UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

Factor Investing and ESG Integration in Regime-switching Models

An Empirical Study on ESG Factor Integration Using Infinite Hidden Markov Models

Arien Haghshenas* Martin Karim†

Supervisor: Dr. Viktor Elliot
Graduate School - Centre for Finance

Abstract

ESG investing is an active area of interest, both for the investment and academic communities. However, research is inconclusive on the financial benefits of integrating ESG factors in portfolio construction. In this thesis, we propose a novel approach to examining the informational content in ESG data using an infinite Hidden Markov framework to capture market regimes. Our objective is to find if ESG factors can increase a portfolio's risk-return characteristics by capturing additional effects that other factors do not. We build a baseline model with the factors Value, Quality, Growth, Momentum, and Risk. Next, we add layers of ESG data to the baseline model and analyze the effect on portfolio performance. Our findings show that the infinite Hidden Markov Model portfolios consistently outperform the market index EURO STOXX 50. However, we do not observe value added by ESG scores in our regime-switching factor investing framework.

Keywords: *ESG, Hidden Markov Models, Factor investing, Machine learning, Portfolio construction, Regime-switching models*

JEL Classification: *C15, E44, G11, G17*

*gushagar@student.gu.se
†gusvkalvma@student.gu.se

Acknowledgements

First and foremost, we want to express our gratitude to our supervisor Dr. Viktor Elliot at School of Business, Economics and Law at the University of Gothenburg. Without his assistance and active participation in every step of the process, as well as guidance in the right direction, the challenge would have been immense. We sincerely appreciate your patience and support. We would like to thank MDOTM Ltd., notably Alessio Fiore and Dr. Francesco Orsini, for their assistance. Thanks to their input and tremendous knowledge in Machine Learning and AI, we were able to successfully develop our model. Thank you for your commitment and availability. Lastly, we would like to thank the School of Business, Economics and Law, and the Center of Finance at the University of Gothenburg for these academic years and for providing us with the necessary resources to conduct our thesis.

Contents

1 Introduction **1**

2 Literature review **4**

2.1 Environmental, Social, and Governance information in Investments 4

2.2 Hidden Markov Models in Finance 5

2.3 Factor Investing 7

3 Theoretical Framework **9**

3.1 Markov Chain 9

3.2 Hidden Markov Model 10

3.3 A Simple Regime Switching Model 13

3.4 Dirichlet Process 14

3.5 Hierarchical Dirichlet Process 15

3.6 Infinite Hidden Markov 15

3.7 Fama McBeth Regression 16

4 Data **17**

4.1 Stock data 17

4.2 Time period 18

4.3 Factor Scores 18

4.4 Macroeconomic data 20

4.5 Refinitiv Eikon's ESG database 21

4.5.1 Limitations of ESG Dataset selection 23

5 Methodology **24**

5.1 An overview of the algorithm 24

5.2 Z-scores 25

5.3 Infinite Hidden Markov Model iterations 26

5.4 Factor scoring and portfolio building 27

5.5 Portfolio performance evaluation metrics 28

6 Results and analysis **31**

6.1 Final iHMM iteration results 31

6.2 Average Factor score weights 31

6.3 Portfolio Performance 32

6.3.1 Full period performance (2003-2022) 32

6.3.2 Recent years performance (2019-2022) 34
6.4 Robustness tests 35

7 Discussion **37**

8 Conclusion **39**
8.1 Further research 40

9 References **41**

A Appendix **46**

1 Introduction

Environmental, Social, and Governance (ESG) factors are increasingly integrated into traditional portfolio management as demand from investors for ESG portfolios has increased. Asset managers deploy ESG strategies, using methodologies from screening-based exclusions to a comprehensive integration of ESG factors into the security selection and portfolio development process. Meanwhile, academia has inconclusive empirical evidence of the financial benefits of integrating ESG factors into the investment process. ESG scores are widely considered the best proxy for companies' ESG behavior. It is, therefore, crucial for asset managers and investors to understand ESG scores' informational content, and researching ESG datasets in asset allocation requires further consideration. While numerous ESG strategies are analyzed to find if they generate alpha, e.g., Nagy, Kassam & Lee (2016); Chen & Mussalli (2020); Bruno, Esakia & Goltz (2021), there is less research on ESG factors' financial performance in regime-switching models, despite evidence of a varying impact of ESG investing in different regimes. In this thesis, we use a regime-switching infinite Hidden Markov Model (iHMM) to build a dynamic stock-picking portfolio using ESG factors and traditional factors[1]. We contribute to the growing body of literature on ESG in asset allocation by examining whether the informational content of ESG scores improves portfolio performance in terms of returns, volatility, and other performance measures when dynamically adjusted for market conditions.

Analyzing ESG data with an iHMM is a subject of interest given empirical evidence that ESG portfolios have displayed discrepancies in financial performance across regimes. Ma (2019) finds that ESG ratings have a larger effect on stock returns in bear markets than in bull markets in the short run. Specifically, Ma (2019) finds that high ESG-rated stocks generally outperform low ESG-rated stocks in terms of abnormal returns in a sluggish market and vice versa. Giese, Nagy & Lee (2020) separates ESG scores into their respective pillars to examine their effect independently, establishing that the E, S, and G pillars have a varying impact on simulated portfolios depending on the investment horizon. These findings imply that the aggregated ESG scores may pick up individual ESG pillars' effects across different regimes. It is, therefore, worth investigating further by using both aggregated ESG scores and the E, S, and G pillar scores as factors separately to break down ESG datas' informational content even further.

This thesis also contributes to the growing literature on factor investing using regime-switching models. The iHMM extends the Hidden Markov Model (HMM) using a Hierarchical Dirichlet Process to make the regime specifications data-driven. Previous research using HMMs, e.g., Fons et al. (2021); Nguyen & Nguyen (2021), has required the number of regimes to be specified in advance, typically done through Bayesian Information Criterion (BIC) or by selecting the model with the highest performance,

[1] "Traditional factors" refer to the five investment style factors Value, Growth, Quality, Momentum, and Risk. We refer to these factors as "traditional factors" for the remainder of the thesis.

resulting in a different number of regimes used through papers. It is noteworthy that the HMM has shown properties that improve model performance more than traditional investment methods only using factor investing (Kim, Jeong & Lee 2019). Fons et al. (2021) and Nguyen & Nguyen (2021) use an HMM with factor investing to fully utilize HMM's ability to capture nonlinearity in time series data, yielding superior financial performance in line with Kim et al. (2019) findings. Therefore, not only do we consider the iHMM as a means to adjust for ESG factors' financial performance across regimes but also as a model that has empirically yielded better financial performance than typical factor investing models.

By deploying an iHMM using traditional factors, we create a baseline stock portfolio model and compare it to an index. The comparison will confirm if the iHMM is a suitable factor model capable of yielding satisfactory financial performance. Next, we add layers of ESG factors to the model. Our aim is not to create ESG portfolios with a certain ESG score threshold but to investigate whether the informational content of our ESG data affects the constructed portfolios and picks up any effects that the traditional factors do not. By comparing performance metrics of the portfolios integrating ESG factors to the baseline portfolio model, we discern the usefulness of ESG data in asset allocation when dynamically adjusting stock allocations to changes in market conditions. The results are of interest for investors focusing on portfolios' risk-return characteristics. To the best of our knowledge, this paper is the first study using mentioned approach to analyze the informational content of ESG data across regimes. Rather, previous research has focused on if ESG factors behave differently across regimes, or compares low to high rated ESG portfolios in different markets, e.g., Ma (2019); Singh (2020).

This thesis is motivated by several articles and research divided into two main groups: (i) regime-switching Hidden Markov Models used for stock selection, and (ii) ESG factor investing. Our primary sources in the first category are the work of Wang, Lin & Mikhelson (2020), Nguyen & Nguyen (2021), and Fons et al. (2021). Nguyen & Nguyen (2021) uses an HMM to create a global stock selection using five stock factors in which the regimes of the six global economic indicators are predicted using an HMM. The author's portfolio outperformed the trading portfolios based on stocks in the All Country World Index, replicating their successful work in their previous paper Nguyen & Nguyen (2015) focusing on the S&P 500. Similarly, Wang et al. (2020) observes that using an HMM for asset allocation can yield superior portfolio returns compared to traditional factor models, e.g., Fama-French Three-factor Model or Carhart Four-Factor Model. As noted, empirical evidence favors the HMM for stock selections. Fons et al. (2021), however, highlights the issue that the number of latent states has to be known in advance or selected through BIC, which is not always effective, or by using a greedy approach, in which the model with the best performance is chosen. Fons et al. (2021) suggests an iHMM to address this. The second source of motivation is papers that analyze if ESG factors can generate alpha. Bruno et al. (2021) researches common strategies empirically shown to generate alpha (e.g., ESG Momentum Alpha, ESG Combined Alpha, and ESG Overall and Component Alpha). Bruno et al. (2021) finds that these strategies, indeed, generate alpha. However, risk-adjusting these results shrink the alphas close to zero. The authors imply that failure to properly risk-adjust the results is the source of alpha in previous literature. Thus, we formulate our research question as follows:

How does ESG data affect portfolios' risk-return characteristics in an infinite Hidden Markov Model?

We examine Western European equities over the period 2003 to 2022 as a leading geographical market regarding ESG questions, both in investments and company behavior. European investors have shown greater sentiment toward ESG investing. Their US counterpart has remained skeptical about the benefits of sustainable investments for financial performance and consequently is less inclined to incorporate ESG factors in their investment process (Kaiser 2020). 22% of US investors, compared to 8% of European investors, believe that ESG data is not material for investment purposes, and enforcing ESG selection on their portfolios would violate their fiduciary duty (Amel-Zadeh & Serafeim 2018).

The iHMM is used to predict the regimes of five macroeconomic variables (stock market index, implied and actual volatility index, inflation index, and GDP per Capita) and finds the time periods in the past during which these indicators have a combination of regimes that is similar to those predicted. We use the factors Value, Quality, Growth, Momentum, and Risk to create the baseline portfolio model before successively adding layers of ESG factors provided by the Refinitiv Eikon ESG dataset. The factors are given weights through Fama-MacBeth regressions. A selection of 50 stocks is used based on how well they score given the weights in regime i. We rebalance the portfolios when the iHMM predicts a new regime before iterating the process again, creating a dynamic stock selection based on market conditions. We do not find evidence of ESG factors providing additional information to the model that increases the portfolio risk-return characteristics. Our results indicate that ESG data do not capture additional information that our traditional factors do not. The results also show that the baseline portfolio model has the best financial performance over the full period 2003 - 2022 and a more recent period 2019 - 2022.

The rest of this study is organized as follows: Section 2 discusses previous literature on regime-switching models, factor investing, and ESG in asset allocation. Section 3 provides the theoretical framework. Section 4 presents the data and ESG scoring methodology. Section 5 outlines the methodology of our study. Section 6 provides an analysis and evaluation of the model, and the results are discussed in Section 7. Finally, the conclusion and potential future research are presented in section 8.

2 Literature review

This section provides a review of the existing literature in the research area. First, we present literature in the field of ESG investments. ESG is a broad concept, and we map out different applications of ESG in investments and research related to the definition and financial performance of ESG metrics. Second, literature on HMMs in finance and stock selection is presented, followed by literature on Factor and Smart beta investing.

2.1 Environmental, Social, and Governance information in Investments

ESG investing is a broad field, and there is no universal measurement for the E, S, and G scores. Discrepancies are also found in ESG portfolio investing, where different investment approaches address various objectives. A high-level breakdown can divide ESG investing into three primary categories depending on the investment objective: First, ESG integration to improve a portfolio's risk-return characteristics. Second, there is value-based investing, aiming to align an investor's portfolio with a set of values, norms, and beliefs. The third is impact investing, in which investors use their capital to push for a better change in ESG areas (Giese et al. 2019). This thesis focus on the first objective.

Amel-Zadeh & Serafeim (2018) finds that the most common reason for using ESG data is to improve investment performance in accordance with the first objective. Secondary motivations were client demand, product strategy, bringing change in companies, and ethical reasons last, coinciding with the three objectives described by Giese et al. (2019). For ESG to be material to investment performance, corporate ESG implementation requires economically meaningful effects, which studies have documented. For instance, Cheng, Ioannou & Serafeim (2014) finds that ESG disclosure is associated with lower capital constraints and lower cost of capital (Dhaliwal, Li, Tsang & Yang 2011). Furthermore, industry-specific materiality classifications identify ESG information that is value relevant and predictive of firms' future financial performance, and the disclosure of such information is linked to less stock price synchronicity Amel-Zadeh & Serafeim (2018).

As studies document significant economic effects, other research studies the relationship between ESG characteristics and traditional risk factors. Melas, Nagy & Kulkarni (2017) discovers that ESG scores have a positive correlation with size, quality, and low volatility. Furthermore, Melas et al. (2017) finds that ESG integration generally improves the historical risk-adjusted performance of many typical passive and factor investment strategies and tilts the strategies toward larger companies with higher profitability, more stable earnings, lower leverage, and lower volatility. These studies imply that ESG datasets may not have additional informational content not already captured by traditional factors.

Academia finds empirical evidence both for and against the financial benefit of ESG factor integration in asset allocation. Halbritter & Dorfleitner (2015) investigates the relationship between corporate social and financial performance based on ESG ratings. The authors examine different portfolios with low and high ESG scores, analyzing compound ESG scores and respective pillar scores. Neither the ESG portfolio nor pillar portfolios show significant return differences between companies featuring high and low ESG rating levels. As such, the relationship between ESG ratings and returns is questioned, and Halbritter & Dorfleitner (2015) concludes that investors should not expect abnormal returns by trading different portfolios of high and low-rated firms.

Giese et al. (2020) hypothesized that ESG pillars are unlikely to have an equal impact on company stock performance. In their publication, the authors find that the E, S, and G pillars relate differently to companies' financial performance. For instance, depending on the time horizon, industry, and weighting scheme, the relationship between the simulated portfolios and performance varied in contrast to Halbritter & Dorfleitner (2015) observations. The Governance pillar proved to be far more significant than Environmental or Social over a short period (one year). Environmental and Social indicators were more significant over long periods and notably visible in companies' stock performance. While research prior to Giese et al. (2020) already examine individual pillar scores' financial performance, e.g., Alessandrini & Jondeau (2020), Bruder et al. (2019), and Drei et al. (2019), Giese et al. (2020) findings have an important implication for research on ESG data, serving as a further foundation for breaking down aggregated ESG scores into its' pillars to analyze ESG data more profoundly. Following these findings, Bruno et al. (2021) uses both aggregated ESG ratings and E, S, and G variables independently in their models, with results varying in magnitude for the different pillars when testing different ESG strategies for stock selection.

2.2 Hidden Markov Models in Finance

Baum & Petrie (1966) described the HMM in a series of statistical publications, and the HMM is since applicated in various fields, notably in finance. Hamilton (1989) introduced the HMM in finance by proposing the use of regime-switching models to identify economic cycles using GNP series. The HMM's popularity in finance stems from the HMM's ability to capture various properties from financial return series, such as time-varying correlations, skewness, and kurtosis while offering a reasonable approximation even if the underlying model is unknown. For instance, Bae, Kim & Mulvey (2014) deploys an HMM to identify market regimes using different asset classes and manages to reduce the risk during left tail events in the portfolio with the regime information. As for regime specification, the HMM requires manual input. Basic intuition of financial markets classifies regimes as states in which the market goes up, down, or remains unchanged. Following this intuition, Liu, Xu & Zhao (2011) and Ma, MacLean, Xu & Zhao (2011) analyze time-varying risk premiums using a regime-switching model with three regimes. However, regime specifications are based on informational data on macroeconomic variables, and classifying regimes could use more complexity. Accordingly, Guidolin & Timmermann (2008) finds evidence of four economic regimes rather than three in size and value factors that capture time variations in mean returns, volatilities, and return correlations. Difficulties selecting the number of states is one of the limitations of the HMM, therefore, other models require consideration, such as the iHMM.

Beal, Ghahramani & Rasmussen (2002) discusses the standard HMM and its limitations. First, the maximum likelihood estimation approach ignores the model's complexity, resulting in difficulties in avoiding over-or underfitting. Secondly, as stated, the model's structure requires definition in advance. Attempts to approximate a full Bayesian analysis of the HMM that integrates across rather than optimizes the parameters are motivated partially by mentioned challenges. For most real-world issues, the basic HMM modeling assumes that data is generated by some discrete state variable and can take one of the multiple values, which is a strong assumption. Beal et al. (2002) proposes an HMM with a limitless number of hidden states, namely the infinite Hidden Markov Model, to address the limitations of the HMM. They show that it is possible to extend an HMM to have a countably infinite number of hidden states. Through the theory of Dirichlet processes, they implicitly integrate out the infinitely many transition parameters, leaving only three hyper-parameters that are data-driven. These three hyper-parameters define a hierarchical Dirichlet process capable of capturing a rich set of transition dynamics.

The iHMM is yet to see a widespread application in financial studies, although the HMM is used in many finance pieces of research, particularly to predict financial market regimes. To date, little work has been done on the impact of regime-switching models on factor investing, though gaining popularity in recent years, e.g., Kim et al. (2019), Wang et al. (2020), Fons et al. (2021), and Nguyen & Nguyen (2021). These studies deploy HMMs and have resulted in more robust models that outperform traditional factor strategies. Specifically, Wang et al. (2020) uses an HMM that rotates between two-factor models. The results yield superior model performance to that of either factor model it is based on and enhances portfolio performance. Fons et al. (2021) addresses the concerns of smart beta strategies, specifically, sensitivity to market fluctuations and often severe short-term drawdown (peak-to-trough decline) with fluctuating financial performance across cycles. They build a dynamic asset allocation system using an HMM to manage the short-term risk of cyclicality and under-performance. Kim et al. (2019) uses an HMM to identify the phases of individual assets and propose an investment strategy using price trends yielding superior financial performance, noting that the HMM reflects the asset selection effect in Jensen's alpha, Fama's Net Selectivity, and Treynot-Mazuy model. Finally, Nguyen & Nguyen (2021) uses an HMM to select stocks from the global stock market that outperform the All Country World Index (ACWI), any single stock factor, or the simple average of a set of five stock factors.

Further use of regime-switching models, albeit scarcely, has been in ESG asset allocation research. For instance, Ma (2019) finds that in a regime-switching model, higher ESG rating stock portfolios significantly outperform their lower counterpart in bear markets, implicating a positive financial effect of ESG in a downward economy. More recent work involves different regimes that capture financial and non-financial crisis periods during the Covid-19 pandemic. Pluciennik & Janik (2022) studies sustainable stock indices using a Markov Switching Approach in which the first regime identifies as normal market conditions and the second regime as pandemic conditions. The two regimes display different mean, standard deviation, kurtosis, and skewness, indicating differences in financial performance over the regimes. Similarly, Singh (2020) analyzes regime switches and finds that ESG asset allocation results in overperformance and improves during the pandemic due to the ESG approach proving to be more resilient in an uncertain environment and a "safer" bet for investors. Other research has also shown discrepancies in the performance of ESG portfolios in comparison to conventional portfolios in times of distress or under Bull/Bear regimes, e.g., Nofsinger & Varma (2014) and la Torre-Torres et al. (2019).

2.3 Factor Investing

Factor investing refers to two main types of factors: macroeconomic and style. Macroeconomic factors and equity returns have been researched extensively with an established correlation, e.g., Ratanapakorn & Sharma (2007) and Sirucek (2012). Integrating macroeconomic factors into models has also led to the identification of different economic regimes by capturing the mean and variance of the market index, typically called bull- and bear market regimes (Kole & Van Dijk 2017). Style factors have earned a premium over long periods, reflecting exposure to systemic risk, and are grounded in the academic literature Bender et al. (2013). The Capital Asset Pricing Model (CAPM) is one of the early factor models that attempts to explain what drives equity returns. The CAPM has a single risk factor to model the risk premium of an asset class and established a foundation of modern financial theory in the 1960s (See Treynor (1961), Sharpe (1964), Lintner (1965), and Mossin (1966)[2]). While the CAPM proposes a single factor, it later set the groundwork for multi-factor models. Ross (1976) developed the Arbitrage Pricing Theory (APT), which infers that a financial asset's projected return is represented as a function of numerous macroeconomic factors or theoretical market indexes. Unlike the CAPM, the APT does not specify what these factors should be. The number and character of these variables are expected to fluctuate over time and between marketplaces. As a result, the issue of developing factor models became, and continues to be, mostly empirical. Fama & French (1992) and Carhart (1997) introduce other common factors models (three-factor and four-factor models) based on the ATP to capture new risk premia. For instance, fundamental factors are most prominent today and are based on the multi-factor models, namely equity factors such as Value, Quality, Growth, Size, and Momentum (Bender et al. 2013).

[2]Sharpe (1964) submitted the initial version of his CAPM to the Journal of Finance. However, it was rejected and not published until after a revision. Other authors' supplementary contributions to Sharpe's (1964) paper resulted in the established CAPM, thus the joint recognition regarding the development of the CAPM.

Smart beta is a low-cost, systematic implementation of factor investing in which securities are chosen based on exposure to a factor historically linked to greater returns. Portfolio development with Smart beta strategies induces opportunities to beat market returns over a longer period compared to cap-weighted index constructions, as shown by Blitz & Swinkels (2008) and Meziani (2014). Similarly, Amenc et al. (2015) finds significant evidence that systematic Smart beta strategies outperform cap-weighted benchmarks in the long run. An issue with Smart betas has been lacking robustness due to data mining and non-robust weighting methodologies that typically arise when back-testing Smart beta strategies. Still, Fons et al. (2021) tests a variety of portfolio construction techniques using smart beta strategies, and the portfolios show an improvement in risk-adjusted returns, especially on more return-oriented portfolios when used in an HMM.

3 Theoretical Framework

This section maps out the theoretical framework that is the foundation of our model. First, we describe Markov Chains that build the framework for the HMM, followed by detailing the HMM's process and simple regime-switching models. Next, the Dirichlet Process (DP) framework is outlined for transition sequences, followed by the Hierarchical Dirichlet Process (HDP) that bridges the HMM and iHMM. Lastly, we outline the theoretical framework for the iHMM and the Fama-McBeth regression framework for factor scoring.

3.1 Markov Chain

Markov Chain is primarily a mathematical tool in stochastic processes and is today employed in various fields. It is a stochastic process that meets the Markov property, which states that while the present is known, the past and future are independent. It suggests that by having knowledge of the current state of a process, extra information about its previous states is not needed to make the best possible forecast about the future. Following this logic, predictions about a stochastic process's future state become simplified (Sericola 2013).

There are two main types of Markov Chains: Discrete and Continuous. The Discrete Markov Chain is an evolving system through time steps in which the changes only occur at a discrete-time value. For example, take a board game such as Snakes and Ladders. The pieces move around the board according to a die roll. It does not matter how the pieces arrive at their current position (i.e., the history of the system) at the beginning of round n. What matters is the current stage of the board and the positioning of the pieces (i.e., the current state of the system). This illustrates how a change in discrete chains only occurs during someone's turn, i.e., at a discrete-time.

Changes in the Continuous Markov Chain can occur at any given time along with a continuous interval. For example, a number of customers visiting a supermarket can occur at any time t during operating hours, and the visits are independent. Knowing the total number of customers at a specific time does not give any predictability for future customer visits (assuming the visitations follow a Poisson Process).

In accordance with Sericola (2013), we define the Discrete Markov chain (*Definition 1*) and the Continuous Markov chain (*Definition 2)* in mathematical terms as:

Definition 1. *A stochastic process* $X = \{X_n, n \in \mathbb{N}\}$ *in a countable space* S *is a discrete-time Markov chain if:*
For all $n \geq 0, X_n \in S$
For all $n \geq 1, X_n$ *and for all* $i_0, ..., i_{n-1}, i_n \in S,$ *we have:*
$P\{X_n = i_n | X_{n-1} = i_{n-1}, ..., X_0 = i_0\} = P\{X_n = i_n | X_{n-1} = i_{n-1}\}$

Definition 2. *A stochastic process* $X = \{X_t, t \in \mathbb{R}^+\}$ *with values in a countable space* S *is a continuous-time Markov chain if:*
For all $n \geq 0, X_n,$ *For all instant* $0 \leq s_0 < ... < s_n < t,$ *and for all states* $i_0, ..., i_n, i, j \in S,$ *we have:*
$P\{X_t = j | X_s = i, X_{s_n} = i_n, ..., X_{s0} = i_0\} = P\{X_t = j | X_s = i\}$

3.2 Hidden Markov Model

An HMM is a stochastic process following a Markov chain/process described in the previous section but in which the underlying Markov process is not observed (hidden). It can, however, be observed through another set of stochastic processes. Let us call the hidden Markov process X and the observable process Y for which outcome is influenced by X in some way. Then, the objective is to learn about X by observing Y (Rabiner & Juang 1986).

To illustrate, we present how a Markov Process (Diagram 1) can be extended into an HMM (Diagram 2) as described by Blunsom (2004). Diagram 1 illustrates a simple model for a stock market index. The model has three states, *Bull, Bear* and *Even*, and three index observations *up, down* and *unchanged*. The model is a finite-state automation with a probabilistic transition between states. Given a sequence of observations, for example, up-down-down, we can easily verify that the state sequence that produced those observations is Bull-Bear-Bear and the probability of the sequence is simply the product of the transitions, in this case, $0.2 \times 0.3 \times 0.3$.

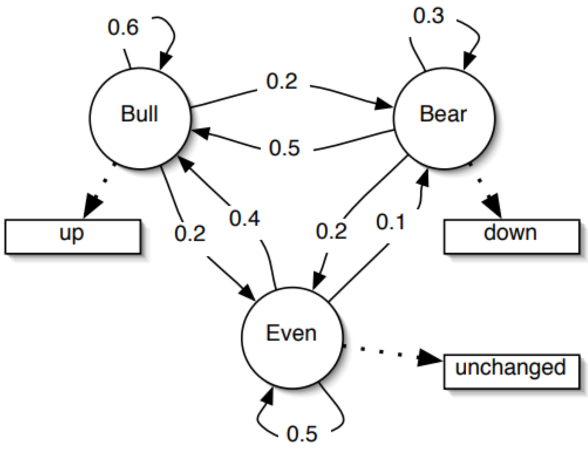

Diagram 1: Markov process example (Blunsom 2004)

Diagram 2 extends the previous example of the Markov process into an HMM. The revised model permits each state to emit all observation symbols with a finite probability. The model becomes more expressive and capable of representing the intuition that, for example, a bull market has both up and down days but has a higher probability of "up". The main distinction is that if we have the same observation up-down-down, we cannot determine which state sequence created these observations, hence the name "Hidden" Markov Model. It is still possible to calculate how likely the model is to generate the sequence and which state sequence was most likely to generate the given observation (Blunsom 2004).

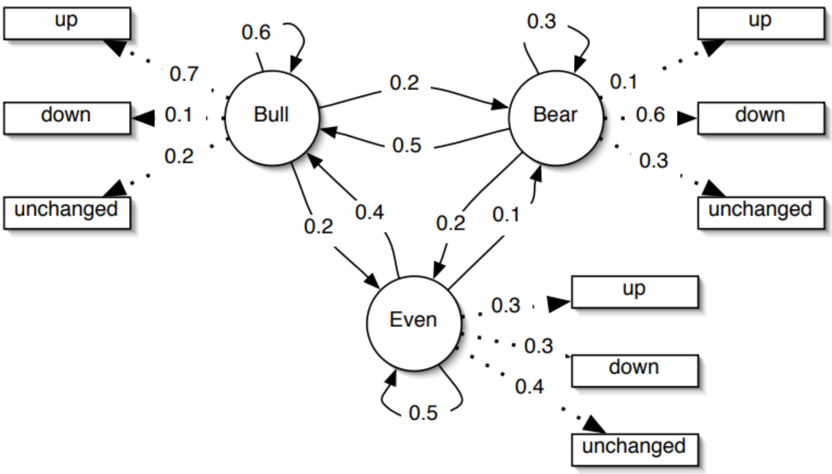

Diagram 2: Hidden Markov Model example (Blunsom 2004)

Rabiner (1989) characterizes an HMM by five elements:

1. There are a finite number N of states in the model that have measurable distinctive properties within the states, even if the states are hidden or not rigorously defined as to what a state is. The states are denoted as $S = \{S_1, S_2, ..., S_N\}$ and the state at time t as q_t.

2. There are a finite number M of distinct observation symbols per state. The individual symbols are denoted $S = \{v_1, v_2, ..., v_M\}$.

3. There is a state transition probability distribution $A = \{a_{ij}\}$ where the increments a_{ij} are defined as:

$$a_{ij} = P(q_{t-1} = S_j | q_t = S_i), 1 \leq i, j \leq N \tag{1}$$

4. There is a corresponding observation symbol probability distribution in state $j, B = \{b_j(k)\}$, where the increments $b_j(k)$ are defined as:

$$b_j(k) = P(v_k \text{ at } t | q_t = S_j), 1 \leq j \leq N, 1 \leq k \leq M \tag{2}$$

5. There is an initial state distribution $\pi = \{\pi_i\}$ where:

$$\pi_i = P(q_1 = S_i), 1 \leq i \leq N \tag{3}$$

In short, the values $\{N, M, A, B, \text{ and } \pi\}$ provide a complete specification of the HMM where N *and* M are the model parameters and $\{A, B, \text{ and } \pi\}$ are the probability measures. Using these provides the observation sequence $O = \{O_1, O_2, ..., O_T\}$ as follows:

1. Choose the initial state $q_1 = S_i$ in accordance with the initial state distribution π

2. Set t = 1

3. Set $O_t = v_k$ in accordance with the symbol probability distribution in state $S_i; b_i(k)$

4. Transit to the new state $q_t + 1 = S_i$ in accordance with the state transition probability distribution for state $S_i; a_{ij}$

5. set $t = t + 1$; and return to step 3 until t = T

3.3 A Simple Regime Switching Model

A basic Markov State Switching model allows parameters to switch states. We call it a Simple Regime Switching Model (SRSM). Put an SRSM in the context of a simple financial market. The mean and variance that are Markov switching would change depending on the state of the market. A classic example is the stock market following a Bull or Bear market pattern. A Bull market has a positive trend with low volatility, and investors typically go long. In contrast, a Bear market has a negative trend and higher volatility than a Bull market, and investors tend to go short or be more defensive. It is known in the SRSM that the markets follow the characteristics described in Bull and Bear markets, in which a positive mean and low volatility is expected in a Bull market, and vice versa. In short, the SRSM provides the mean, volatility, and probability for the two different states (Hardy 2001).

The return for an SRSM is assumed to be as follows:

$$r_t = \mu_{s_t} + \sigma_{s_t}\varepsilon_t \tag{4}$$

where r_t is a time series of returns, S_t is a Markov chain with k possible states, and the error term ε_t is an i.i.d process with $t = 1, ..., T$. Assume $k = 2$, meaning there are two different regimes or states. We define S_t as:

$$S_t = \begin{cases} 1 \text{ with probability } \pi, \\ 2 \text{ with probability } 1-\pi, \end{cases} \tag{5}$$

The transition matrix for the Markov chain is:

$$P^* = \begin{pmatrix} p_{11} & p_{21} \\ p_{12} & p_{22} \end{pmatrix} \tag{6}$$

Looking at equation 6, the diagonal value p_{11} and p_{22} represent the probability of staying in regime 1 or 2, respectively. This implies that $p_{12} = 1 - p_{11}$ and $p_{21} = 1 - p_{22}$, which in turn represents the probabilities of switching from regime 1 to 2 and regime 2 to 1. r_t is modeled as follows:

$$r_t = \begin{cases} \mu_1 + \sigma_1\varepsilon_t \text{ if } S_t = 1, \\ \mu_2 + \sigma_2\varepsilon_t \text{ if } S_t = 2, \end{cases} \tag{7}$$

for the two states. The error terms ε_t are i.i.d $N(0, 1)$ and:

$$\varepsilon_t \sim \begin{cases} N(\mu_1, \sigma_1^2) \text{ if } S_t = 1, \\ N(\mu_2, \sigma_2^2) \text{ if } S_t = 2, \end{cases} \tag{8}$$

The outlined framework is a basic version that has been used with a different number of regimes to capture time variations in mean returns, volatility, and return correlations. Furthermore, traditional factor investing has also deployed the model as discussed in the literature review. The model has a shortcoming in which it is necessary to know the structure in advance, which can be addressed using an iHMM. We continue this section with the theoretical framework building up the iHMM.

3.4 Dirichlet Process

First, we examine the statistics of the hidden state transitions from a specific state $s_t = i$ with k finite number of states. The transition probabilities given the *i:th* row in the transition matrix can be interpreted as mixing proportions for $s_t + 1$ that we define as $\pi = \{\pi_1, ..., \pi_k\}$. Drawing n samples $\{c_1, ..., c_n\}$ from a discrete indicator variable can have the values $\{1, ..., l\}$ with the properties π. The joint distribution of these indicators is multinomial:

$$P(c_1, ..., c_n|\pi) = \prod_{j=1}^{k} \pi_j^{n_j}, \text{ with } n_j = \sum_{n'=1}^{n} \delta(c_{n'}, j) \tag{9}$$

where the Kronecker delta function ($\delta(a, b) = 1 \iff a = b$ and otherwise 0) is used to count the number n_j times that $s_{t+1} = j$ has been drawn. It becomes possible to see what happens to the distribution of these indicators when integrating out the mixing proportions π under a conjugate prior. Providing the mixing proportions a symmetric Dirichlet prior with positive concentration hyperparameter β yields:

$$P(\pi|\beta) \sim Dirichlet(\beta/k \dots \beta/K) = \frac{\Gamma(\beta)}{\Gamma(\beta/k)^k} \prod_{j=1}^{k} \pi_j^{\beta/k-1} \tag{10}$$

where π is restricted to be on the simplex of mixing proportions that sum to 1. Integrate out π under this prior to get:

$$P(c_1 \dots, c_n|\beta) = \int d\pi P(c_1 \dots c_n|\pi) = \frac{\Gamma(\beta)}{\Gamma(n+\beta)} \prod_{j=1}^{k} \frac{\Gamma(n_j + \beta/k)}{\Gamma(\beta/k)} \tag{11}$$

Thus, the probability of a particular sequence of indicators is only a function of the counts $\{n_1 \dots n_k\}$. The conditional probability of an indicator c_d given the setting of all other indicators (c_{-d}) is given by:

$$P(c_d = j|c_{-d}, \beta) = \frac{n_{-d,j} + \beta/k}{n - 1 + \beta} \tag{12}$$

where $n_{-d,j}$ is the count as in equation 9 with the *d:th* indicator removed. The self-reinforcing property of equation 12 (c_d) is more likely to choose a state that is already popular. A key property of the Dirichlet Process that is at the very heart of the model is the expression for equation 13 where we take the limit as the number of hidden states k goes to infinity:

$$P(c_d = j | c_{-d}, \beta) = \begin{cases} \frac{n_{-d,j}}{n-1+\beta} \; j \in 1 \dots K \text{ i.e., represented,} \\ \\ \frac{\beta}{n-1+\beta} \text{ for all unrepresented } j \text{ combined} \end{cases} \tag{13}$$

where K is the number of represented states that cannot be infinite since n is finite. β is interpreted as the number of pseudo observations of $\pi = \{1/k \dots 1/k\}$, i.e., the strength of belief in the symmetric prior. β is the "innovation" or error parameter in the infinite limit that controls for the tendency of the model to populate a previously unrepresented state.

3.5 Hierarchical Dirichlet Process

The Hierarchical Dirichlet Process is a set of DPs coupled through a shared random base measure also drawn from a DP. More specifically, $G_k \sim DP(\alpha, G_o)$ with a shared base measure G_0, interpreted as the mean of G_k, and $\alpha > 0$ as the concentration parameter which governs the variability around G_0. A smaller α implies a larger variability. The shared base measure itself is given a DP prior: $G_0 \sim DP(\gamma, H)$, with H as the global base measure. The stick-breaking construction for the HDP demonstrates that the random measures can be expressed as:

$$G_0 = \sum_{k'=1}^{\infty} \beta_{k'} \delta_{\phi_{k'}} \tag{14}$$

and

$$G_k = \sum_{k'=1}^{\infty} \pi_{kk'} \delta_{\phi_{k'}} \tag{15}$$

where $\delta \sim GEM(\gamma)$ is the stick-breaking construction for DPs $\pi_k \sim DP(\alpha, \beta)$ and each $\phi_{k'} \sim H$ independently.

3.6 Infinite Hidden Markov

By identifying each G_k (see equation 15) by describing both the transition probabilities $\pi_{kk'}$ from state k to k' and the emission distributions parameterized by ϕ_{k*}, we can formally define the iHMM as:

$$\begin{aligned} \beta &\sim GEM(\gamma), \\ \pi_k | \beta &\sim DP(\alpha, \beta), \\ \phi_k &\sim H, \end{aligned} \tag{16}$$

$$\begin{aligned} s_t | s_{t-1} &\sim Multinomial(\pi_{s_{t-1}}), \\ y_t | s_t &\sim F(\phi_{s_t}), \end{aligned} \tag{17}$$

which is illustrated graphically in Figure 3.

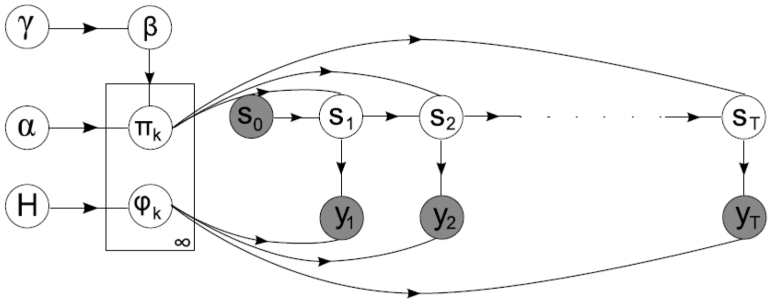

Figure 3: Visualisation of the iHMM model (Van Gael et al. 2008)

Thus, it becomes possible to move from a model with a finite number of states, which has to be pre-specified, to a model with an infinite possible number of states. The iHMM opens for possibilities beyond the basic bull-bear framework, using a data-driven approach to set the number of states.

3.7 Fama McBeth Regression

We use cross-section regression as the base for our scoring system in our algorithm through the Fama-MacBeth two-step approach outlined by Fama & MacBeth (1973). We use the method to investigate the effect of factors on the returns. The two-step parameter estimation is as follows:

1. Regress each n asset returns against the m proposed risk factors:

$$\begin{aligned}
&R_{1,t} = \alpha_1 + \beta_{1,F_1} F_{1,t} + \beta_{1,F_2} F_{2,t} + \cdots + \beta_{1,F_m} F_{m,t} + \varepsilon_{1,t} \\
&R_{2,t} = \alpha_2 + \beta_{2,F_1} F_{1,t} + \beta_{2,F_2} F_{2,t} + \cdots + \beta_{2,F_m} F_{m,t} + \varepsilon_{2,t} \\
&\vdots \\
&R_{n,t} = \alpha_n + \beta_{n,F_1} F_{1,t} + \beta_{n,F_2} F_{2,t} + \cdots + \beta_{n,F_m} F_{m,t} + \varepsilon_{n,t}
\end{aligned} \tag{18}$$

2. Regress all asset returns for each time period T against the previous estimates to determine the risk premium for each factor:

$$\begin{aligned}
&R_{i,1} = \gamma_{1,0} + \gamma_{1,1}\hat{\beta}_{i,F_1} + \gamma_{1,2}\hat{\beta}_{i,F_2} + \cdots + \gamma_{1,m}\hat{\beta}_{i,F_m} + \varepsilon_{i,1} \\
&R_{i,2} = \gamma_{2,0} + \gamma_{2,1}\hat{\beta}_{i,F_1} + \gamma_{2,2}\hat{\beta}_{i,F_2} + \cdots + \gamma_{2,m}\hat{\beta}_{i,F_m} + \varepsilon_{i,2} \\
&\vdots \\
&R_{i,T} = \gamma_{T,0} + \gamma_{T,1}\hat{\beta}_{i,F_1} + \gamma_{T,2}\hat{\beta}_{i,F_2} + \cdots + \gamma_{T,m}\hat{\beta}_{i,F_m} + \varepsilon_{i,T}
\end{aligned} \tag{19}$$

4 Data

This section follows a description of the data and the motivation behind the selected data. To summarize, the datasets are collected mainly from Refinitiv Eikon, from which ESG data, factor scores, stock data, and stock index data build our panel data. We use two different sets of weekly data. (i) Stock data with factors for the portfolio construction, and (ii) Macroeconomic data for the iHMM. The ESG data is collected on a yearly basis as reported. Complementary macroeconomic data is collected from Eurostat. The stock and factor data ranges from 2003 until 2022 over 17 countries in Western Europe, totaling 1839 different equities across 29 sectors. The equity data includes companies that have been delisted, thus reducing survivorship bias. The macroeconomic data for the iHMM ranges from 1999 to 2022. The handling of the data has been through Python.

4.1 Stock data

The stock data consists of equities from 17 countries based on Refinitiv's definition of Western Europe, with 1839 equities across 29 sectors (see Table 1 for country details). More European investors believe that ESG is material to investment decisions compared to their US counterparts, as noted by Amel-Zadeh & Serafeim (2018), hence the use of European stock data in this thesis. The geographical area of Western Europe is considered to be the more developed area of Europe with robust financial markets (Refinitiv Eikon 2022*b*). Thus, equities are liquid and available to most institutional and other investors whereas other parts of Europe, such as eastern Europe, have other types of risks and may be more restrictive to investors (for example, the Russian/Ukraine conflict in 2022 that resulted in regulated markets in Russia). The countries in focus are also classified as high-income and developed countries by the World Bank, further increasing the robustness of the dataset. For the same reasons, the included equities have a market capitalization of a minimum of 100 million euros to use a dataset of liquid equities available for all investors.

Table 1 presents changes in the dataset from before to after the screening of companies. Pre-screening consists of all listed companies with over 100 million euros in market capitalization. Post-screening companies have an ESG rating and fulfill the requirements to obtain factor scores (for example, companies without earnings are removed as they cannot get a Value score). 2275 companies in total are removed from the final dataset. While it is a substantial percentage of the companies (-55%), we deem the remaining set sufficient for our study. Moreover, we include all companies that fulfill our requirements in the dataset for the period they do, even if they are delisted at a later stage since we use dynamic stock picking. See Table A.1 in the appendix for sector details.

Table 1: Distribution of companies per country

Country	# Companies pre-screening	# Companies post-screening	Change
Austria	47	29	-38%
Belgium	93	45	-52%
Denmark	84	59	-30%
Finland	102	65	-36%
France	492	152	-69%
Germany	675	239	-65%
Iceland	22	4	-82%
Ireland	20	13	-35%
Italy	205	101	-51%
Luxembourg	12	1	-92%
Netherlands	118	53	-55%
Norway	216	79	-63%
Portugal	23	14	-39%
Spain	143	68	-52%
Sweden	369	225	-39%
Switzerland	281	158	-44%
United Kingdom	1212	534	-56%
Total	4114	1839	-55%

4.2 Time period

The main concern of this study is the effect of ESG data and is the primary motivation for the time period selected for data gathering. While stock and factor data is available for an extensive period, ESG data is more restrictive. Still, Refinitiv's ESG database stretches back to 2002 and provides a sufficient time frame for this study, including a range of possible market regimes for the iHMM. More specifically, the data consists of points from 2002-01-01 up to 2022-04-01. Moreover, a uniform ESG disclosure has not been and still is not a requirement across companies (some European countries have imposed requirements in recent years), and ESG data may in general not be adequately available nor robust before the beginning of the 2000s. We use macroeconomic data ranging from 1999 to 2022. Macroeconomic data is used to train the iHMM model. Thus, it becomes advantageous to use more data if it is available to have a larger training set, resulting in more robust regime predictions and motivating the wider range of data used compared to the ESG and stock data.

4.3 Factor Scores

We use six different factor scores; the five traditional scores Value, Quality, Growth, Momentum, and Risk, calculated using individual factors, and ESG factor scores. Fons et al. (2021) uses 25 different factors from the five factor families (Value, Growth, Momentum, Quality, Size, and Volatility/Risk/Beta). Similarly, Nguyen & Nguyen (2015) and Nguyen & Nguyen (2021) also use factors but a different set (Free cash flow/Enterprise value, Earnings/Price, Sales/Enterprise, and Long-term sale growth). The selected composition of the traditional factors is based mainly on Fons et al. (2021) by taking a smart beta investing approach. See Table 2 for the scores and factor breakdown selected for this thesis.

Table 2: Score factor components

Score	Components
ESG:	ESG, Environmental, Social, and Governance
Value:	Price to Earnings, Price to Book, Price to Cash flow, and Free Cash flow Yield
Quality:	Return on Assets, Return on Equity, EBIT margin, and Operating Margin
Growth:	Price to Sales growth, Enterprice Value to Sales growth, Free Cash flow growth, and Earnings per Share growth
Momentum:	1, 3, 6 and 12 months stock price momentum
Risk:	1, 3, 6, and 12 months stock price volatility

In contrast to Fons et al. (2021), we do not include "Size" in our factors, even though it is one of the most prominent in factor models. Since we are using a Z-score methodology (described in detail in section 5, methodology) for all our variables, it means that in the case of company size, we have to take a stance on if larger companies or smaller companies should receive a higher score. Some investors prefer larger and safer companies with a lower expected return, while others prefer smaller companies that are riskier but have a larger expected return. Thus, to keep the analysis more neutral and avoid any bias, we decide not to include company size as a factor. We investigate whether we exclude vital information from the model when removing Size by conducting robustness tests in section 6.4. The results yield worse performance for all portfolios and verify that we can proceed with our method.

Table 3 displays a correlation table for all factor scores. Overall, the correlation between the traditional factors is low, indicating that the indicators are independent in representing the economics, while the ESG factors show a higher correlation within each pillar. See the Table A.3 in Appendix A for descriptive statistics.

Table 3: Factor scores correlation table

Variables	Momentum	Risk	Value	Quality	Growth	ESG	Environmental	Social	Governance
Momentum	1.00								
Risk	0.03	1.00							
Value	-0.15	0.01	1.00						
Quality	0.06	0.30	-0.05	1.00					
Growth	0.22	0.01	-0.04	0.06	1.00				
ESG	-0.04	0.11	0.09	-0.03	-0.01	1.00			
Environmental	-0.04	0.10	0.13	-0.03	0.00	0.84	1.00		
Social	-0.04	0.11	0.09	-0.03	-0.01	0.89	0.70	1.00	
Governance	-0.03	0.05	0.04	-0.02	-0.01	0.71	0.40	0.43	1.00

4.4 Macroeconomic data

We acknowledge the potential challenge of deciding on proper representative macroeconomic factors due to the diversity of Europe compared to, for example, a single country such as the United States. There is also the issue of selecting x number of variables to use in the model. Using few macroeconomic indicators may result in the model being too basic. However, use too many, and it may become difficult for the model to identify patterns in the historical data as more regime combinations become possible, especially with the use of an iHMM.

Nguyen & Nguyen (2015) finds that the S&P 500 performs significantly differently across various states of four macroeconomic variables: inflation (consumer price index), industrial production index (INDPRO), stock market index (S&P 500), and market volatility (VIX). The underlying explanation is that these macroeconomic variables exhibit at least two states: bull/bear for the stock index, inflation/deflation for inflation, and low/high volatility for market volatility. In accordance with Nguyen & Nguyen (2015), but with a focus on European markets, three macroeconomic variables are selected based on their significant effects on stock prices; one for the European stock market, the corresponding volatility index, and European inflation. We also add a fourth variable measuring the actual market volatility in order to give the model more depth. Furthermore, following Zhou et al. (2020) findings that there is a correlation between firm-level ESG practices and GDP per capita, we add an additional macroeconomic variable, GDP per Capita, as a proxy for the "ESG market", ending up at a total of five macroeconomic factors. See table 4 for details.

Table 4: Macroeconomic factor details

Macroeconomic factors	Details
Stock Market Index:	EURO STOXX 50 index created by STOXX. Owned by Deutsche Börse Group. Weekly close prices of the index are used as the stock market indicator.
Implied Volatility Index:	Weekly data. VSTOXX index based on EURO STOXX 50 real-time option prices. Reflects the market expectations of near-term to long-term volatility by measuring the square root of the implied variance across all options of a given time to expiration.
Actual Volatility:	Used to capture the actual volatility to complement implied volatility. We calculate the weekly volatility of the EURO STOXX 50 index.
Inflation Index:	Calculated monthly. Based on the Harmonized Index of Consumer Prices for the European Economic Area. Used by ECB to calculate inflation levels.
GDP per Capita:	GDP per Capita for the EU. Calculated on quaterly basis.

We scale the data by normalizing it. We calculate the percentage difference for all the variables besides inflation, which is already the change in the Harmonized Index of Consumer Prices, to have all macroeconomic data in the same unit.

4.5 Refinitiv Eikon's ESG database

ESG scores by Refinitiv Eikon are widely used, both in academic research and the investment management industry. Refinitiv ESG data has been mentioned in over 1,200 academic papers by October 2020, and assets managers such as BlackRock use the database. The World Economic forum referenced Refinitiv ESG data in a whitepaper in 2019 and the ESG data was analyzed as one of three key rating providers (Berg, Fabisik & Sautner 2020); (Boffo & Patalano 2020); (Eltogby, Brown & Corrigan 2019). Accordingly, Refinitiv ESG offers "one of the most comprehensive ESG databases in the industry", covering over 80% of the global market cap across more than 10,000 global companies in 76 countries (Refinitiv 2022). Given the richness of Refinitiv Eikon's ESG database, we select the database for this thesis.

The ESG scores build upon relative performance. Scores are based on Environmental and Social performance in a given company's sector, and the Governance score is based on the country of incorporation. Given the variety of importance of the different E, S, and G factors across industries, Refinitiv map each metrics' materiality for each industry. The core of Refinitivs' methodology is company disclosure with applied weighting based on material or immaterial data points to either negatively or positively impact company reporting. The data is continuously updated on a weekly basis. However, ESG disclosure by companies is typically done on a yearly basis, resulting in the ESG data updating once a year. Following this, we use yearly ESG data points in this thesis. An overview of the ESG methodology is presented in Figure 4.

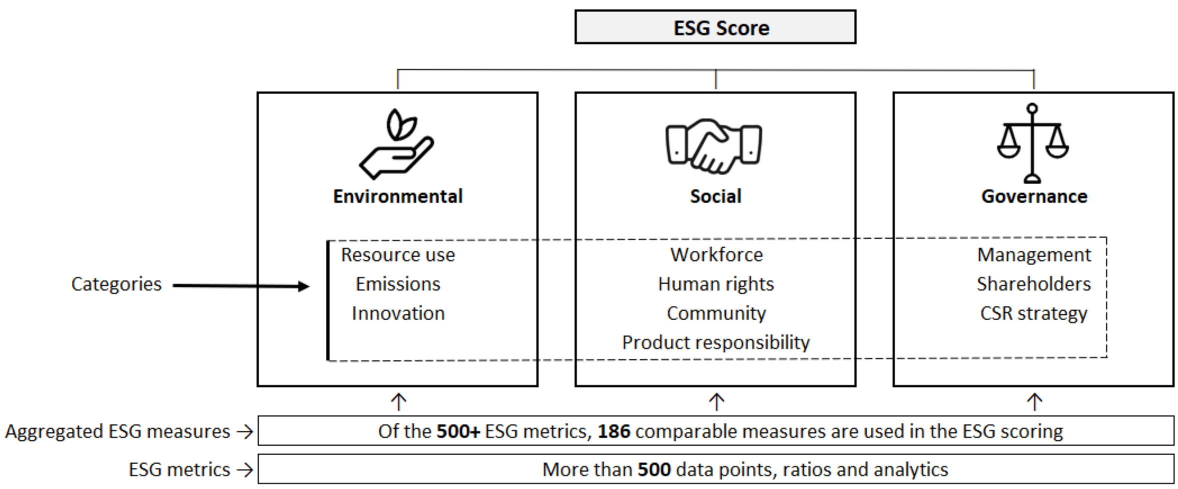

Figure 4: Refinitiv ESG Score structure, based on Refinitiv Eikon (2022*a*)

The score structure is as follows: over 500 company-level measures with a subset of 186 of the most comparable and material measures per industry compiles the overall company assessment and scoring process[3]. These data points are grouped into ten categories on which each ESG pillar score is based upon. For Environmental, we have Resource use, Emissions, and Innovation. Social in ESG covers the categories Workforce, Human rights, Community, and Product responsibility. Lastly, for Governance, we have Management, Shareholders, and Corporate social responsibility (CSR) strategy (see Table A.2 in the appendix score definitions). The respective ESG pillar score is normalized to percentages ranging between 0 and 100, and a relative sum of the category weights is calculated based on the sector a given company operates within. Lastly, an aggregated ESG score is given to reflect a company's overall ESG performance.

Figure 5 presents the monthly development of ESG scores. The y-axis is the average ESG score across our stocks and the x-axis shows datestamps. There is a clear positive development in the average ESG score until 2019, where the average score somewhat declines. Notably, the average ESG score has increased more than 30% from 37.5 to north of 50 during 2002 to 2022. This is in line with the positive trajectory of ESG integration and awareness. We take this into account and create two portfolios to compare the results, one with the full period and another portfolio using more recent data.

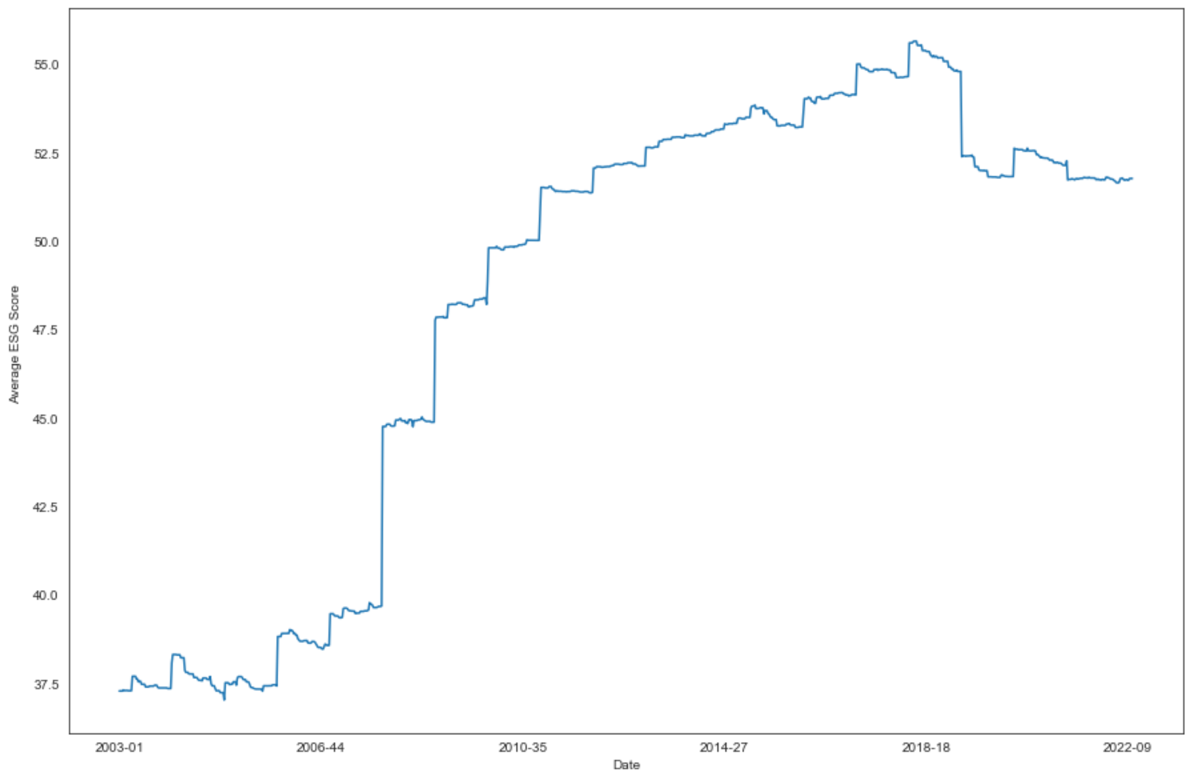

Figure 5: Average ESG score development

[3]Details of the 500 company-level ESG measures and the subset of 186 measures is available upon request at Refinitiv.

4.5.1 Limitations of ESG Dataset selection

A key challenge in conducting ESG activities, either research or portfolio construction, prevails in measuring a firm's "ESG quality". There is a need for quantifiability of how well a firm performs with respect to ESG criteria, hence the use of ESG scores constructed by professional data providers. While ESG scores are the best proxy available, they are not without issues. Amel-Zadeh & Serafeim (2018) discusses the important limitation to using ESG information, namely, the lack of reporting standards. Consequently, there is a lack of comparability, reliability, quantifiability, and timeliness. Other papers are written on this, for example, Billio et al. (2021) finds that there is a lack of commonality in the definition of ESG (i) characteristics, (ii) attributes, and (iii) standards in defining E, S, and G components. Billio et al. (2021) also provides evidence that heterogeneity in rating criteria can lead to agencies having opposite opinions on the same evaluated companies and consequently creating different benchmarks. These issues are evident in Refinitiv Eikon's ESG database. Berg, Fabisik & Sautner (2020) compares ESG scores with a sample between 2011 and 2017 to the same sample extracted in September 2020 in Refinitiv Eikon's ESG database. During this period, Refinitiv Eikon adjusted its scoring methodology, which came into effect on April 6, 2020, and retroactively modified historical scores in the database. 13% of the sample was subject to an upgrade, and 87% downgraded. No score remained constant. The overall ESG scores in the rewritten version showed on average 20,6% lower scores than the initial version.

The results presented by Berg et al. (2020) give weight to the other literature mentioned on the issue of ESG databases. The implications for this research are a lack of comparability to research before Refinitiv Eikons re-scoring and other ESG databases commonly used (such as the MSCI ESG database). Since the ESG data vendors do not provide a full breakdown of the scoring process and are ever-changing, the reliability of empirical studies may be questionable in the long run. The method proposed in this paper will add ESG data as a layer of information to find if it is informational or not, thus, will be able to process and compare different datasets.

5 Methodology

This section presents the methodology of our study. First, the reader is given an overview of the algorithm to better understand the different steps taken in the model. Next, we break up the different parts of the model and present them more in-depth, starting with Z-scoring, followed by a description of the iHMM iteration process, factor scoring through Fama-MacBeth regression, portfolio building, and finally, portfolio evaluation metrics.

5.1 An overview of the algorithm

We describe the algorithm as follows: The algorithm starts with five years (1999-2003) worth of weekly data and runs the iHMM with our five macroeconomic variables as inputs for initial model training. The iHMM predicts all the regimes for every week of data input based on the macroeconomic variables. For each week of regime prediction, the portfolio rebalances if the predicted regime is different from the prior regime. The model segments and takes all the available stocks when rebalancing occurs up until that point and only keeps the data in which the historical regimes are the same as the current regime, building a temporary dataset. The model runs a Fama-MacBeth regression on the factors and returns to give them a value based on their effect on the returns. The factors are ranked, with the highest values receiving the highest scores. For example, in the baseline model, we have five stock factors ranked through Fama-Macbeth regression from highest to lowest: [5,4,3,2,1] and divided by $5+4+3+2+1 = 15$ in order to yield a factor weight for a given time point and is used to calculate a composite score for each stock. The model then calculates a composite score for all the stocks based on the latest 12 weeks of data available. The top 50 stocks based on the composited score are selected in the portfolio and weighted equally. EURO STOXX 50 is the market index for our comparison and compromises of 50 stocks, hence the same number of stocks in our portfolio to enable comparability. Finally, the algorithm iterates the process described so far by adding one week of data until the model predicts a regime change, followed by rebalancing the portfolio before iterating again. We apply a 3% trading cost each time the portfolio is rebalanced. See Figure 6 for an illustration of the algorithm.

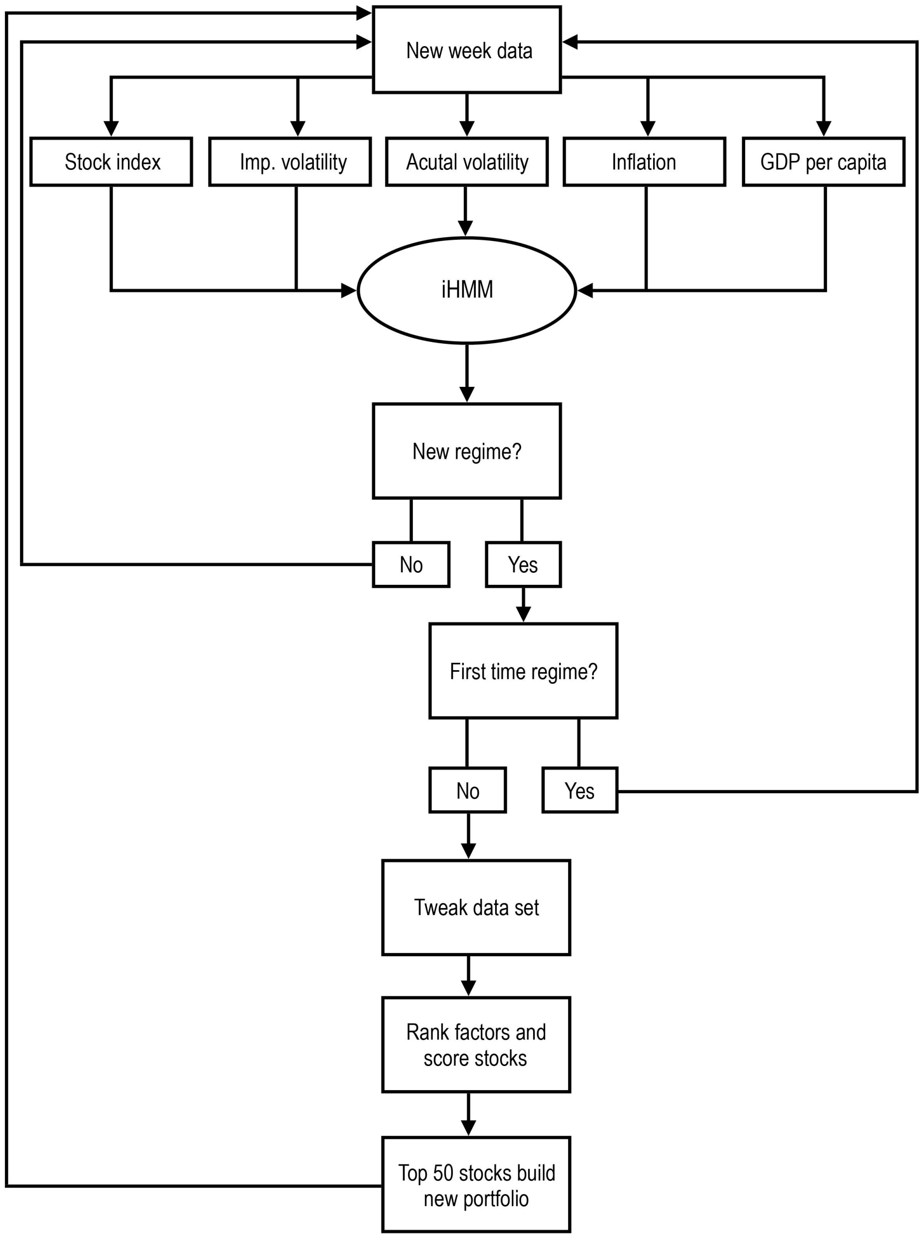

Figure 6: Illustration of the algorithm

5.2 Z-scores

First, we optimize our large amount of raw data before using it in our model. We optimize by using a methodology in which we transform our six standard factors and four ESG factors. We follow Asness et al. (2019) framework and standardize factor variables and group them into a Z-score, resulting in one score for each of our factors.

In accordance with Frey (2018), the Z-score calculation is as follows:

$$Z_N^i = \frac{V_N^i - \mu_i}{\sigma_i} \tag{20}$$

where Z_N^i is the standardized score for factor i of stock N, V_N^i is the factor value for for factor i of stock N, μ_i is the cross-sectional mean, and σ_i is the standard deviation of factor i. The factor score equals:

$$Score_N^i = \frac{Z_N^1 + Z_N^2, ..., Z_N^i}{n} \tag{21}$$

where n is the number of factors. Each Z-score is calculated based on each week in the data to avoid future "unobserved" data affecting observed data. We build the portfolios based on the Z-scores and historical performance in the different regimes.

5.3 Infinite Hidden Markov Model iterations

After optimizing our data, we feed it to the iHMM, starting with the five macroeconomic variables using weekly data input from 1999 to 2003. Every weekly datapoint will generate one regime based on the five macroeconomic variables. Predicting a new regime compared to the previous regime results in a rebalancing of the portfolio and potentially changes the stock composition of the portfolio based on the scores. A continuation of a regime results in no portfolio adjustment, and another week of data is input as the model loops. We utilize an expanding window, meaning that we continuously add new data to the model while keeping the older data. This process allows the iHMM to get more robust predictions as more macroeconomic data is processed while the factor data stays relevant as we cannot tell how far back the data for the same regime will be. Figure 7 displays an illustration of the expanding window.

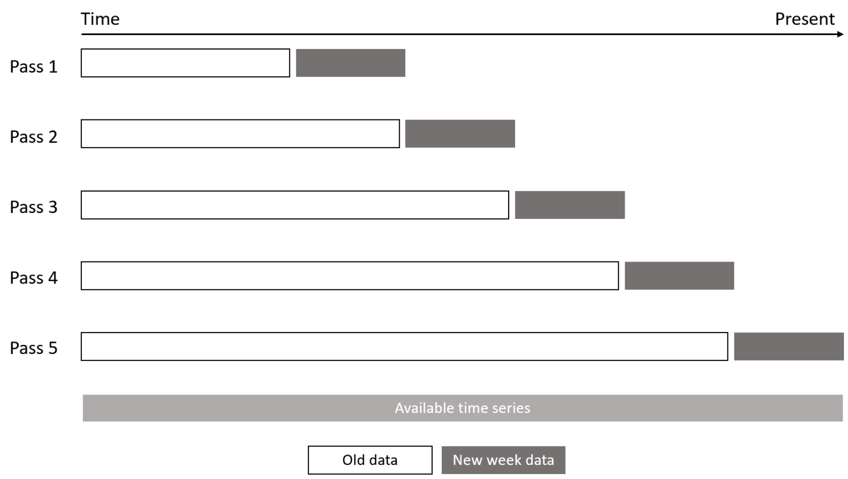

Figure 7: Illustration of an expanding window

We utilize the iHMM's ability to decide the optimal number of regimes based on data rather than predetermining them in contrast to previous research using an HMM. The result is that each iteration can have a different number of regimes. For example, iteration n_0 may have three different regimes [1,2,3,1,3], and n_1 can have five [5,4,3,2,1,3]. It is a result of the expanding window in which newly available macroeconomic data continuously trains the model, and future iterations may alter the previously defined regimes. We do not rebalance the portfolio for changes in the previously defined regime. In the mentioned example, the portfolio stays unchanged as the latest regime in n_0 (3) is the same as the predicted regime in iteration n_1 (3), even though the previously defined regimes are changed. We also have one security feature built into the model. If the model detects a new regime, but that regime has not yet been found historically, the model will not rebalance the portfolio. Another week of data is added instead, and the model is rerun. This process eliminates noise in the model, and because once the model scores the stocks, it would only have one week of data available, i.e., the latest week added, and may result in poor allocations due to the lack of data.

5.4 Factor scoring and portfolio building

Once a new regime is predicted, the model temporarily strips the dataset of all stock and factor data for periods that do not match the regime. The model runs a Fama-MacBeth regression on the temporary dataset. Each week yields a regression output only used for scoring and is thus not interesting to analyze from a statistically significant point of view since all factors are utilized in the model regardless. However, the signs are of interest as they are used to rank the values from smallest to largest. To account for the possibility of negative values in the Fama-MacBeth regression, we allow weights to be negative, showcasing that certain factors have a negative impact on returns. The values are divided by the total sum, resulting in weights that add up to 100%. Similarly to Nguyen & Nguyen (2021), we keep the 12 most recent week's data (one quarter) to avoid look-back bias and avoid having the same stock selection continuously in the portfolio. This step addresses the risk of having stocks selected in the coming portfolio constructions due to receiving an exceedingly high score early in the model, even if they would be deemed a poor pick in later stages. The result is a panel data set with stocks, their respective factor scores, and the calculated factor weight. Next, each factor score is multiplied by the corresponding factor weight, and each stock receives a composite score. The model picks the top 50 stocks with the highest composite scores based on the results and adjusts the portfolio. Since new data is added by the end of each trading week, the rebalancing of the portfolios occurs at the beginning of the following trading week. Each stock is equally weighted independent of the composite scores. Table 5 presents the portfolio models.

Table 5: Portfolio models

Model	Model breakdown
Base Model:	Momentum, Risk, Value, Quality and Growth factors
ESG Model:	Base + ESG, Environmental, Social and Governance factors
E,S,G Model:	Base + Environmental, Social and governance factors
ESG, no E,S,G Model:	Base + ESG factor
Environmental Model:	Base + Environmental factor
Social Model:	Base + Social factor
Governance Model:	Base + Governance factor

By comparing the five different ESG models with the Base model, we observe if different ESG factor compositions can enhance the portfolio performance and if the ESG factors add value to the portfolio building framework.

5.5 Portfolio performance evaluation metrics

We select the following seven metrics to compare and evaluate the portfolio's performances:

- **Compounded annual growth rate (CAGR):** CAGR is the rate of return required for an investment to increase from its initial balance to its final balance, providing profits are reinvested at the conclusion of each period of the investment's life span. The formula for CAGR is as follows:

$$CAGR = \left(\frac{\text{Ending value}}{\text{Start value}}\right)^{\frac{1}{n}} - 1 \tag{22}$$

- **Annual Volatility** The pace at which a stock's price rises and falls over a given period is referred to as volatility. It is the same thing as the weekly returns' standard deviation. The riskier an investment is, the higher the stock volatility. Volatility is often measured in years. The formula for Annual volatility is as follows:

$$\sigma_A = \text{Weekly standard deviation} * \sqrt{52} \tag{23}$$

- **Sharpe Ratio** The Sharpe ratio measures the return on investment per unit of risk. A greater Sharpe ratio indicates superior risk-to-reward performance. We divide the annualized return by the annualized standard deviation to get the annualized Sharpe ratio:

$$\text{SR} = \frac{R_p - Rf}{\sigma_A} \tag{24}$$

- **Information Ratio** Similar to the Sharpe Ratio, the Information ratio is an indicator of risk-adjusted returns. However, the information ratio uses a benchmark return such as an index instead of a risk-free return (in this case, the EURO STOXX 50 index). A higher Information Ratio implies a better portfolio with a higher return in excess of the benchmark, given the risk taken.

$$\text{IR} = \frac{\text{Portfolio return} - \text{Benchmark return}}{\sigma(\text{Portfolio return} - \text{Benchmark return})} \tag{25}$$

- **T-test** For the t-test, we use a Dependent t-test for paired samples. This test is utilized when the samples are dependent, that is, when only one sample has been examined twice (repeated measurements) or when two samples have been matched or paired. We test the null hypothesis that the two samples, the ESG factor portfolio models and the Base portfolio model, have identical average (expected) values. The formula for the t-test is:

$$t = \frac{\bar{X}_D - \mu_0}{S_D/n} \tag{26}$$

- **Average ESG score** We look at the average ESG score for the portfolios for the whole period to examine if adding ESG factors will result in a portfolio with higher ESG scores. The formula for the average ESG score is:

$$\mu_{ESG} = \frac{ESG_1 + ESG_2 + ... + ESGn}{n} \tag{27}$$

- **Multi-factor Alpha** Following Bruno et al. (2021), we use a multi-factor model to analyze portfolio alphas. A positive and significant alpha implies improved risk-adjusted returns. that the portfolios enable investors to improve their risk-adjusted returns. We deploy a standard time series regression to measure the level of returns of the ESG models that do not come from exposure to traditional factors. The multi-factor model includes seven factors: the market (MRK), value (HML), size (SMB), robust (RMW), conservative (CMA), and momentum (WML) factors. We estimate alpha in time-series regressions of weekly ESG portfolio returns (in excess of the risk-free rate) on factor returns:

$$
\begin{aligned}
r_{S,t} - r_{f,t} = \alpha_s + \beta_{MKT,S}(r_{M,t} - r_{f,t}) + \beta_{SMB,S}(SMB_t) + \beta_{HML,S}(HML_t) \\
+ \beta_{RMW,S}(RWM_t) + \beta_{CMA,S}(CMA_t) + \beta_{WML,S}(WML_t) + \varepsilon_{S,t} \quad (28)
\end{aligned}
$$

The six factor's data is obtained from French, Kenneth R (2022).

6 Results and analysis

This section presents the results and analysis of our study using two periods, one for the full range of data (2003-2022) and one using recent years (2019-2022) to discern any differences in more recent years given the increase in ESG awareness. We need historical stock data to create the composite scores, thus the first year of data (2002) is not used for portfolio construction. First, we present the regimes iterated through the iHMM. Next, the Fama McBeth regression outputs are presented, which describe the weight of the returns allocated to each factor in the portfolios. Finally, the performance of our constructed portfolios is presented based on our metrics as described in the methodology.

6.1 Final iHMM iteration results

We run the iHMM weekly and experience regime switches frequently. On average, the model results in 14-15 regimes per iteration. Table 6 presents the final iteration and the regime breakdown. Regimes 0, 2, and 6 are the most typical regimes, while 5, 11, and 14 are the most uncommon. Furthermore, we can see the different average levels of growth for each of our macroeconomic variables in the given regimes. Figure A.1 in appendix A illustrates the regime switches and the stock index performance.

Table 6: Final iHMM iretation results

Regime	Count	Stock index	Implied volatility	Actual volatility	Inflation	GDP/Cap
0	205	0.11%	-0.67%	1.23%	0.17%	-5.16%
1	50	2.77%	-10.83%	4.11%	0.23%	-6.82%
2	338	0.41%	-0.70%	0.75%	0.21%	4.60%
3	32	-4.80%	29.96%	3.18%	0.30%	-2.01%
4	23	-3.23%	19.83%	5.26%	-0.10%	-2.65%
5	1	-12.33%	168.11%	8.40%	0.20%	-8.33%
6	240	0.16%	-0.31%	1.53%	0.06%	-0.27%
7	118	2.06%	-11.00%	2.99%	0.22%	4.57%
8	20	-3.96%	7.42%	3.40%	0.22%	-6.72%
9	8	6.95%	-16.42%	9.78%	1.19%	3.36%
1	23	3.55%	-6.10%	6.39%	0.05%	1.10%
1	6	-1.48%	23.13%	1.29%	-0.32%	-1.91%
1	11	0.15%	-5.44%	1.57%	-0.24%	10.95%
1	132	-2.24%	12.49%	2.54%	0.21%	4.52%
1	4	-10.70%	51.45%	10.89%	0.00%	0.28%

6.2 Average Factor score weights

Table 7 presents the average allocated factor weights through the Fama-Macbeth regressions for the entire duration of our study (2003-2022). The top row displays the different constructed portfolios, and the leftmost column displays the factors. We can see that Momentum and Growth are allocated the largest weight on average in all portfolios, while the Risk and Quality factors have the lowest overall.

The allocated ESG weights are relatively low compared with the traditional factors.

Table 7: 2003-2022 Average Factor score weights

2003-2022	Base	ESG	ESG, no E,S,G	E,S,G	Environmental	Social	Governance
Momentum	66.67%	80.00%	42.86%	44.44%	42.86%	42.86%	42.86%
Risk	-66.67%	-40.00%	-28.57%	-22.22%	-28.57%	-28.57%	-28.57%
Value	33.33%	60.00%	28.57%	33.33%	28.57%	28.57%	28.57%
Quality	-33.33%	-20.00%	-14.29%	-11.11%	-14.29%	-14.29%	-14.29%
Growth	100.00%	100.00%	57.14%	55.56%	57.14%	57.14%	57.14%
ESG	-	20.00%	14.29%	-	-	-	-
Environmental	-	-60.00%	-	-33.33%	14.29%	-	-
Social	-	-80.00%	-	11.11%	-	14.29%	-
Governance	-	40.00%	-	22.22%	-	-	14.29%

Table 8 displays the average allocated factor weights from the Fama-MacBeth regression for 2019-2022. Momentum and Growth still have the largest weights on average, except for Growth in the "ESG" model, in which the ESG pillar factors have the largest weights. Similarly, Risk and Quality are on the lower end. The results are generally unchanged for the ESG factor's weights in the other models compared to the entire period results, except for the Environmental pillar in the "E,S,G" portfolio, which changes sign and weight with the Social pillar. Initially, it would seem that ESG scores are more prominent in recent years for portfolio construction, given their weights in the Fama-MacBeth regression. The following section will present the model's performance in detail.

Table 8: 2019-2022 Average Factor score weights

2019-2022	Base	ESG	ESG, no E,S,G	E,S,G	Environmental	Social	Governance
Momentum	66.67%	-60.00%	42.86%	44.44%	42.86%	42.86%	42.86%
Risk	-33.33%	20.00%	-14.29%	-11.11%	-14.29%	-14.29%	-14.29%
Value	33.33%	-40.00%	28.57%	33.33%	28.57%	28.57%	28.57%
Quality	-66.67%	40.00%	-28.57%	-22.22%	-28.57%	-28.57%	-28.57%
Growth	100.00%	-80.00%	57.14%	55.56%	57.14%	57.14%	57.14%
ESG	-	-20.00%	14.29%	-	-	-	-
Environmental	-	80.00%	-	11.11%	14.29%	-	-
Social	-	60.00%	-	-33.33%	-	14.29%	-
Governance	-	100.00%	-	22.22%	-	-	14.29%

6.3 Portfolio Performance

6.3.1 Full period performance (2003-2022)

Figure 8 illustrates the different portfolio's performances for 2003-2022. We note that all the constructed portfolios outperform the market index EURO STOXX 50. The Base portfolio outperforms ESG factor portfolios in cumulative returns. The ESG factor portfolios also show discrepancies, notably in the "E,S,G" and "Environmental" portfolio models, in which we can observe relatively poor performance.

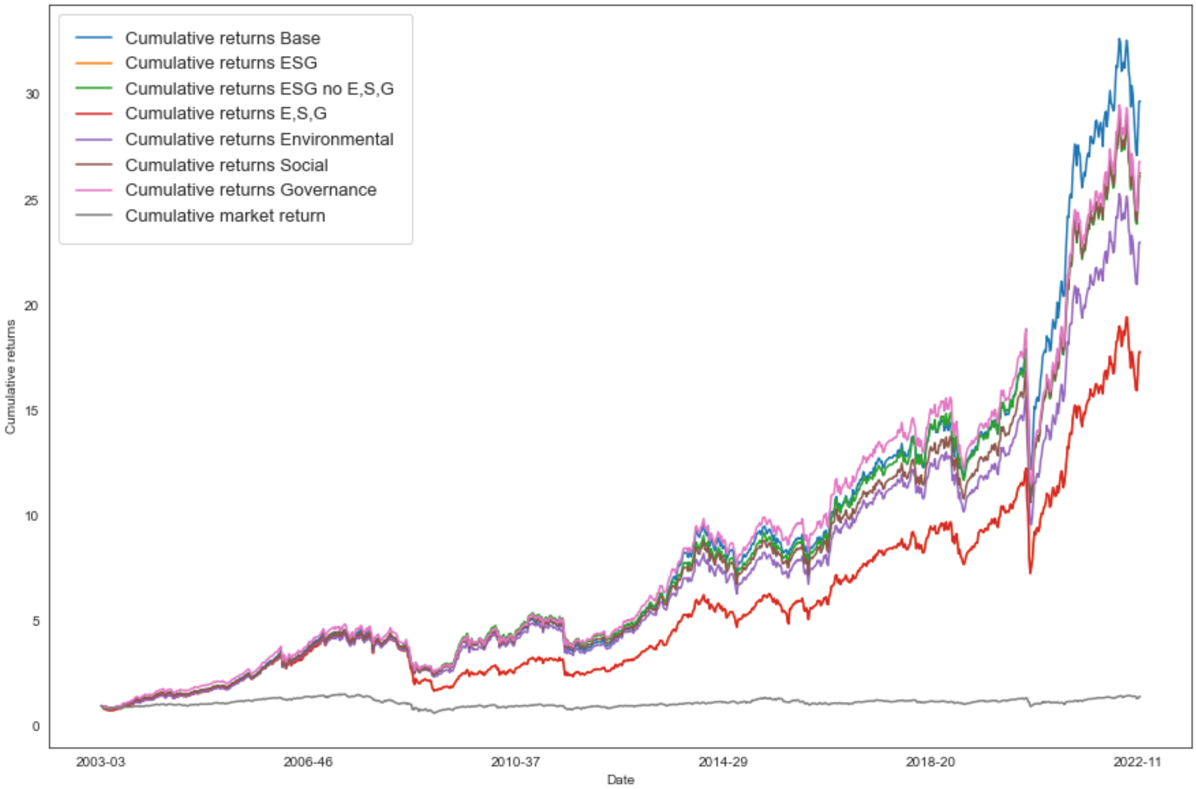

Figure 8: 2003-2022 Portfolio performance

Table 9 presents the results in detail and outlines the differences. The Base portfolio model has the highest CAGR of 19.09%, and the closest performing portfolio is "Governance" with an 18.50% CAGR. The volatilities of the portfolios are similar, ranging from "Environmental" 19.79%, which is lower than the Base portfolio's volatility of 20.36%, to "E,S,G" 21.81%. The Base portfolio also displays the best Sharpe ratio of 0.91 and the highest Information Ratio of 0.77. However, none of the ESG factor portfolio return's t-tests against the Base portfolios' return are statistically significant, i.e., we cannot reject the null hypothesis that the two samples have identical average (expected) values.

Table 9: 2003-2022 Portfolio Performance

2003-2022	CAGR	Volatility	SR	IR	t-test P-value	α	α P-value	ESG score
Base	19.09%	20.36%	0.93	0.77	-	0.32%	0.000	43.67
ESG	16.09%	21.81%	0.73	0.63	0.43	0.26%	0.012	43.55
ESG, no E,S,G	18.33%	20.19%	0.90	0.75	0.53	0.31%	0.001	43.73
E,S,G	16.09%	21.81%	0.73	0.63	0.43	0.26%	0.012	43.55
Environmental	17.55%	19.79%	0.88	0.72	0.15	0.29%	0.001	44.41
Social	18.36%	20.07%	0.91	0.75	0.51	0.30%	0.001	43.75
Governance	18.50%	20.14%	0.91	0.76	0.61	0.30%	0.001	43.72
Market	2.38%	21.10%	0.11	-	-	-	-	-

We observe positive and statistically significant alphas, implying that the returns in the ESG factor portfolios not arising from the traditional factors are significant. The alphas, however, are low in terms of raw percentage, ranging from 0.26% in "E,S,G" to 0.31% in "ESG, no E,S,G". Notably, we can observe positive and significant alphas in the Base portfolio model too with an alpha of 0.32%. Following these results, we consider that the alphas in the ESG factor portfolio models may not be due to ESG information as the Base portfolio model with the highest performance also generates similar alpha. Thus, the returns may instead be a result of the regime-switching iHMM.

Finally, we observe that the average ESG scores for the portfolios are similar without a substantial deviation to the Base portfolio model. Stocks with higher ESG scores do not seem to be allocated more to the portfolios on average. Furthermore, the "ESG" and "E,S,G" portfolios results are identical, i.e., the algorithm has allocated the same stocks to both models. A possible explanation is a high correlation in the two sets of ESG data, where ESG is a compounded score of the E, S, and G pillars, meaning that models do not benefit from including both sets of factors simultaneously.

6.3.2 Recent years performance (2019-2022)

Figure 9 displays the performance of our constructed portfolios for the period 2019-2022. All portfolios beat the market index EURO STOXX 50, similar to the results for the full period, where the Base portfolio outperforms the other portfolios. Table 10 presents the results in detail.

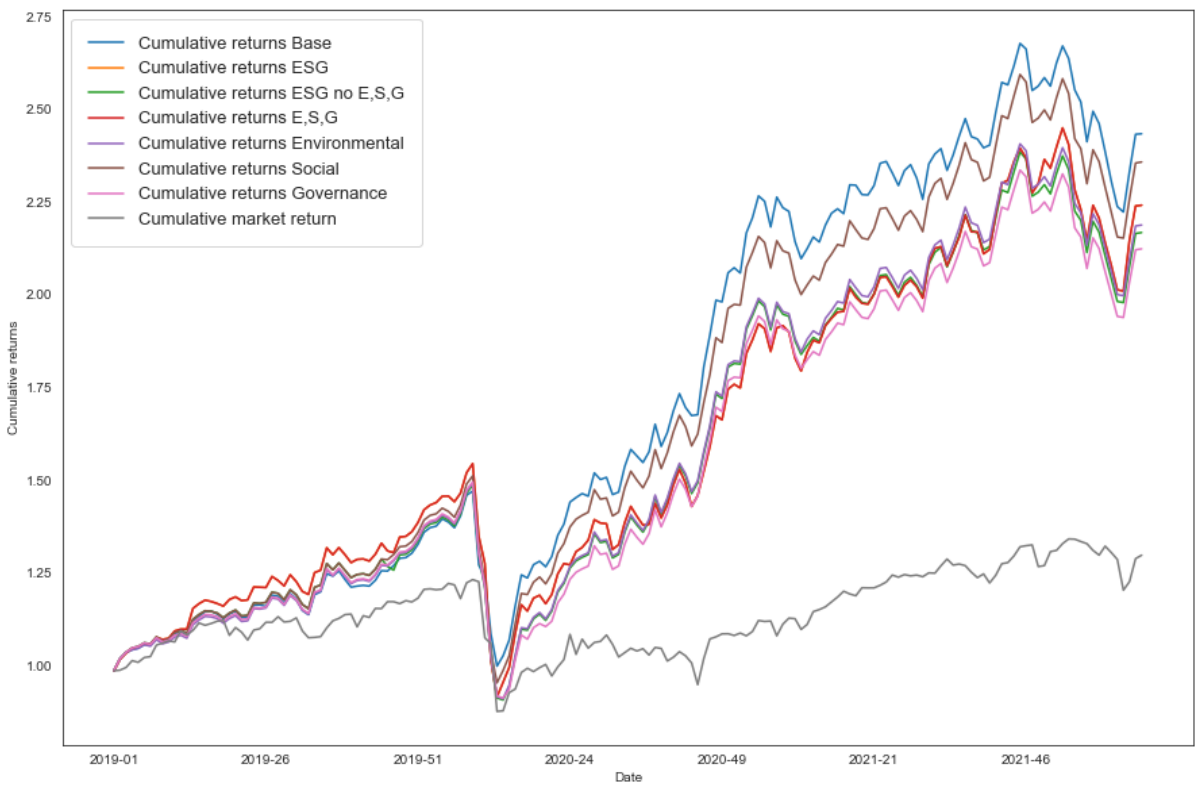

Figure 9: 2019-2022 Portfolio performance

Table 10: 2019-2022 Portfolio Performance

2019-2022	CAGR	Volatility	SR	IR	t-test P-value	α	α P-value	ESG score
Base	30.92%	20.28%	1.52	0.84	-	0.46%	0.027	45.25
ESG	28.01%	23.30%	1.20	0.82	0.91	0.64%	0.013	44.98
ESG, no E,S,G	26.39%	21.60%	1.22	0.68	0.26	0.43%	0.052	45.03
E,S,G	28.01%	23.30%	1.20	0.82	0.93	0.65%	0.012	44.98
Environmental	26.75%	21.53%	1.24	0.70	0.28	0.43%	0.049	45.09
Social	29.68%	21.23%	1.39	0.79	0.76	0.45%	0.037	45.02
Governance	25.60%	21.40%	1.19	0.66	0.16	0.42%	0.053	45.12
Market	8.66%	23.56%	0.36	-	-	-	-	-

The Base portfolio model has the highest CAGR of 30.92%, followed by the "Social" portfolio with 29.68%. Interestingly, the worst-performing portfolio is the "Governance" with a CAGR of 25.60%, whereas it has the second-best performance in the full period. The Base model portfolio displays lower volatility of 20.28% compared with the benchmark market with a volatility of 23.56%, and ESG factor portfolios have volatilities in between.

In the other performance metrics, we can observe that all portfolios display a Sharpe ratio excess of 1 compared with the full period in which all portfolios have a Sharpe ratio less than 1. The Base portfolio model has the highest Sharpe ratio and Information ratio of 1.52 and 0.84, respectively. We note that none of the ESG portfolio's returns t-tests against the base model returns are significant, i.e., we cannot reject the null hypothesis that the two samples have identical average (expected) values. As for average ESG scores for the portfolios, the portfolios have a similar score for the period 2019-2022 as 2003-2022 despite the higher average ESG scores in our dataset (see Figure 5). The average ESG score of the portfolio is approximately 45 and is significantly lower than the average of the dataset for the same period, which is slightly above 50.

All portfolios display positive and significant alphas at a 10% significance level or lower. Accordingly, this implies that the returns in the portfolios not coming from the traditional factors are significant. Based on the same reasoning for the alphas as in the full period results, we do not conclude that the ESG scores' informational content explains the alpha generated.

6.4 Robustness tests

We perform additional tests to analyze the robustness of our results. First, we have discussed reasons to exclude size as a factor in our risk scores. We run a model including a Size factor score to analyze if model performance improves and ensure that our model does not exclude vital information. Unlike the other factors, we do not imply if larger or smaller companies shall receive a better score. By having a factor score built on a single factor (Size), the model itself will determine the factor effect. This methodology is used solely for the purpose of robustness checking. Investors typically follow an investment thesis and would select if a smaller or larger company would receive a higher weight in the Score factor. Second, we have highlighted the issue of selecting macroeconomic variables to represent Western Europe. For example, the United Kingdom is no longer part of the EU post-Brexit, hence not included in all GDP calculations, but UK data is still a large part of our dataset presented in Section 4, Data. Therefore, we run additional tests using only UK equities and macroeconomic variables for the UK. Finally, we run a final analysis including only Sweden as they, in general, are perceived as front-runners in terms of ESG (Morningstar 2021). Note that Sweden keeps more companies post-screening in Table 1 relative to the other major economies (a 39% reduction compared with the 55% average). These robustness checks will be analyzed using full-period data. The tables of results for the robustness tests can be found in section A.

Size The inclusion of Size as a factor results in all portfolios having worse performance in the selected metrics (see Appendix A, Table A.4 and Table A.5). Furthermore, the allocated weights from the Fama-MacBeth regressions are on the lower end for all portfolios. Our results imply that other factors are more important in our model, whereas Size added noise.

United Kingdom We tailor the macroeconomic factors for running a UK model. We use FTSE 100 for the stock index, FTSE Implied Volatility Index Series, inflation, and GDP per Capita. The results are presented in (Appendix A, table A.6). Most portfolios outperform the baseline model except for the "ESG" and "E,S,G" portfolio models. Overall, the portfolios with only UK stocks do not outperform the original portfolios, implying that the effect of including the UK is not dominant in our results.

Sweden We tailor the data for Sweden similar to the UK tests. We use the macroeconomic variables OMX30 as the stock index, inflation, and GDP per Capital for Sweden. There is no implied volatility index for Sweden, thus, the VSTOXX 50 is selected as we use it in the original models. The results are presented in appendix A, table A.7. The results align with the original dataset in which the base portfolio model has the best performance. Identical to the UK tests, using a Swedish dataset with Swedish equities yields lower performance than the original model for all portfolios.

7 Discussion

Our findings show that the iHMM portfolios consistently outperform the market index EURO STOXX 50, both for the full period portfolios (2003-2022) and the more recent period portfolios (2019-2022). The results verify that the iHMM is capable of satisfactory financial performance, similar to the work on HMM in stock selection, e.g., Nguyen & Nguyen (2021) and Wang et al. (2020). In terms of ESG factor performance, we conclude that the regime-switching iHMM does not utilize ESG factors in a way that generates ESG-specific alphas. To analyze the portfolios, we first consider the full period. We observe that the models using ESG factors fail to outperform the base model constructed using the traditional factors in the performance metrics, albeit insignificant t-tests for all ESG factor portfolios. Giese et al. (2020) finds evidence that the environmental and social pillars have a significant positive impact on companies' stock performance in the long run and governance in the short run. Based on our main results, we do not find the suggested effect to be captured in ESG scores' informational content in addition to what may already be explained by traditional factors. However, we observe slightly different results in the robustness tests, notably for the UK. The respective pillar model portfolios outperform the base model portfolio, suggesting that country-specific effects may have an impact. Still, we do not draw any conclusions based on the UK results considering the t-tests' insignificance and the small alphas. The results could be of interest to more niche investors in specific markets and researched further.

The results for the more recent period do not differ from the full period results, only in magnitude. We constructed two sets of results based on an increase in ESG activities and the positive movement in recent years, which could affect high-ESG score companies' stock prices. We do not document such effects. A possible explanation is that heavy recent ESG investments have a drawback on financial performance in the short run given the costs of implementing company ESG measures. Such explanation would have additional weight if, in contrast, the full period portfolios generated meaningful alphas. It is, therefore, more reasonable to conclude that ESG scores' informational content is already captured by traditional factors, and ESG scores do not provide additional meaningful effects that improve the portfolio models' financial performance. Indeed, Melas et al. (2017) document a positive correlation between ESG scores and traditional factors, and the implications that ESG strategies tilt towards risk factor style investments could explain our findings.

We use Bruno et al. (2021) approach to risk-adjust the results and analyze the alphas. We observe statistically significant alphas for all portfolios in contrast to Bruno et al. (2021) results, albeit low alphas. We theorize that the iHMM generates the alphas and not the integration of ESG scores, aligning our findings with Bruno et al. (2021). Furthermore, the results suggest that the additional information provided in our model through the ESG dataset is not significant, i.e., the ESG factors do not explain additional returns. These findings also align with Halbritter & Dorfleitner (2015), who fails to observe a relationship between ESG ratings and returns, and we conclude that ESG integration is not a source of outperformance. Neither does the inclusion of ESG scores seem to provide lower volatility as a "safer bet" per Ma (2019) and Melas et al. (2017), who find that ESG factors have a positive relationship with low volatility. In contrast, the "ESG" and "E,S,G" models have the highest volatilities for both periods with a lower CAGR.

8 Conclusion

This thesis investigates whether ESG data affects portfolios' risk-return characteristics in an infinite Hidden Markov Model. We analyze ESG scores' informational content by developing a baseline portfolio model using traditional factors (Value, Quality, Growth, Momentum, and Risk), then expand the model by adding layers of ESG factors. This method allows us to observe whether ESG scores capture additional effects that traditional factors do not. The study uses equity data for Western European stocks for the period 2003-2022, and the results are divided into a full period study and a study using more recent year's data (2019-2022) to investigate if the ESG phenomenon changes the results.

Firstly, consistent with previous research on stock selection using regime-switching models, we find that the iHMM is capable of satisfactory financial performance when compared to EURO STOXX 50 as a benchmark. This study has contributed to the existing literature by expanding on the HMM into the iHMM for stock selection and factor investing. With the use of an iHMM, the regime specification becomes data-driven, overcoming previous challenges of proper specification, which has used inefficient methods or greedy approaches.

Secondly, we investigate the performance metrics of constructed portfolios using different layers of ESG factor data compared with the baseline model. The Base portfolio model outperforms all portfolios constructed with ESG factors in terms of CAGR with lower volatility than almost all portfolios. The ESG portfolios' returns t-tests against the Base portfolio's return are not significant, i.e., we cannot reject the null hypothesis that the two samples have identical average (expected) values.

Lastly, we risk-adjust our results and analyze the alpha. We observe statistically significant alphas, albeit low. The Base portfolio model display similar alpha as the other ESG models, thus, we conclude that ESG scores are not the source and contribute the alpha to the iHMM. Furthermore, we conclude that ESG scores do not capture effects that improve portfolios' risk-return characteristics that traditional factors do not.

The main implication of our study is that in an iHMM regime-switching framework, investors in the Western European market should not utilize ESG scores to attempt to improve the risk-return characteristics of their portfolios. Rather, investors may want to consider other objectives, for example, aligning their portfolios with their values and norms or as a means to impact investing. Furthermore, investors cannot imply that they have "ESG portfolios" simply by integrating ESG factors in their models. Evidently, doing so did not significantly impact our portfolio's average ESG score, which hardly increased compared to the baseline model.

8.1 Further research

We note that our results rely on the ESG data provider Refinitiv Eikon for this thesis. The methodology for the ESG scoring differs among the providers, and Billio et al. (2021) provides evidence of heterogeneity in rating criteria. We suggest recreating this research using another prominent ESG dataset, such as the ESG dataset provided by MSCI, to find whether the results are robust. Additionally, further research may analyze the ESG scores in-depth, using data that constructs the compounded score, such as emission data. It may neutralize bias in rating providers' methodologies and value gains possibly found in respective company metrics. Furthermore, based on the results of the robustness tests, notably for the UK, our model can be applied to other geographical markets to verify further the robustness of the results and possibly capture country-specific effects.

We are entering unprecedented times where climate risks are increasingly taking physical form as drought, floods, and other natural disasters. It may have an impact on ESG investing, where it becomes more material as regulatory interventions increase. Companies well-positioned may face a lower risk of being affected by regulations and disruptions. Consequently, we cannot say that our results have timelessness, as noted by Amel-Zadeh & Serafeim (2018). Future research may want to replicate this study if the conditions have drastically changed and expect different results.

9 References

Alessandrini, F. & Jondeau, E. (2020), 'Esg investing: From sin stocks to smart beta', *The Journal of Portfolio Management* **46**(3), 75–94.

Amel-Zadeh, A. & Serafeim, G. (2018), 'Why and how investors use esg information: Evidence from a global survey', *Financial Analysts Journal* **74**(3), 87–103.

Amenc, N., Goltz, F., Sivasubramanian, S. & Lodh, A. (2015), 'Robustness of smart beta strategies', *The Journal of Index Investing* **6**(1), 17–38.

Asness, C. S., Frazzini, A. & Pedersen, L. H. (2019), 'Quality minus junk', *Review of Accounting Studies* **24**(1), 34–112.

Bae, G. I., Kim, W. C. & Mulvey, J. M. (2014), 'Dynamic asset allocation for varied financial markets under regime switching framework', *European Journal of Operational Research* **234**(2), 450–458.

Baum, L. E. & Petrie, T. (1966), 'Statistical inference for probabilistic functions of finite state markov chains', *The annals of mathematical statistics* **37**(6), 1554–1563.

Beal, M. J., Ghahramani, Z. & Rasmussen, C. E. (2002), 'The infinite hidden markov model', *Advances in neural information processing systems* **1**, 577–584.

Bender, J., Briand, R., Melas, D. & Subramanian, R. A. (2013), 'Foundations of factor investing', *Available at SSRN 2543990* .

Berg, F., Fabisik, K. & Sautner, Z. (2020), 'Rewriting history ii: The (un) predictable past of esg ratings', *European Corporate Governance Institute–Finance Working Paper* **708**(2020), 10–2139.

Billio, M., Costola, M., Hristova, I., Latino, C. & Pelizzon, L. (2021), 'Inside the esg ratings:(dis) agreement and performance', *Corporate Social Responsibility and Environmental Management* **28**(5), 1426–1445.

Blitz, D. & Swinkels, L. (2008), 'Fundamental indexation: An active value strategy in disguise', *Journal of Asset Management* **9**(4), 264–269.

Blunsom, P. (2004), 'Hidden markov models', *Lecture notes, August* **15**(18-19), 48.

Boffo, R. & Patalano, R. (2020), 'Esg investing: Practices, progress and challenges', *Éditions OCDE, Paris* .

Bruder, B., Cheikh, Y., Deixonne, F. & Zheng, B. (2019), 'Integration of esg in asset allocation', *Available at SSRN 3473874* .

Bruno, G., Esakia, M. & Goltz, F. (2021), '"honey, i shrunk the esg alpha": Risk-adjusting esg portfolio returns', *The Journal of Investing* .

Carhart, M. M. (1997), 'On persistence in mutual fund performance', *The Journal of finance* **52**(1), 57–82.

Chen, M. & Mussalli, G. (2020), 'An integrated approach to quantitative esg investing', *The Journal of Portfolio Management* **46**(3), 65–74.

Cheng, B., Ioannou, I. & Serafeim, G. (2014), 'Corporate social responsibility and access to finance', *Strategic management journal* **35**(1), 1–23.

Dhaliwal, D. S., Li, O. Z., Tsang, A. & Yang, Y. G. (2011), 'Voluntary nonfinancial disclosure and the cost of equity capital: The initiation of corporate social responsibility reporting', *The accounting review* **86**(1), 59–100.

Drei, A., Le Guenedal, T., Lepetit, F., Mortier, V., Roncalli, T. & Sekine, T. (2019), 'Esg investing in recent years: New insights from old challenges', *Available at SSRN 3683469* .

Eltogby, M., Brown, K. & Corrigan, G. (2019), Seeking return on esg: Advancing the reporting ecosystem to unlock impact for business and society, *in* 'White Paper. World Economic Forum'.

Fama, E. F. & French, K. R. (1992), 'The cross-section of expected stock returns', *the Journal of Finance* **47**(2), 427–465.

Fama, E. F. & MacBeth, J. D. (1973), 'Risk, return, and equilibrium: Empirical tests', *Journal of political economy* **81**(3), 607–636.

Fons, E., Dawson, P., Yau, J., Zeng, X.-j. & Keane, J. (2021), 'A novel dynamic asset allocation system using feature saliency hidden markov models for smart beta investing', *Expert Systems with Applications* **163**, 113720.

French, Kenneth R (2022), 'Kenneth r. french data library', http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. [Online; accessed 03-May-2022].

Frey, B. B. (2018), *The SAGE encyclopedia of educational research, measurement, and evaluation*, Sage Publications.

Giese, G., Lee, L.-E., Melas, D., Nagy, Z. & Nishikawa, L. (2019), 'Foundations of esg investing: How esg affects equity valuation, risk, and performance', *The Journal of Portfolio Management* **45**(5), 69–83.

Giese, G., Nagy, Z. & Lee, L.-E. (2020), 'Deconstructing esg ratings performance', *Journal of Portfolio Management* .

Guidolin, M. & Timmermann, A. (2008), 'Size and value anomalies under regime shifts', *Journal of Financial Econometrics* **6**(1), 1–48.

Halbritter, G. & Dorfleitner, G. (2015), 'The wages of social responsibility—where are they? a critical review of esg investing', *Review of Financial Economics* **26**, 25–35.

Hamilton, J. D. (1989), 'A new approach to the economic analysis of nonstationary time series and the business cycle', *Econometrica: Journal of the econometric society* pp. 357–384.

Hardy, M. R. (2001), 'A regime-switching model of long-term stock returns', *North American Actuarial Journal* **5**(2), 41–53.

Kaiser, L. (2020), 'Esg integration: value, growth and momentum', *Journal of Asset Management* **21**(1), 32–51.

Kim, E.-c., Jeong, H.-w. & Lee, N.-y. (2019), 'Global asset allocation strategy using a hidden markov model', *Journal of Risk and Financial Management* **12**(4), 168.

Kole, E. & Van Dijk, D. (2017), 'How to identify and forecast bull and bear markets?', *Journal of Applied Econometrics* **32**(1), 120–139.

la Torre-Torres, D., Oscar, V., Galeana-Figueroa, E. & Álvarez-García, J. (2019), 'Efficiency of the public pensions funds on the socially responsible equities of mexico', *Sustainability* **11**(1), 178.

Lintner, J. (1965), 'Security prices, risk, and maximal gains from diversification', *The journal of finance* **20**(4), 587–615.

Liu, P. P., Xu, K. & Zhao, Y. (2011), 'Market regimes, sectorial investments, and time-varying risk premiums', *International Journal of Managerial Finance* .

Ma, Y., MacLean, L., Xu, K. & Zhao, Y. (2011), 'A portfolio optimization model with regime-switching risk factors for sector exchange traded funds', *Pac J Optim* **7**(2), 281–296.

Ma, Z. (2019), 'Three essays on asset pricing in regime and esg environments'.

Melas, D., Nagy, Z. & Kulkarni, P. (2017), Factor investing and esg integration, *in* 'Factor Investing', Elsevier, pp. 389–413.

Meziani, S. (2014), 'Smart beta etfs: A bird's-eye view of the market and analysis of its performance trends', *The Journal of Index Investing* **5**(1), 146–157.

Morningstar (2021), 'Which countries lead on esg?', https://www.morningstar.co.uk/uk/news/211424/which-countries-lead-on-esg.aspx. [Online; accessed 25-May-2022].

Mossin, J. (1966), 'Equilibrium in a capital asset market', *Econometrica: Journal of the econometric society* pp. 768–783.

Nagy, Z., Kassam, A. & Lee, L.-E. (2016), 'Can esg add alpha? an analysis of esg tilt and momentum strategies', *The Journal of Investing* **25**(2), 113–124.

Nguyen, N. & Nguyen, D. (2015), 'Hidden markov model for stock selection', *Risks* **3**(4), 455–473.

Nguyen, N. & Nguyen, D. (2021), 'Global stock selection with hidden markov model', *Risks* **9**(1), 9.

Nofsinger, J. & Varma, A. (2014), 'Socially responsible funds and market crises', *Journal of Banking & Finance* **48**, 180–193.

Pluciennik, P. & Janik, B. (2022), The influence of covid-19 pandemic on sustainable stock indices-the markov switching approach, *in* 'Economic and Social Development (Book of Proceedings), 77th International Scientific Conference on Economic and Social', p. 63.

Rabiner, L. & Juang, B. (1986), 'An introduction to hidden markov models', *ieee assp magazine* **3**(1), 4–16.

Rabiner, L. R. (1989), 'A tutorial on hidden markov models and selected applications in speech recognition', *Proceedings of the IEEE* **77**(2), 257–286.

Ratanapakorn, O. & Sharma, S. C. (2007), 'Dynamic analysis between the us stock returns and the macroeconomic variables', *Applied Financial Economics* **17**(5), 369–377.

Refinitiv (2022), 'Environmental, social and corporate governance - esg', https://www.refinitiv.com/en/financial-data/company-data/esg-data#:~:text=Refinitiv%20offers%20one%20of%20the,history%20going%20back%20to%202002. [Online; accessed 14-April-2022].

Refinitiv Eikon (2022*a*), 'Environmental, social and corporate governance scores from refinitiv', https://www.refinitiv.com/content/dam/marketing/en_us/documents/methodology/refinitiv-esg-scores-methodology.pdf. [Online; accessed 14-February-2022].

Refinitiv Eikon (2022*b*), 'Refinitiv eikon terminal', https://emea1.identity.ciam.refinitiv.net/auth/UI/Login. [Online; accessed 13-Mars-2022].

Ross, S. (1976), 'The arbitrage theory of capital asset pricing', *Journal of Economic Theory* **13**(3), 341–360.

Sericola, B. (2013), *Markov chains: theory and applications*, John Wiley & Sons.

Sharpe, W. F. (1964), 'Capital asset prices: A theory of market equilibrium under conditions of risk', *The journal of finance* **19**(3), 425–442.

Singh, A. (2020), ‘Covid-19 and safer investment bets’, *Finance research letters* **36**, 101729.

Sirucek, M. (2012), ‘Macroeconomic variables and stock market: Us review’.

Treynor, J. L. (1961), ‘Market value, time, and risk’, *Time, and Risk (August 8, 1961)* .

Van Gael, J., Saatci, Y., Teh, Y. W. & Ghahramani, Z. (2008), Beam sampling for the infinite hidden markov model, *in* ‘Proceedings of the 25th international conference on Machine learning’, pp. 1088–1095.

Wang, M., Lin, Y.-H. & Mikhelson, I. (2020), ‘Regime-switching factor investing with hidden markov models’, *Journal of Risk and Financial Management* **13**(12), 311.

Zhou, X., Caldecott, B., Harnett, E. & Schumacher, K. (2020), ‘The effect of firm-level esg practices on macroeconomic performance’, *University of Oxford— Working Paper* (20-03).

A Appendix

Table A.1: Distribution of industries

Industry	# Companies pre-screening	# Companies post-screening	Change
Academic Educational Services	6	3	-50%
Applied Resources	52	39	-25%
Automobiles Auto Parts	59	45	-24%
Banking Investment Services	377	175	-54%
Chemicals	68	48	-29%
Collective Investments	236	7	-97%
Consumer Goods Conglomerates	11	11	0%
Cyclical Consumer Products	151	99	-34%
Cyclical Consumer Services	162	103	-36%
Energy - Fossil Fuels	107	63	-41%
Financial Technology (Fintech)	15	7	-53%
Food Beverages	143	76	-47%
Food Drug Retailing	32	21	-34%
Healthcare Services Equipment	114	69	-39%
Industrial Commercial Services	235	160	-32%
Industrial Goods	254	183	-28%
Insurance	60	46	-23%
Investment Holding Companies	71	13	-82%
Mineral Resources	90	63	-30%
Personal Household Products	17	11	-35%
Pharmaceuticals Medical Research	145	72	-50%
Real Estate	291	110	-62%
Renewable Energy	31	9	-71%
Retailers	91	60	-34%
Software IT Services	252	123	-51%
Technology Equipment	125	77	-38%
Telecommunications Services	45	32	-29%
Transportation	95	65	-32%
Utilities	87	49	-44%
Other	692	-	
Total	4114	1839	-55%

Table A.2: Refinitiv ESG Score definition (Refinitiv Eikon 2022*a*)

Score	Definition
Refinitiv ESG resource use score	The resource use score reflects a company's performance and capacity to reduce the use of materials, energy or water, and to find more eco-efficient solutions by improving supply chain management.
Refinitiv ESG emissions reduction score	The emission reduction score measures a company's commitment and effectiveness towards reducing environmental emissions in its production and operational processes.
Refinitiv ESG innovation score	The innovation score reflects a company's capacity to reduce the environmental costs and burdens for its customers, thereby creating new market opportunities through new environmental technologies and processes, or eco-designed products.
Refinitiv ESG workforce score	The workforce score measures a company's effectiveness in terms of providing job satisfaction, a healthy and safe workplace, maintaining diversity and equal opportunities, and development opportunities for its workforce.
Refinitiv ESG human rights score	The human rights score measures a company's effectiveness in terms of respecting fundamental human rights conventions.
Refinitiv ESG community score	The community score measures the company's commitment to being a good citizen, protecting public health and respecting business ethics.
Refinitiv ESG product responsibility score	The product responsibility score reflects a company's capacity to produce quality goods and services, integrating the customer's health and safety, integrity and data privacy.
Refinitiv ESG management score	The management score measures a company's commitment and effectiveness towards following best practice corporate governance principles.
Refinitiv ESG shareholders score	The shareholders score measures a company's effectiveness towards equal treatment of shareholders and the use of anti-takeover devices.
Refinitiv ESG CSR strategy score	The CSR strategy score reflects a company's practices to communicate that it integrates economic (financial), social and environmental dimensions into its day-to-day decision-making processes.

Table A.3: Descriptive statistics

Variable	count	Mean	Std. Dev.	min	max
Momentum 1mo	1655170	0.01	0.20	-0.98	80.87
Momentum 3mo	1655170	0.03	0.38	-0.98	98.00
Momentum 6mo	1655170	0.07	0.81	-0.99	579.00
Momentum 12mo	1655170	0.15	1.63	-0.99	476.00
Volatility 1mo	1655170	3.56	10.12	0.00	4011.61
Volatility 3mo	1655170	4.34	11.07	0.00	2467.61
Volatility 6mo	1655170	4.59	11.61	0.00	1780.67
Volatility 12mo	1655170	4.80	12.24	0.00	1414.30
P/E	1537736	24.35	21.78	4.72	88.69
P/B	1537736	2.67	2.67	0.30	14.13
Price to cash flow	1537736	427.36	751.11	2.10	2452.74
Free cash flow yield	1537736	4.51	13.60	-33.27	51.65
ROE	1636413	11.30	18.79	-55.00	63.40
ROA	1636413	4.92	8.18	-22.35	27.12
EBIT Margin	1636413	16.06	28.80	-88.36	92.46
Operating Profit Margin	1636413	14.58	33.02	-109.67	108.11
Price to Sales per Share Growth	1509674	0.00	0.05	-0.16	0.18
EV to Sales Growth	1509674	0.00	0.05	-0.19	0.21
EPS Share Growth	1509674	0.06	1.70	-7.00	9.25
Free cash flow growth	1509674	-0.14	4.90	-23.48	26.86
ESG Score	829339	50.30	20.82	0.43	95.71
Environmental Pillar Score	829339	46.79	28.30	0.00	99.14
Social Pillar Score	829339	52.45	23.84	0.12	98.63
Governance Pillar Score	829339	50.61	22.87	0.29	99.33

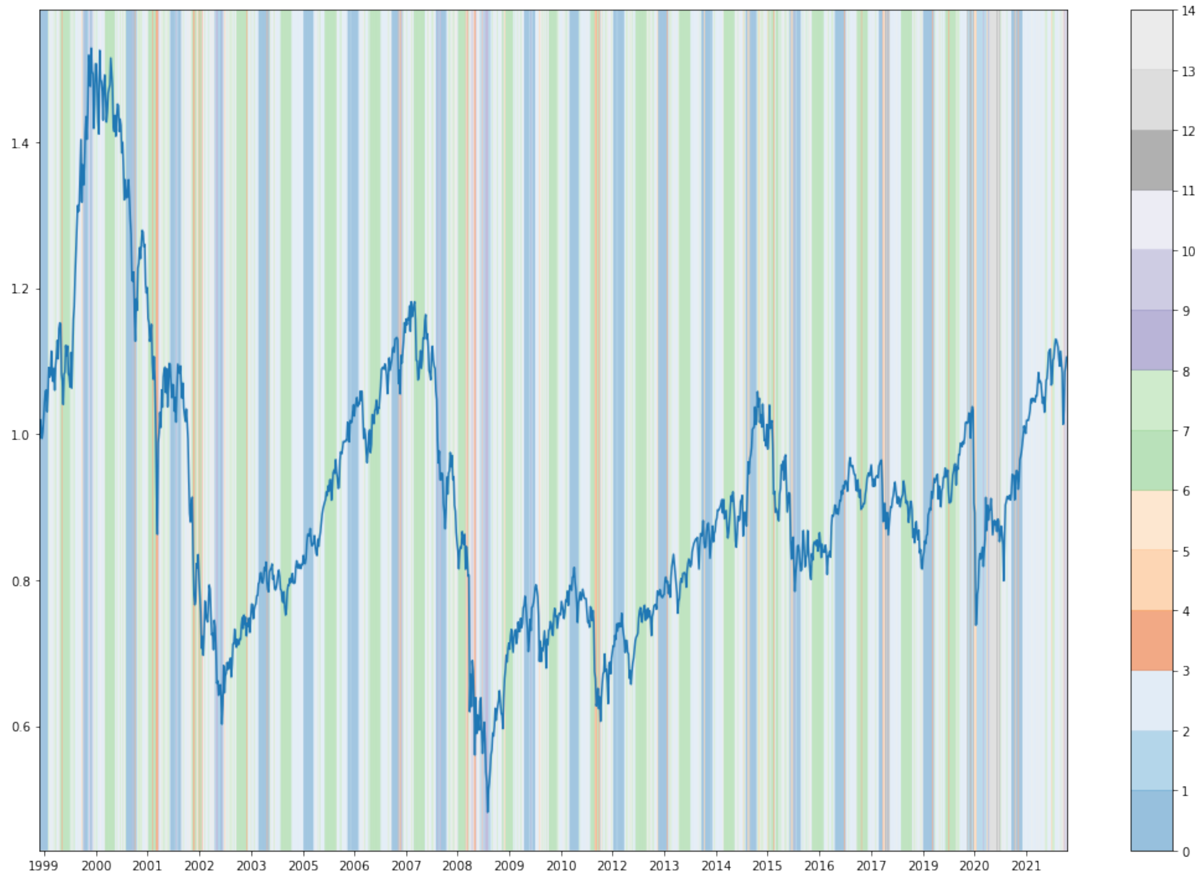

Figure A.1: Regime switches

Table A.4: Results including size factor

2003-2022	CAGR	Volatility	SR	IR	t-test P-value	α	α P-value	ESG score
Base	15.04%	30.71%	0.49	0.40	-	0.0039	0.055	44.73
ESG	5.89%	22.47%	0.26	0.19	0.12	0.0029	0.054	44.67
ESG, no E,S,G	14.51%	32.24%	0.45	0.36	0.85	0.0027	0.071	44.46
E,S,G	5.89%	22.47%	0.26	0.19	0.12	0.0029	0.054	44.67
Environmental	13.65%	32.42%	0.42	0.34	0.49	0.0024	0.119	44.46
Social	14.29%	32.18%	0.44	0.36	0.72	0.0026	0.075	44.46
Governance	13.47%	32.64%	0.41	0.33	0.39	0.0024	0.114	44.46
Market	2.38%	21.10%	0.11	-	-	-	-	-

Table A.5: Fama-MacBeth results including size factor

2003-2022	Base	ESG	ESG, no E,S,G	E,S,G	Environmental	Social	Governance
Momentum	42.86%	45.45%	33.33%	33.33%	33.33%	33.33%	33.33%
Risk	-28.57%	-18.18%	-16.67%	-13.33%	-16.67%	-16.67%	-16.67%
Value	28.57%	36.36%	25.00%	26.67%	25.00%	25.00%	25.00%
Quality	-14.29%	-9.09%	-8.33%	-6.67%	-8.33%	-8.33%	-8.33%
Growth	57.14%	54.55%	41.67%	40.00%	41.67%	41.67%	41.67%
Size	14.29%	18.18%	16.67%	20.00%	16.67%	16.67%	16.67%
ESG	-	27.27%	8.33%	-	-	-	-
Environmental	-	-27.27%	-	-20.00%	8.33%	-	-
Social	-	-36.36%	-	6.67%	-	8.33%	-
Governance	-	9.09%	-	13.33%	-	-	8.33%

Table A.6: Results for UK

2003-2022	CAGR	Volatility	SR	IR	t-test P-value	α	α P-value	ESG score
Base	15.77%	26.39%	0.59	0.37	-	0.0053	0.000	44.45
ESG	11.85%	28.12%	0.42	0.24	0.25	0.0044	0.006	42.88
ESG, no E,S,G	16.06%	26.94%	0.59	0.37	0.86	0.0055	0.000	42.79
E,S,G	11.85%	28.12%	0.42	0.24	0.25	0.0044	0.006	42.88
Environmental	16.61%	26.58%	0.62	0.39	0.70	0.0057	0.000	42.79
Social	16.54%	26.67%	0.62	0.39	0.73	0.0057	0.000	42.79
Governance	16.09%	26.61%	0.60	0.38	0.87	0.0055	0.000	42.79
Market	3.63%	17.17%	0.21	-	-	-	-	-

Table A.7: Results for Sweden

2003-2022	CAGR	Volatility	SR	IR	t-test P-value	α	α P-value	ESG score
Base	15.14%	28.54%	0.53	0.20	-	0.0049	0.003	48.86
ESG	11.97%	29.81%	0.40	0.11	0.22	0.0040	0.024	49.42
ESG, no E,S,G	13.99%	28.52%	0.49	0.17	0.14	0.0046	0.007	49.19
E,S,G	11.97%	29.81%	0.40	0.11	0.22	0.0040	0.024	49.42
Environmental	14.42%	28.50%	0.50	0.18	0.39	0.0047	0.005	49.19
Social	13.01%	28.41%	0.45	0.14	0.03	0.0043	0.010	49.19
Governance	14.79%	28.73%	0.51	0.19	0.58	0.0048	0.005	49.19
Market	7.96%	19.65%	0.40	-	-	-	-	-