86*	-0.15	-0.16	-0.01
Scalable (UK)										0.00	0.45***	2.35***	0.33***	0.33***	0.48***
Schwab											0.00	1.9**	-0.11	-0.12	0.03
Betterment												0.00	-2.01**	-2.02**	-1.87*
Wealthfront													0.00	-0.01	0.14
SigFig														0.00	0.15
Axos Invest															0.00

Medium risk level

Investment strategy 2	Scalable (Ger)	Cominvest	Liqid	Quirion	Growney	Nutmeg	Moneyfarm	Wealthsimple	Wealthify	Scalable (UK)	Schwab	Betterment	Wealthfront	SigFig	Axos Invest
Scalable (Ger)	0.00	-0.34**	-0.32**	-0.25*	-0.29*	-0.11	-0.42***	-0.32**	0.03	-0.08	0.17	-0.28*	-0.27*	-0.35**	-0.34**
Cominvest		0.00	0.01	0.08	0.05	0.23*	-0.08	0.02	0.36**	0.26*	0.5***	0.06	0.06	-0.02	0.00
Liqid			0.00	0.07	0.03	0.21*	-0.10	0.00	0.35**	0.25*	0.49***	0.04	0.05	-0.03	-0.01
Quirion				0.00	-0.04	0.14	-0.17*	-0.06	0.28*	0.18*	0.42**	-0.03	-0.02	-0.10	-0.08
Growney					0.00	0.18*	-0.13*	-0.03	0.32**	0.21*	0.46***	0.01	0.02	-0.06	-0.04
Nutmeg						0.00	-0.31**	-0.21*	0.14	0.03	0.28*	-0.17*	-0.16*	-0.24*	-0.22*
Moneyfarm							0.00	0.10	0.44***	0.34***	0.59***	0.14*	0.15*	0.07	0.08
Wealthsimple								0.00	0.34**	0.24*	0.48***	0.04	0.04	-0.03	-0.02
Wealthify									0.00	-0.10	0.14	-0.31*	-0.3*	-0.38**	-0.36**
Scalable (UK)										0.00	0.24*	-0.2*	-0.2*	-0.28*	-0.26*
Schwab											0.00	-0.45**	-0.44**	-0.52***	-0.5***
Betterment												0.00	0.01	-0.07	-0.06
Wealthfront													0.00	-0.08	-0.06
SigFig														0.00	0.02
Axos Invest															0.00

Highest risk level

Investment strategy 2	Scalable (Ger)	Cominvest	Liqid	Quirion	Growney	Nutmeg	Moneyfarm	Wealthsimple	Wealthify	Scalable (UK)	Schwab	Betterment	Wealthfront	SigFig	Axos Invest
Scalable (Ger)	0.00	-0.43***	-0.37**	-0.3*	-0.35**	0.05	-0.45***	-0.31*	-0.21*	-0.06	-0.03	-0.41***	-0.4***	-0.39***	-0.46***
Cominvest		0.00	0.06	0.13*	0.08	0.48***	-0.02	0.12	0.22*	0.37***	0.39***	0.02	0.03	0.03	-0.03
Liqid			0.00	0.07	0.02	0.42***	-0.08	0.06	0.16*	0.31**	0.33**	-0.04	-0.03	-0.03	-0.09
Quirion				0.00	-0.05	0.35**	-0.15*	-0.01	0.09	0.24*	0.27*	-0.11	-0.10	-0.09	-0.16*
Growney					0.00	0.4**	-0.10	0.04	0.14	0.29**	0.31**	-0.06	-0.05	-0.05	-0.11*
Nutmeg						0.00	-0.5***	-0.36**	-0.26*	-0.11	-0.08	-0.46***	-0.45***	-0.44***	-0.51***
Moneyfarm							0.00	0.14*	0.24**	0.39***	0.41***	0.04	0.05	0.05	-0.01
Wealthsimple								0.00	0.10	0.25*	0.27*	-0.10	-0.10	-0.09	-0.16*
Wealthify									0.00	0.15	0.17	-0.2*	-0.2*	-0.19*	-0.26**
Scalable (UK)										0.00	0.02	-0.35**	-0.34**	-0.34**	-0.4***
Schwab											0.00	-0.37***	-0.37**	-0.36**	-0.43***
Betterment												0.00	0.01	0.01	-0.05
Wealthfront													0.00	0.01	-0.06
SigFig														0.00	-0.07
Axos Invest															0.00

TABLE 12 Comparison of the VaR ratio differences

This table reports the performance comparison parameters based on the VaR ratio (VR) for the 15 robo-managers listed in Table 1. These parameters include the return; the standard deviation; the number of positive performance differences for the VR (ΔVR), the maximum performance difference, the minimum performance difference and the resulting spread.

Risk level															
Low	Scalable (Ger)	Cominvest	Liqid	Quirion	Growney	Nutmeg	Moneyfarm	Wealthsimple	Wealthify	Scalable (UK)	Schwab	Betterment	Wealthfront	SigFig	Axos Invest
Average return	2.25%	4.31%	2.14%	0.87%	2.50%	1.10%	0.73%	3.13%	5.55%	0.53%	2.13%	0.81%	5.77%	4.67%	4.23%
$VaR_{95\%}$	-7.79%	-14.47%	-8.25%	-4.23%	-7.90%	-11.87%	-9.80%	-17.36%	-9.87%	-12.59%	-12.19%	-0.36%	-14.22%	-11.89%	-7.91%
# positive ΔVR	7	6	5	4	8	2	1	3	13	0	11	14	10	9	12
Max. difference	0.24	0.23	0.22	0.16	0.26	0.05	0.03	0.12	0.49	-0.03	0.45	2.35	0.33	0.33	0.48
Min. difference	-2.11	-2.11	-2.13	-2.18	-2.08	-2.30	-2.32	-2.22	-1.86	-2.35	-1.90	1.86	-2.01	-2.02	-1.87
Spread	2.35	2.35	2.35	2.35	2.35	2.35	2.35	2.35	2.35	2.32	2.35	0.49	2.35	2.35	2.35
Medium	Scalable (Ger)	Cominvest	Liqid	Quirion	Growney	Nutmeg	Moneyfarm	Wealthsimple	Wealthify	Scalable (UK)	Schwab	Betterment	Wealthfront	SigFig	Axos Invest
Average return	6.15%	6.06%	4.60%	4.36%	4.35%	6.18%	3.67%	4.44%	6.46%	5.70%	6.56%	5.70%	7.35%	6.11%	5.92%
$VaR_{95\%}$	-10.19%	-22.81%	-15.18%	-12.90%	-14.50%	-12.56%	-22.04%	-16.40%	-10.22%	-10.86%	-8.49%	-17.92%	-21.83%	-24.63%	-22.33%
# positive ΔVR	12	2	4	9	6	10	0	5	13	11	14	7	8	1	3
Max. difference	0.42	0.08	0.10	0.17	0.13	0.31	-0.07	0.10	0.44	0.34	0.59	0.14	0.15	0.07	0.08
Min. difference	-0.17	-0.50	-0.49	-0.42	-0.46	-0.28	-0.59	-0.48	-0.14	-0.24	0.14	-0.45	-0.44	-0.52	-0.50
Spread	0.59	0.59	0.59	0.59	0.59	0.59	0.52	0.59	0.59	0.59	0.44	0.59	0.59	0.59	0.59
High	Scalable (Ger)	Cominvest	Liqid	Quirion	Growney	Nutmeg	Moneyfarm	Wealthsimple	Wealthify	Scalable (UK)	Schwab	Betterment	Wealthfront	SigFig	Axos Invest
Average return	7.12%	7.55%	7.55%	6.96%	7.36%	10.33%	6.84%	6.52%	6.62%	6.85%	8.79%	7.11%	7.71%	6.85%	6.34%
$VaR_{95\%}$	-11.47%	-38.83%	-28.66%	-21.75%	-26.88%	-14.80%	-40.80%	-21.23%	-16.02%	-12.16%	-14.58%	-33.60%	-34.81%	-30.75%	-41.45%
# positive ΔVR	13	2	6	9	7	14	1	8	10	11	12	3	4	5	0
Max. difference	0.46	0.03	0.09	0.16	0.11	0.51	0.01	0.16	0.26	0.40	0.43	0.05	0.06	0.07	-0.01
Min. difference	-0.05	-0.48	-0.42	-0.35	-0.40	0.05	-0.50	-0.36	-0.26	-0.11	-0.08	-0.46	-0.45	-0.44	-0.51
Spread	0.51	0.51	0.51	0.51	0.51	0.46	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.50

TABLE 13 Maximum deviation between the number of positive performance differences
This table reports the maximum deviations between the performance measures regarding the numbers of positive performance differences when comparing the results from the Sharpe, VaR and Sortino ratios.

	Risk classRobo-manager	Low	Medium	High
Germany	Scalable (Germany)	1	0	0
	Cominvest	1	2	1
	Liqid	0	1	1
	Quirion	0	0	0
	Growney	0	0	1
United Kingdom	Nutmeg	0	0	0
	Moneyfarm	0	0	1
	Wealthsimple	0	3	2
	Wealthify	1	0	0
	Scalable (UK)	0	0	0
United States	Schwab Intelligent	1	0	0
	Betterment	0	1	1
	Wealth Front	1	0	1
	SigFig	1	0	0
	Axos Invest	2	0	0

deviation in the number of positive performance differences when comparing the results from the Sharpe, VaR and Sortino ratios. The largest deviation is for Wealthsimple at the medium risk level, with a value of three. All the other deviations between the risk measures range from zero to two.

The previous analysis aimed to realistically replicate the observable characteristics of the individual robo-managers. As part of the analysis, specific rebalancing methods and company-specific fees are integrated into the study when the information is available. In a further step, we focus on how the results change when the differences in costs and rebalancing are eliminated for the individual robo-managers.

4.3 | Impact of rebalancing on PM

To investigate the effects of rebalancing, we assume two standardized rebalancing methods for all robo-managers. At first, we analyze how the results change with no rebalancing. In this scenario, after capital has been invested, no interim transactions take place until the account is closed, regardless of market fluctuations (Dichtl et al., 2016). This procedure is also known as the buy-and-hold approach (Cuthbertson et al., 2016). As a second approach, we investigate how the results change when all the robo-managers use the most frequently observed procedure of threshold-based rebalancing with a limit value of 10%. Both approaches will enable us to show what effects the rebalancing methods have on the results. Essentially, the intent is to determine the influence of each rebalancing method on the benefits of the robo-managers.

TABLE 14 Deviation of the mean values

This table reports the deviation of the mean values of the descriptive statistics, calculated as the difference between scenarios without rebalancing and the normal scenario. The descriptive statistics include the annual return, the annual volatility, the annual Value at Risk with a confidence level of 95% ($VaR_{95\%}$), the maximum drawdown (MD), the Sharpe ratio (SR), the VaR ratio (VR) and the Sortino ratio (SO). For each statistic, the country means is also reported in an additional row.

		Δr (%)	$\Delta\sigma$ (%)	$\Delta VaR_{95\%}$ (%)	ΔMD (%)	ΔSR	ΔVR	ΔSO
Germany	Scalable (Germany)	0.6	0.3	−0.9	0.4	0.03	0.01	0.01
	Cominvest	1.4	0.8	−3.2	1.0	0.06	0.01	0.01
	Liqid	0.8	0.8	−3.4	1.6	−0.05	−0.02	−0.04
	Quirion	0.4	0.4	−1.7	0.9	−0.05	−0.01	−0.04
	Growney	0.8	0.9	−3.7	2.1	−0.07	−0.02	−0.06
	Mean	0.8	0.6	−2.6	1.2	−0.02	−0.01	−0.03
United Kingdom	Nutmeg	2.0	0.7	−1.5	0.3	0.26	0.08	0.19
	Moneyfarm	0.9	1.1	−4.4	2.3	0.00	0.00	−0.00
	Wealthsimple	0.9	0.1	−0.3	−0.5	0.16	0.04	0.11
	Wealthify	1.2	0.8	−2.5	1.4	−0.04	−0.02	−0.07
	Scalable (UK)	1.5	0.7	−1.5	0.4	0.17	0.05	0.12
	Mean	1.3	0.7	−2.0	0.8	0.11	0.03	0.07
United States	Schwab Intelligent	1.1	0.7	−2.5	1.0	−0.15	−0.07	−0.20
	Betterment	0.6	0.8	−3.1	2.1	−0.14	−0.04	−0.11
	Wealth Front	0.4	0.9	−3.0	2.8	−0.11	−0.03	−0.08
	SigFig	0.4	1.1	−4.7	3.4	−0.15	−0.04	−0.10
	Axos Invest	0.5	0.9	−3.7	2.5	−0.12	−0.04	−0.10
	Mean	0.6	0.9	−3.4	2.4	−0.13	−0.04	−0.12

First, changes resulting from assuming a uniform buy-and-hold approach are examined. Table 14 shows the spread between the scenario with a buy-and-hold approach compared to the normal scenario from the previous Section 4.2. The calculation is based on the average value across all scenarios and risk levels.

It is clear from these results that the returns and the risk parameters have increased significantly. Without rebalancing measures, the average returns increase in a range from 0.4% for Quirion, Wealthfront and SigFig to 2.0% for Nutmeg. Comparing the countries, we find that a lack of rebalancing in the United Kingdom would lead to an increase in returns of 1.3%. In Germany and the United States, the effect is much less distinctive, with respective increases of 0.8% and 0.6%.

Volatility also increases significantly, ranging from 0.1% for Wealthsimple to 1.1% for Moneyfarm and SigFig. The volatility of German robo-managers increases only up to 0.6%. However, the UK robo-managers follow very closely, with a 0.7% increase, as do the US robo-managers, with 0.9%. Similar effects can be observed for the VaR results. Again, the range extends from −0.3% for Wealthsimple to −4.7% for SigFig. This time, the UK robo-managers show the smallest risk increase, with −2.0%, followed by the German robo-managers with −2.6% and the US robo-managers with

TABLE 15 Changes in the number of positive performance differences

This table reports the changes in the number of positive performance differences in the lowest-, medium- and highest-risk classes, calculated by the differences between the scenarios without rebalancing and the standard scenario. The performance measures (PM) that are considered in this table are the Sharpe ratio (SR), the VaR ratio (VR) and the Sortino ratio (SO).

PM	SR			VR			SO		
Risk class (L = low, M = medium, H = high)	L	M	H	L	M	H	L	M	H
Scalable (Germany)	0	0	0	−1	1	0	0	1	0
Cominvest	0	−1	4	2	1	2	1	−1	3
Liqid	0	0	0	0	1	0	0	0	−1
Quirion	−2	0	0	−1	−1	−1	−2	0	0
Growney	0	−2	0	−1	−2	0	1	−2	−1
Nutmeg	1	3	0	0	2	0	1	2	0
Moneyfarm	−1	0	0	−1	0	0	−1	0	0
Wealthsimple	1	6	0	1	4	1	1	6	2
Wealthify	0	−2	0	−1	−2	0	−1	−2	0
Scalable (UK)	1	−1	0	1	−1	1	1	−1	0
Schwab Intelligent	2	0	0	2	0	−1	1	0	0
Betterment	0	0	0	0	−1	2	0	−1	1
Wealth Front	1	−1	−2	0	−1	−2	1	−1	−2
SigFig	−1	0	−2	0	0	−2	−3	0	−2
Axos Invest	−2	−1	0	−1	−1	0	0	−1	0

−3.4%. It is worth noting that the robo-managers with the highest increases in returns are not those with the greatest increases in risk. The effect on the maximum drawdown is clear, except for the case of Wealthsimple, where not rebalancing would result in a −0.5% reduction in the maximum drawdown. For the other companies, the increase ranges from 0.3% for Nutmeg to 3.4% for SigFig.

In the area of PM, however, the effect is unclear. The performance of some robo-managers increases consistently across all the measures considered, whereas the performance of others is reduced. Some country-specific characteristics can be determined. All the UK robo-managers, except for Wealthify, increase their performance. The effect is unclear for the German robo-managers. Scalable (Germany) and Cominvest would experience an increase in performance, whereas Liqid, Quirion and Growney would undergo a reduction. Overall, it is clear that rebalancing has a considerable influence on a portfolio's risk–return positioning.

The next step is to evaluate the effect no rebalancing will have on the comparison of the robo-managers' performance. Table 15 shows the changes in the numbers of positive performance differences. It outlines the results for all the PM for SR, VR and SO.

The results show that, in most cases, a lack of rebalancing leads to a change in the numbers of positive performance differences ranging from zero to two. Wealthsimple is a clear outlier, as Wealthsimple exhibits a significant increase in medium-risk levels across all PM. Overall, for most robo-managers, stopping rebalancing does not lead to a change in performance ranking.

TABLE 16 Deviation of the mean values

This table reports the deviation of the mean values of the descriptive statistics, calculated as the difference between the scenarios with uniform threshold rebalancing with a trigger of 10% and the normal scenario. The descriptive statistics include the annual return, the annual volatility, the annual Value at Risk with a confidence level of 95% ($VaR_{95\%}$), the maximum drawdown (MD), the Sharpe ratio (SR), the VaR ratio (VR) and the Sortino ratio (SO). For each statistic, the country means is also reported in an additional row.

		Δr (%)	$\Delta\sigma$ (%)	$\Delta VaR_{95\%}$ (%)	ΔMD (%)	ΔSR	ΔVR	ΔSO
Germany	Scalable (Germany)	0.0	0.0	0.0	0.0	0.00	0.00	0.00
	Cominvest	0.0	0.0	0.0	0.0	0.00	0.00	0.00
	Liqid	0.2	0.1	−0.5	0.2	0.00	0.00	0.00
	Quirion	0.0	0.0	0.0	0.0	0.00	0.00	0.00
	Growney	0.2	0.2	−0.7	0.3	0.00	0.00	0.00
	Mean	0.1	0.1	−0.2	0.1	0.00	0.00	0.00
United Kingdom	Nutmeg	0.6	0.1	−0.1	−0.1	0.12	0.04	0.09
	Moneyfarm	0.0	0.0	0.0	0.0	0.00	0.00	−0.00
	Wealthsimple	0.3	0.0	0.2	−0.4	0.07	0.02	0.05
	Wealthify	0.4	0.2	−0.4	0.2	0.03	0.01	0.02
	Scalable (UK)	0.0	0.0	0.0	0.0	0.00	0.00	0.00
	Mean	0.3	0.1	−0.1	−0.1	0.05	0.01	0.03
United States	Schwab Intelligent	0.4	0.2	−0.7	0.3	−0.02	−0.01	−0.04
	Betterment	0.2	0.1	−0.5	0.1	−0.02	−0.01	−0.03
	Wealth Front	0.1	0.1	−0.3	0.1	−0.01	−0.00	−0.01
	SigFig	0.1	0.1	−0.5	0.2	−0.02	−0.01	−0.02
	Axos Invest	0.1	0.1	−0.4	0.2	−0.02	−0.01	−0.02
	Mean	0.2	0.1	−0.5	0.2	−0.02	−0.01	−0.02

Schwab Intelligent and Wealthify are still the only two robo-managers with double-digit numbers of positive performance differences across all risk levels.

In addition to omitting rebalancing, we further investigate the effect of standardizing the rebalancing method. A threshold-based rebalancing with a trigger criterion of 10% was thus assumed for all robo-managers. The results in Tables 16 and 17 clearly show that the effect is only very small in both the differences between the mean values and the comparison of the positive performance differences. In the case of robo-managers with this rebalancing procedure in the standard scenario, we find no changes in the descriptive statistics. This applies to Scalable (Germany), Cominvest, Moneyfarm and Scalable (UK). The return, risk measures and PM all change only very slightly. Thus, the choice and specification of rebalancing method have only a minor effect on the risk–return positioning of a portfolio, the only significant factor being that there is some form of rebalancing. The extent to which positive performance differences change is also only in the range of zero to two.

In a final step, in the following section, we analyze the effects of fees.

TABLE 17 Changes in the number of positive performance differences

This table presents the changes in the number of positive performance differences in the lowest-, medium- and highest-risk classes, calculated by the differences between the scenarios with uniform threshold rebalancing with a trigger of 10% and the normal scenario. The performance measures (PM) that are considered in this table are the Sharpe ratio (SR), the VaR ratio (VR) and the Sortino ratio (SO).

PM	SR			VR			SO		
Risk class (L = low, M = medium, H = high)	L	M	H	L	M	H	L	M	H
Scalable (Germany)	0	0	0	0	0	0	0	0	0
Cominvest	0	−1	0	0	1	0	0	−1	0
Liqid	0	1	0	0	1	0	0	1	0
Quirion	0	0	0	0	0	0	0	0	0
Growney	0	−1	0	0	−2	0	0	−1	0
Nutmeg	0	1	0	0	1	0	0	1	0
Moneyfarm	0	0	0	0	0	0	0	0	0
Wealthsimple	0	2	0	0	1	0	0	2	0
Wealthify	0	0	0	−1	0	0	0	0	0
Scalable (UK)	0	−1	0	0	−1	0	0	−1	0
Schwab Intelligent	2	0	0	2	0	0	0	0	0
Betterment	0	1	0	0	0	0	0	0	1
Wealth Front	0	0	0	0	0	0	0	0	−1
SigFig	0	0	0	0	0	0	0	0	0
Axos Invest	−2	−1	0	−1	−1	0	0	−1	0

4.4 | Impact of fees on PM

The costs of an investment are often a decisive criterion for its performance. The different fees of the 15 robo-managers were discussed in Section 2. The following analysis assesses how strongly fees influence the performance of the individual robo-managers. As a general rule, the percentage costs on the invested assets reduce a portfolio's return. The higher the costs, the greater the effect. To check the significance of the costs for our results, we show in Tables 18 and 19 the changes in the mean values of the descriptive statistics and in the numbers of positive performance differences when costs are neglected.

The impact on the portfolio statistics in Table 18 is clear. The returns increase in each case by exactly the amount of fees expressed as a percentage of AUM. There are no changes for the companies Schwab Intelligent and Axos Invest, as they do not charge fees, even in the standard scenario. Furthermore, volatility does not change for all the robo-managers, as the fees do not represent a risk-influencing variable. The changes in VaR and the maximum drawdown reflect the shift in the distribution of returns regarding costs. The VaR and maximum drawdown, therefore, also increase by approximately the same amount as the costs, although there is not a complete match, due to the random influences from the simulation. For PM, neglecting the costs leads to an increase in benefits. The increases in performance tend to be the largest for companies with relatively high costs.

TABLE 18 Deviation of the mean values

This table reports the deviation of the mean values of the descriptive statistics, calculated as the difference between the scenarios without cost and the normal scenario. The descriptive statistics include the annual return, the annual volatility, the annual Value at Risk with a confidence level of 95% ($VaR_{95\%}$), the maximum drawdown (MD), the Sharpe ratio (SR), the VaR ratio (VR) and the Sortino ratio (SO). For each statistic the country mean is also reported in an additional row.

		Δr (%)	Δσ (%)	$\Delta VaR_{95\%}$ (%)	ΔMD (%)	ΔSR	ΔVR	ΔSO
Germany	Scalable (Germany)	0.8	0.0	0.7	−0.3	0.28	0.11	0.27
	Cominvest	1.0	0.0	0.7	−0.9	0.20	0.05	0.13
	Liqid	0.5	0.0	0.4	−0.5	0.15	0.04	0.11
	Quirion	0.5	0.0	0.4	−0.3	0.22	0.07	0.17
	Growney	1.0	0.0	0.8	−0.8	0.33	0.10	0.26
	Mean	0.8	0.0	0.6	−0.5	0.24	0.08	0.19
United Kingdom	Nutmeg	0.5	0.0	0.4	−0.3	0.15	0.05	0.12
	Moneyfarm	0.7	0.0	0.5	−0.9	0.18	0.04	0.10
	Wealthsimple	0.7	0.0	0.6	−0.6	0.17	0.05	0.10
	Wealthify	0.8	0.0	0.6	−0.3	0.24	0.10	0.25
	Scalable (UK)	0.8	0.0	0.7	−0.6	0.27	0.09	0.22
	Mean	0.7	0.0	0.6	−0.5	0.20	0.07	0.16
United States	Schwab Intelligent	0.0	0.0	0.0	0.0	0.00	0.00	0.00
	Betterment	0.3	0.0	0.2	−0.2	0.15	6.03	0.85
	Wealth Front	0.3	0.0	0.2	−0.2	0.05	0.02	0.04
	SigFig	0.2	0.0	0.2	−0.2	0.05	0.01	0.03
	Axos Invest	0.0	0.0	0.0	0.0	0.00	0.00	0.00
	Mean	0.2	0.0	0.1	−0.1	0.05	1.21	0.19

However, there are exceptions to this basic trend, which become clear from comparing the companies Growney, Scalable (Germany) and Cominvest. This effect results from the mathematical fact that performance, as a relative indicator, always establishes a relation between a return and a risk indicator. Comparing two robo-managers whose returns increase but whose risk remains the same leads to the following situation, If the returns increase at the same rate, the increase in performance of the robo-manager with the lower initial risk level is greater.

There are only minor shifts in the changes in the numbers of positive performance differences, analogous to the previous results. The largest differences are for Growney's low and medium risk levels for the VR and SO with a value of three.

For all the other robo-managers, the change is in a range from zero to two. Schwab Intelligent and Wealthify are still the two companies with consistently double-digit numbers of positive performance differences. An influence of costs on the results of the study can thus clearly be established, but the effect is comparatively small. The fee differences are, therefore, not so great that this criterion would dominate the other factors developed in Section 2.2. Even if a robo-manager is associated with higher fees, they are mainly compensated for by better performance.

TABLE 19 Changes in the number of positive performance differences

This table presents the changes in the number of positive performance differences in the lowest-, medium- and highest-risk classes, calculated by the differences between the scenario without cost and the normal scenario. The performance measures (PM) that are considered in this table are the Sharpe ratio (SR), the VaR ratio (VR) and the Sortino ratio (SO).

PM	SR			VR			SO		
Risk class (L = low, M = medium, H = high)	L	M	H	L	M	H	L	M	H
Scalable (Germany)	2	0	0	2	0	1	3	0	0
Cominvest	−1	1	0	−1	2	0	−1	1	0
Liqid	−1	−2	0	−1	−1	0	−1	−1	−1
Quirion	2	0	0	2	−1	−1	1	−1	0
Growney	2	2	1	2	3	0	3	3	0
Nutmeg	−1	0	0	−1	0	−1	−1	0	0
Moneyfarm	1	0	2	1	0	0	1	0	0
Wealthsimple	0	2	−1	0	0	1	0	1	1
Wealthify	1	0	0	0	0	0	0	0	0
Scalable (UK)	0	0	1	0	0	1	0	0	1
Schwab Intelligent	1	0	−1	1	0	−1	0	0	−1
Betterment	0	0	−1	0	−1	0	0	−1	0
Wealth Front	−2	−1	−1	−2	−1	0	−2	−1	0
SigFig	−2	0	0	−2	0	0	−2	0	0
Axos Invest	−2	−1	0	−1	−1	0	−1	−1	0

5 | CONCLUSION

The following statements summarize our main results:

- The qualitative comparison of the robo-managers clearly shows considerable differences, based on the criteria developed.
- In particular, the portfolio models and the different asset structures explain different performance trends.
- The differences between the robo-managers are primarily company specific and are driven by the business model and focus on a specific customer clientele. Our study shows that country-specific characteristics, such as regulation, play only a minor role.
- Comparable risk–return exposures to low-, medium- and high-risk levels do not exist according to our quantitative evaluation.
- Across all PM, risk levels and robustness tests, only two robo-managers offer consistently double-digit numbers of positive performance differences in our simulation. These results can be described as significant.

- Different risk and PM were used to assess the risk and performance of the robo-managers, but these play only a minor role in the results, especially in terms of the numbers of positive performance differences.
- Each country has a particularly beneficial robo-manager, which is why each country is represented in the top three robo-managers based on our ranking.
- For risk–return positioning, a concrete rebalancing measure has only a minor effect; it is overall more important that some form of rebalancing be carried out.
- The fees have an impact on the results, but only a minor effect on the performance ranking.

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

REFERENCES

Annaert, J., van Osselaer, S., & Verstraete, B. (2009). Performance evaluation of portfolio insurance strategies using stochastic dominance criteria. *Journal of Banking & Finance*, *33*, 272–280. https://doi.org/10.1016/j.jbankfin.2008.08.002

Aziz, S., Dowling, M., Hammami, H., & Piepenbrink, A. (2021). Machine learning in finance: A topic modeling approach. *European Financial Management*, *50*, 1–27. https://doi.org/10.1111/eufm.12326

Bacon, C. R. (2013). *Practical risk-adjusted performance measurement*. Wiley.

BaFin. (2017). *Annual report 2017*. Federal Financial Supervisory Authority. https://www.bafin.de/DE/PublikationenDaten/Jahresbericht/jahresbericht_node.html

Baker, T., & Dellaert, B. G. C. (2018). Regulating robo advice across the financial services. *Iowa Law Review*, *103*, 713–750. https://doi.org/10.2139/ssrn.2932189

Bayón, P. S., & Vega, L. G. (2018). *Automated investment advice: Legal challenges and regulatory questions* (Policy Report, 37). Banking & Financial Services. https://ssrn.com/abstract=3226651

Behr, P., Güttler, A. & Miebs, F. (2008). Is minimum-variance investing really worth the while? An analysis with robust performance inference. Centre for Financial Research, Universität zu Köln. October 31, 2008. https://www.cfr-cologne.de/download/kolloquium/2009/behr_et_al.pdf

Belanche, D., Casaló, L. V., & Flavián, C. (2019). Artificial intelligence in FinTech: Understanding robo-advisors adoption among customers. *Industrial Management & Data Systems*, *119*, 1411–1430. https://doi.org/10.1108/IMDS-08-2018-0368

Chen, C. Y.-H., & Chiang, T. C. (2016). Empirical analysis of the intertemporal relationship between downside risk and expected returns: Evidence from time-varying transition probability models. *European Financial Management*, *22*, 749–796. https://doi.org/10.1111/eufm.12079

Cheng, X., Guo, F., Chen, J., Li, K., Zhang, Y., & Gao, P. (2019). Exploring the trust influencing mechanism of robo-advisor service: A mixed-method approach. *Sustainability*, *11*, 4917–4937. https://doi.org/10.3390/su11184917

Cogneau, P., & Zakamouline, V. (2013). Block bootstrap methods and the choice of stocks for the long run. *Quantitative Finance*, *13*, 1443–1457. https://doi.org/10.1080/14697688.2012.713115

Cuthbertson, K., Hayley, S., Motson, N., & Nitzsche, D. (2016). What does rebalancing really achieve? *International Journal of Finance & Economics*, *21*, 224–240. https://doi.org/10.1002/ijfe.1545

D'Acunto, F., Prabhala, N., & Rossi, A. G. (2019). The promises and pitfalls of robo-advising. *Review of Financial Studies*, *32*, 1983–2020. https://doi.org/10.1093/rfs/hhz014

D'Acunto, F., & Rossi, A. G. (2021). Robo-advising. In R. Rau, R. Wardrop, & L. Zingales (Eds.), *Handbook of technological finance*. Palgrave Macmillan. https://www.palgrave.com/gp/book/9783030651169

D'Hondt, C., Winne, R. D., Ghysels, E., & Raymond, S. (2019). Artificial intelligence alter egos: Who benefits from robo-investing? *SSRN Electronic Journal*. Advance online publication. 1–75. https://doi.org/10.2139/ssrn.3415981

Davison, A. C., & Hinkley, D. V. (1997). Bootstrap methods and their application, *Cambridge series in statistical and probabilistic mathematics*. Cambridge University Press.

Dichtl, H., Drobetz, W., & Wambach, M. (2016). Testing rebalancing strategies for stock-bond portfolios across different asset allocations. *Applied Economics*, *48*, 772–788. https://doi.org/10.1080/00036846.2015.1088139

Dowd, K. (2000). Adjusting for risk. *International Review of Economics & Finance*, *9*, 209–222. https://doi.org/10.1016/S1059-0560(00)00063-0

European Securities and Markets Authority. (2016). Report on automation in financial advice. https://esas-joint-committee.europa.eu/Publications/Reports/EBA%20BS%202016%20422%20(JC%20SC%20CPFI%20Final%20Report%20on%20automated%20advice%20tools).pdf

European Securities and Markets Authority. (2018). Final report. Guidelines on certain aspects of the MiFID II suitability requirements. https://www.esma.europa.eu/sites/default/files/library/esma35-43-869-_fr_on_guidelines_on_suitability.pdf

Fein, M. L. (2015). Robo-advisors: A closer look. *SSRN Electronic Journal*. Advance online publication. 1–33. https://doi.org/10.2139/ssrn.2658701

Ferson, W. E. (2013). Ruminations on investment performance measurement. *European Financial Management*, *19*, 4–13. https://doi.org/10.1111/j.1468-036X.2012.00657.x

Financial Conduct Authority (2017a). Finalised guidance. FG17/8: Streamlined advice and related consolidated guidance. September 2017. https://www.fca.org.uk/publications/finalised-guidance/fg17-8-streamlined-advice-consolidated-guidance

Financial Conduct Authority (2017b). Robo advice: An FCA perspective. 11 October 2017. https://www.fca.org.uk/news/speeches/robo-advice-fca-perspective

Financial Industry Regulatory Authority. (2016). Report on digital investment advice. https://www.finra.org/sites/default/files/digital-investment-advice-report.pdf

Foerster, S., Linnainmaa, J. T., Melzer, B. T., & Previtero, A. (2017). Retail financial advice: Does one size fit all? *The Journal of Finance*, *72*, 1441–1482. https://doi.org/10.1111/jofi.12514

Glaser, F., Iliewa, Z., Jung, D., & Weber, M. (2019). Towards designing robo-advisors for unexperienced investors with experience sampling of time-series data. *Information Systems and Neuroscience*, *2018*, 133–138. https://doi.org/10.1007/978-3-030-01087-4_16

Grischuk, P. (2017). Robo-advice: Automatisierte Anlageberatung in der Aufsichtspraxis, *BaFin Journal* (Vol. , pp. 18–22). Bonn and Frankfurt am Main: BaFin.

Gulden, J. (2019). *Automatisierte Geldanlage*. Springer Gabler. https://doi.org/10.1007/978-3-658-24054-7

Hall, P. (1992). *The bootstrap and Edgeworth expansion*. Springer.

Hambuckers, J., & Heuchenne, C. (2016). Estimating the out-of-sample predictive ability of trading rules: A robust bootstrap approach. *Journal of Forecasting*, *35*, 347–372. https://doi.org/10.1002/for.2380

Harvey, E. B., & Thel, S. (2019). *Investment management law and regulation* (3rd ed.). Wolters Kluwer Legal & Regulatory U.S.

Hock, M. (2020, February 2). Geldanlage vom Roboter. *Frankfurter Allgemeine Zeitung*, 26.

Hölscher, R., & Nelde, M. (2016). Performancemessung im Vermögensverwaltungsgeschäft. In P. Hoppe, & F. Keuper (Eds.), *Strategische Vermögensverwaltung: Konzepte–Instrumente–Entwicklung* (pp. 37–89). Logos.

James, J., & Yang, L. (2010). Stop-losses, maximum drawdown-at-risk and replicating financial time series with the stationary bootstrap. *Quantitative Finance*, *10*, 1–12. https://doi.org/10.1080/14697680903545596

Ji, M. (2017). Are robots good fiduciaries? Regulating robo-advisors under the Investment Adviser Act of 1940. *Columbia Law Review*, *117*, 1543–1583. www.jstor.org/stable/44392957

Jobson, J. D., & Korkie, B. M. (1981). Performance hypothesis testing with the Sharpe and Treynor measures. *The Journal of Finance*, *36*, 889–908. https://doi.org/10.2307/2327554

Jung, D., Dorner, V., Weinhardt, C., & Pusmaz, H. (2018). Designing a robo-advisor for risk-averse, low-budget consumers. *Electronic Markets*, *28*, 367–380. https://doi.org/10.1007/s12525-017-0279-9

Keppler, M. (1990). Risiko ist nicht gleich Volatilität. *Die Bank*, *11*, 610–614.

Kleijnen, J. P. C. (2012). Design and analysis of Monte Carlo experiments. In J. E. Gentle, W. K. Härdle, & Y. Mori (Eds.), *Handbook of computational statistics: Concepts and methods* (2nd ed., pp. 529–547). Springer. https://doi.org/10.1007/978-3-642-21551-3_18

Lazaro, C. (2019). The regulation of digital investment advice. Legal Studies Research Paper Series. *SSRN Electronic Journal*. Advance online publication. 1–13. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3314271

Linsmeier, T. J., & Pearson, N. D. (2000). Value at risk. *Financial Analysts Journal*, *56*(2), 47–67. https://doi.org/10.2469/faj.v56.n2.2343

Lückoff, P. (2011). *Mutual fund performance and performance persistence: The impact of fund flows and manager changes* (1st ed.). Gabler. https://doi.org/10.1007/978-3-8349-6527-1

Mansini, R., Ogryczak, W., & Speranza, M. G. (2015). Linear and mixed integer programming for portfolio optimization, *EURO advanced tutorials on operational research*. Springer International Publishing. https://doi.org/10.1007/978-3-319-18482-1

Memmel, C. (2003). Performance hypothesis testing with the Sharpe ratio. *Finance Letters*, *1*, 21–23. https://doi.org/10.1016/j.frl.2013.08.001

Möller, I. (2003). *Nichtlineare Abhängigkeiten bei finanzwirtschaftlichen Zeitreihen: Aktuelle Testverfahren am Beispiel einer Wechselkursanalyse* (1st ed.). Deutscher Universitäts-Verlag. https://doi.org/10.1007/978-3-322-81594-1

Monaco, S. M., Pershkow, A. W., Cruz, L. S., McCamman, P. M., Getsinger, A. D., & Kanter, A. (2017). US Securities and Exchange Commission's division of investment management issues guidance regarding robo-advisers. *Journal of Investment Compliance*, *18*(3), 26–33. https://doi.org/10.1108/JOIC-06-2017-0035

Nelde, M. M. (2019). *Anlageentscheidungen privater Investoren: Robo-Advisor als Instrument zur Rationalitätssicherung*. Shaker.

Patton, A., Politis, D. N., & White, H. (2009). Correction to "automatic block-length selection for the dependent bootstrap. *Econometric Reviews*, *28*, 372–375. https://doi.org/10.1080/07474930802459016

Pfiffelmann, M., Roger, T., & Bourachnikova, O. (2016). When behavioral portfolio theory meets Markowitz theory. *Economic Modelling*, *53*, 419–435. https://doi.org/10.1016/j.econmod.2015.10.041

Politis, D. N., & Romano, J. P. (1994). The stationary bootstrap. *Journal of American Statistical Association*, *89*, 1303–1313. https://doi.org/10.1080/01621459.1994.10476870

Politis, D. N., & White, H. (2004). Automatic block-length selection for the dependent bootstrap. *Economic Reviews*, *23*, 53–70. https://doi.org/10.1081/ETC-120028836

Puhle, M. (2019). The performance and asset allocation of German robo-advisors. *Society and Economy*, *41*, 331–351. https://doi.org/10.1556/204.2019.41.3.4

Rossi, A. G., & Utkus, S. P. (2020). The needs and wants in financial advice: Human versus robo-advising. *SSRN Electronic Journal*. Advance online publication. 1–76. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3759041

Rossi, A. G. & Utkus, S. P. (2021). Who benefits from robo-advising? Evidence from machine learning. SSRN Electronic Journal. Advance online publication. 1-73. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3552671

Scalable Capital. (2016). *Der Investmentprozess von Scalable Capital* [White paper]. https://de.scalable.capital/assets/3odztfgndkxn/53ulOKNnzGQ4gak2Ce44E8/39456e4e4052067151635333172b249b/Scalable_Capital_Whitepaper_WP04_DE.pdf

Scheurle, S. (2017). Essays in empirical personal finance: Individual investor reactions to low interest rates and robo-advice (Doctoral thesis). Johann Wolfgang Goethe—Universität Frankfurt am Main. https://www.econbiz.de/Record/essays-in-empirical-personal-finance-individual-investor-reactions-to-low-interest-rates-and-robo-advice-scheurle-sebastian/10011637827

Sharpe, W. F. (1994). The Sharpe ratio. *Journal of Portfolio Management*, *21*(1), 49–58. https://doi.org/10.3905/jpm.1994.409501

Stanek, B. (2020, May 1). *How much does a financial advisor cost?* Smart Asset. https://smartasset.com/financial-advisor/financial-advisor-cost

Tertilt, M., & Scholz, P. (2019). To advise, or not to advise: How robo-advisors evaluate the risk preferences of private investors. *SSRN Electronic Journal*, *107*, 158–62. https://doi.org/10.3905/jwm.2018.21.2.070

Tsay, R. S. (2010). *Analysis of financial time series* (3rd ed.). Wiley.

U.S. Securities and Exchange Commission (2017). Guidance update robo-advisers. February 2017. https://www.sec.gov/investment/im-guidance-2017-02.pdf

Vinod, H. D., & Morey, M. R. (2000). Confidence intervals and hypothesis testing for the Sharpe and Treynor performance measures: A bootstrap approach. In Y. S. Abu-Mostafa, B. LeBaron, A. W. Lo, & A. S. Weigend (Eds.), *Computational finance 1999* (pp. 25–39). MIT Press.

Zeisberger, S., Langer, T., & Trede, M. (2007). A note on myopic loss aversion and the equity premium puzzle. *Finance Research Letters*, *4*, 127–136. https://doi.org/10.1016/j.frl.2006.11.002

How to cite this article: Helms, N., Hölscher, R., & Nelde, M. (2022). Automated investment management: Comparing the design and performance of international robo-managers. *European Financial Management*, *28*, 1028–1078. https://doi.org/10.1111/eufm.12333

APPENDIX A

Table A1.

TABLE A1 Performance differences based on the Sortino ratio (SO)

This table reports the performance differences based on the Sortino ratio (SO) between all the robo-managers (categorized according to the lowest, medium and highest risk levels). These performance differences are calculated as the mean of 400,000,000 SO differences equal to SO(investment strategy 1) − SO(investment strategy 2). *, **, ***Significance at the 10%, 5%, 1% levels, respectively.

Lowest risk level

ΔSO = SO(1) − SO(2)	Investment strategy 1														
Investment strategy 2	Scalable (Ger)	Cominvest	Liqid	Quirion	Growney	Nutmeg	Moneyfarm	Wealthsimple	Wealthify	Scalable (UK)	Schwab	Betterment	Wealthfront	SigFig	Axos Invest
Scalable (Ger)	0.00	0.01	-0.05	-0.19	0.08	-0.45*	-0.5**	-0.27	0.65*	-0.57**	0.52*	4.74**	0.19	0.23	0.52*
Cominvest		0.00	-0.06	-0.20	0.07	-0.46**	-0.51**	-0.28*	0.64*	-0.57**	0.51*	4.73**	0.18	0.22	0.51*
Liqid			0.00	-0.14	0.14	-0.4*	-0.44**	-0.21	0.7**	-0.51**	0.57*	4.79***	0.25	0.28	0.57*
Quirion				0.00	0.27	-0.26*	-0.31*	-0.08	0.84**	-0.37*	0.71**	4.93***	0.38*	0.42*	0.71**
Growney					0.00	-0.53**	-0.58***	-0.35*	0.56*	-0.65***	0.44*	4.66**	0.11	0.14	0.44*
Nutmeg						0.00	-0.05	0.18	1.1***	-0.12	0.97***	5.19***	0.64**	0.68**	0.97***
Moneyfarm							0.00	0.23*	1.14***	-0.07	1.02***	5.24***	0.69***	0.72***	1.02***
Wealthsimple								0.00	0.91***	-0.3*	0.79***	5.01***	0.46*	0.49*	0.79**
Wealthify									0.00	-1.21***	-0.13	4.09*	-0.45*	-0.42	-0.13
Scalable (UK)										0.00	1.09***	5.31***	0.76***	0.79***	1.09***
Schwab											0.00	4.22*	-0.33	-0.30	0.00
Betterment												0.00	-4.55**	-4.52**	-4.22*
Wealthfront													0.00	0.03	0.33
SigFig														0.00	0.29
Axos Invest															0.00

Medium risk level

Investment strategy 2	Scalable (Ger)	Cominvest	Liqid	Quirion	Growney	Nutmeg	Moneyfarm	Wealthsimple	Wealthify	Scalable (UK)	Schwab	Betterment	Wealthfront	SigFig	Axos Invest
Scalable (Ger)	0.00	-0.78**	-0.76**	-0.59*	-0.68*	-0.25	-0.98***	-0.82**	0.09	-0.19	0.43	-0.66*	-0.64*	-0.83**	-0.79**
Cominvest		0.00	0.02	0.19	0.10	0.53*	-0.20	-0.03	0.88**	0.59*	1.21***	0.12	0.14	-0.05	-0.01
Liqid			0.00	0.17	0.08	0.51*	-0.21	-0.05	0.86**	0.57*	1.19***	0.10	0.12	-0.07	-0.03
Quirion				0.00	-0.09	0.35	-0.38*	-0.22	0.69*	0.4*	1.02**	-0.07	-0.05	-0.24	-0.20
Growney					0.00	0.43*	-0.3*	-0.13	0.78**	0.49*	1.11***	0.02	0.04	-0.15	-0.11
Nutmeg						0.00	-0.73**	-0.57*	0.34	0.06	0.68*	-0.41*	-0.39*	-0.58*	-0.54*
Moneyfarm							0.00	0.16	1.07***	0.79***	1.41***	0.32*	0.34*	0.15	0.19
Wealthsimple								0.00	0.91**	0.62*	1.24***	0.15	0.17	-0.02	0.02
Wealthify									0.00	-0.29	0.33	-0.76*	-0.74*	-0.93**	-0.89**
Scalable (UK)										0.00	0.62*	-0.47*	-0.45*	-0.64*	-0.6*
Schwab											0.00	-1.09**	-1.07**	-1.26***	-1.22***
Betterment												0.00	0.02	-0.17	-0.13
Wealthfront													0.00	-0.19	-0.15
SigFig														0.00	0.04
Axos Invest															0.00

Highest risk level

Investment strategy 2	Scalable (Ger)	Cominvest	Liqid	Quirion	Growney	Nutmeg	Moneyfarm	Wealthsimple	Wealthify	Scalable (UK)	Schwab	Betterment	Wealthfront	SigFig	Axos Invest
Scalable (Ger)	0.00	-0.99***	-0.83**	-0.67*	-0.79**	0.18	-1***	-0.85**	-0.44*	-0.13	-0.02	-0.94***	-0.94**	-0.92**	-1.07***
Cominvest		0.00	0.15	0.31*	0.19	1.17***	-0.01	0.13	0.54*	0.86***	0.97***	0.04	0.05	0.07	-0.08
Liqid			0.00	0.16	0.04	1.01***	-0.16	-0.02	0.39*	0.71**	0.82**	-0.11	-0.10	-0.08	-0.23
Quirion				0.00	-0.12	0.85**	-0.32*	-0.18	0.23	0.55*	0.66*	-0.27	-0.26	-0.24	-0.39*
Growney					0.00	0.97***	-0.21	-0.06	0.35*	0.67**	0.78**	-0.15	-0.15	-0.12	-0.27*
Nutmeg						0.00	-1.18***	-1.03***	-0.62*	-0.30	-0.20	-1.12***	-1.12***	-1.09***	-1.25***
Moneyfarm							0.00	0.14	0.55**	0.87***	0.98***	0.06	0.06	0.08	-0.07
Wealthsimple								0.00	0.41*	0.73**	0.84**	-0.09	-0.08	-0.06	-0.21
Wealthify									0.00	0.32	0.43*	-0.5*	-0.49*	-0.47*	-0.62**
Scalable (UK)										0.00	0.11	-0.82**	-0.81**	-0.79**	-0.94***
Schwab											0.00	-0.92***	-0.92***	-0.9***	-1.05***
Betterment												0.00	0.00	0.03	-0.13
Wealthfront													0.00	0.02	-0.13
SigFig														0.00	-0.15
Axos Invest															0.00

Table A2.

TABLE A2 Comparison of the Sortino ratio differences

This table reports the performance comparison parameters based on the Sortino ratio (SO) for the 15 robo-managers listed in Table 1. These parameters include the return, the standard deviation, the number of positive performance differences for the Sortino ratio (ΔSO), the maximum performance difference, the minimum performance difference and the resulting spread.

Risk level															
Low	Scalable (Ger)	Cominvest	Liqid	Quirion	Growney	Nutmeg	Moneyfarm	Wealthsimple	Wealthify	Scalable (UK)	Schwab	Betterment	Wealthfront	SigFig	Axos Invest
# positive ΔSO	6	7	5	4	8	2	1	3	13	0	12	14	9	10	11
Max. difference	0.57	0.57	0.51	0.37	0.65	0.12	0.07	0.30	1.21	-0.07	1.09	5.31	0.76	0.79	1.09
Min. difference	-4.74	-4.73	-4.79	-4.93	-4.66	-5.19	-5.24	-5.01	-4.09	-5.31	-4.22	4.09	-4.55	-4.52	-4.22
Spread	5.31	5.31	5.31	5.31	5.31	5.31	5.31	5.31	5.31	5.24	5.31	1.21	5.31	5.31	5.31
Medium	Scalable (Ger)	Cominvest	Liqid	Quirion	Growney	Nutmeg	Moneyfarm	Wealthsimple	Wealthify	Scalable (UK)	Schwab	Betterment	Wealthfront	SigFig	Axos Invest
# positive ΔSO	12	4	5	9	6	10	0	2	13	11	14	7	8	1	3
Max. difference	0.98	0.20	0.21	0.38	0.30	0.73	-0.15	0.16	1.07	0.79	1.41	0.32	0.34	0.15	0.19
Min. difference	-0.43	-1.21	-1.19	-1.02	-1.11	-0.68	-1.41	-1.24	-0.33	-0.62	0.33	-1.09	-1.07	-1.26	-1.22
Spread	1.41	1.41	1.41	1.41	1.41	1.41	1.26	1.41	1.41	1.41	1.07	1.41	1.41	1.41	1.41
High	Scalable (Ger)	Cominvest	Liqid	Quirion	Growney	Nutmeg	Moneyfarm	Wealthsimple	Wealthify	Scalable (UK)	Schwab	Betterment	Wealthfront	SigFig	Axos Invest
# positive ΔSO	13	2	7	9	8	14	1	6	10	11	12	3	4	5	0
Max. difference	1.07	0.08	0.23	0.39	0.27	1.25	0.07	0.21	0.62	0.94	1.05	0.13	0.13	0.15	-0.07
Min. difference	-0.18	-1.17	-1.01	-0.85	-0.97	0.18	-1.18	-1.03	-0.62	-0.30	-0.20	-1.12	-1.12	-1.09	-1.25
Spread	1.25	1.25	1.25	1.25	1.25	1.07	1.25	1.25	1.25	1.25	1.25	1.25	1.25	1.25	1.18