Expert Systems with Applications 163 (2021) 113720

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

A novel dynamic asset allocation system using Feature Saliency Hidden Markov models for smart beta investing

Elizabeth Fons [a,c,*], Paula Dawson [a], Jeffrey Yau [b], Xiao-jun Zeng [c], John Keane [c]

[a] AllianceBernstein, London, UK
[b] University of California, Berkeley, California, USA
[c] School of Computer Science, University of Manchester, Manchester, UK

ARTICLE INFO

Article history:
Received 5 November 2019
Revised 4 May 2020
Accepted 2 July 2020
Available online 15 July 2020

Keywords:
Hidden Markov Model
Dynamic asset allocation
Portfolio optimization
Feature selection
Smart beta

ABSTRACT

The financial crisis of 2008 generated interest in more transparent, rules-based strategies for portfolio construction, with smart beta strategies emerging as a trend among institutional investors. Whilst they perform well in the long run, these strategies often suffer from severe short-term drawdown (peak-to-trough decline) with fluctuating performance across cycles. To manage short term risk (cyclicality and underperformance), we build a dynamic asset allocation system using Hidden Markov Models (HMMs). We use a variety of portfolio construction techniques to test our smart beta strategies and the resulting portfolios show an improvement in risk-adjusted returns, especially on more return-oriented portfolios (up to 50% of return in excess of market adjusted by relative risk annually). In addition, we propose a novel smart beta allocation system based on the Feature Saliency HMM (FSHMM) algorithm that performs feature selection simultaneously with the training of the HMM, to improve regime identification. We evaluate our systematic trading system with real life assets using MSCI indices; further, the results (up to 60% of return in excess of market adjusted by relative risk annually) show model performance improvement with respect to portfolios built using full feature HMMs.

1. Introduction

Smart beta is a relatively new term that has become ubiquitous in asset management over the last few years. The financial theory underpinning smart beta, known as factor investing, emerged in the 1960s, when factors were first identified as being drivers of equity returns (Agather & Gunthorp, 2017). These factor returns may be a source of risk and/or improved return, hence understanding whether any additional risk is adequately compensated with higher returns is important (Ang, 2014).

By selecting stocks based on their factor exposures, active managers can build portfolios with particular factor exposures and so use factor investing to improve portfolio returns and/or lower risk, depending on their objectives. Smart beta aims to achieve these goals at a reduced cost by utilising a transparent, systematic, rules-based approach, bringing down costs significantly when compared to active management (Asness, 2016).

While smart beta strategies have shown strong performance in the long run, they often suffer from severe short-term drawdown (peak-to-trough decline) with fluctuating performance across cycles (Arnott, Beck, Kalesnik, & West, 2016). These fluctuations can arise from extreme macroeconomic conditions, elevated volatility, heightened correlations across multiple markets and uncertainty monetary and fiscal policy responses.

This paper addresses these concerns by building a regime switching model using Hidden Markov Models (HMMs). An introduction to regime switching models applied to macroeconomics is given inHamilton (2010), Kim and Nelson (1999) present regime switching models with application examples to business cycles and finance. HMMs have become a mainstream technique to model nonlinearity in time series data that other methods such as autoregressive models cannot do (Baum, Petrie, Soules, & Weiss, 1970; Rabiner, 1989). We study how a regime switching framework can be used to detect regimes across factors and, if so, add value to smart beta strategies. The dominant approach in regime switching frameworks for asset allocation has been to specify a static decision rule dependent on the predicted state in advance (Nystrup, Madsen, & Lindstrom, 2017a; Reus & Mulvey, 2016; Ang & Sorensen, 2012; Nguyen, 2017; Erlwein, Mamon, & Davison, 2011; Axelsson, 2017).

* Corresponding author.
E-mail addresses: elizabeth.fons@postgrad.manchester.ac.uk (E. Fons), paula.dawson@alliancebernstein.com (P. Dawson), jyau@berkeley.edu (J. Yau), x.zeng@manchester.ac.uk (X.-j. Zeng), john.keane@manchester.ac.uk (J. Keane).

https://doi.org/10.1016/j.eswa.2020.113720
0957-4174/

An alternative approach is to dynamically optimise a portfolio using information from the inferred regime parameters. We follow this approach and use the regime information combined with a variety of portfolio construction strategies to generate different types of portfolios (more return-oriented, where the return is part of the optimisation process and more risk- focused, where only the volatility is taken into account to optimise the portfolios). Initially we build a dynamic asset allocation (DAA) system to construct portfolios using a regime switching model and perform a systematic analysis using hundreds of combinations of factors by training the HMM with the same factors that will be used for allocation in the portfolio. Our study shows that using regime information from the HMM has a better performance than a single regime allocation and we find that more return-oriented portfolios yield better risk-adjusted returns than their benchmarks, while the performance of more risk-focused portfolios show some improvement.

Finally, the common factor in the majority of the work on regime-switching models in finance considers either a single or a small set of assets to build the model, with the selection criteria for the assets usually coming from domain knowledge (Mulvey & Liu, 2016; Ang & Timmermann, 2012; Nystrup, Hansen, Madsen, & Lindstrom, 2016; Nystrup, Hansen, Larsen, Madsen, & Lindstrom, 2017). The reason being that unsupervised feature selection for HMMs is very limited, with wrapping methods exhibiting high computational cost or with very few methods specific for HMMs (Adams & Beling, 2017). In most applications of HMMs, features are either pre-selected based on expert knowledge or feature selection is omitted entirely. One of the few feature selection algorithms developed for HMMs is the feature saliency HMM (FSHMM) Adams, Beling, and Cogill (2016), where the feature selection process is embedded within the HMM training. We incorporate this FSHMM into our DAA system yielding two benefits: (1) by selection during training we expect to improve regime identification by selecting features that are state dependent and rejecting state independent features; (2) it allows incorporation of many features in a model and allows the algorithm to decide which contribute to regime identification, thus avoiding the need for expert knowledge.

The main contributions of this paper are as follows:

1. We build a DAA system using an HMM for regime detection and perform a systematic study using multiple combinations of assets and compare performance with two benchmarks: their single-regime portfolio counterpart and an equally weighted portfolio. We show the DAA system consistently out-performs these benchmarks;
2. We extend our DAA system by incorporating a Feature Saliency HMM for feature selection, to reduce noise from spurious data (state independent) that may worsen model performance, leading to improved regime identification;
3. We test the DAA system with embedded feature selection on real life investable indices using MSCI indices and show an improvement in risk-adjusted return on strategies built using the DAA system with FSHMM with respect to strategies built using DAA without feature selection.

This paper is organized as follows: Section 2 overviews previous work on HMM in finance; Section 3 introduces HMMs and FSHMMs; data and index construction are described in Section 4; Section 5 introduces our DAA system, the feature saliency algorithm and outlines its incorporation into the DAA system; Section 6 shows the experimental results of the DAA system, and the incorporation of embedded feature selection; conclusions and further work are presented in Section 7.

2. Previous work

HMMs have been used extensively in finance to build regime-based models, since Hamilton proposed use of a regime-switching model to identify economic cycles using the GNP series (Hamilton, 1989). As pointed out by Ang and Timmermann (2012) HMMs simultaneously capture multiple characteristics from financial return series such as time-varying correlations, skewness and kurtosis, while also providing good approximations even in processes for which the underlying model is unknown (Ang & Bekaert, 2003; Bulla, Mergner, Bulla, Sesboue, & Chesneau, 2011; Bulla & Bulla, 2006; Nystrup, Madsen, & Lindstrom, 2015; Nystrup, Madsen, & Lindstrom, 2017b). Further, HMMs allow for good interpretability of results, as thinking in terms of regimes is natural in finance. Examples of DAA are Reus and Mulvey (2016) that uses a HMM to build a dynamic portfolio using currency futures and Bae, Kim, and Mulvey (2014) that uses a HMM to identify market regimes using different asset classes, with regime information helping portfolios to avoid risk during left-tail events.

Guidolin (2012) provides an extensive review of applications of Markov switching models in empirical finance covering stock returns, term structure of default-free interest rates, exchange rates and joint processes of stock and bond returns.

Outside of asset allocation, HMMs have been used to capture energy prices dynamics (Dias & Ramos, 2014) to build credit risk systems, for example Petropoulos, Chatzis, and Xanthopoulos (2016) build a credit rating system using a students'-t HMM, addressing two problems in current systems: their heavy-tailed actual distribution and their time-series nature; Elliott, Siu, and Fung (2014) build a model using a double HMM to extract information about true credit qualities of firms. Dabrowski, Beyers, and de Villiers (2016) study HMMs and other Bayesian networks to build early warning systems to detect systemic banking crises and find that Bayesian methods provide superior performance on early warning compared with traditional signal extraction logic models; and Zhou and Mamon (2012) investigate three popular short-rate models and extend them to capture the switching of economic regimes using a finite-state Markov chain.

To date, little work has been done on the impact of regime switching models on factor investing. Among this work, Guidolin and Timmermann (2008) found evidence of four economic regimes in size and value factors that capture time-variations in mean returns, volatilities and return correlations. Liu, Xu, and Zhao (2011) and Ma, Maclean, Xu, and Zhao (2011) study time-varying risk premiums using a regime switching model with 3 states. The regime switching model is trained using six well-established factors found in the literature and the assets used for allocation are 9 sector ETFs (Exchange Traded Fund). They achieve a Sharpe ratio of 0.18 but the work covers a short period of testing time (9 months) and does not consider transaction costs.

3. Theoretical background

This section presents the FSHMM that can simultaneously train the model and perform feature selection. An introduction on HMMs and associated notation can be found in A.

3.1. FSHMM

The FSHMM considers a feature relevant if its distribution is dependent on the underlying state and irrelevant if it is independent. Given a set of binary variables $\mathbf{z} = \{z_1, \ldots, z_L\}$ that indicate the relevance of the feature, i.e. $z_l = 1$ if the lth feature is relevant and $z_l = 0$ if its irrelevant, the feature saliency ρ_l is defined as the probability that the lth feature is relevant. Assuming the features are conditionally independent given the state enables the multivariate Gaussian to be written as a multiplication of univariate Gaussians, and the conditional distribution of y_t given $\mathbf{z}$ and $\mathbf{x}$ can be written as follows:

$$p(y_t|\mathbf{z}, x_t = i, \bar{\Lambda}) = \prod_{l=1}^{L} r(y_{lt}|\mu_{il}, \sigma_{il}^2)^{z_l} q(y_{lt}|\epsilon_l, \tau_l^2)^{1-z_l} \tag{1}$$

where $r(y_{lt}|\mu_{il}, \sigma_{il}^2)$ is the Gaussian conditional feature distribution for the l-th feature and $q(y_{lt}|\epsilon_l, \tau_l^2)$ is the state-independent feature distribution. The Gaussian FSHMM model parameters are $\bar{\Lambda} = (\pi, A, \mu, \sigma, \rho, \epsilon, \tau)$ where the first four parameters correspond to the regular HMM, ρ is the feature saliency and ϵ and τ are the mean and variance of the state independent Gaussian feature distribution. Fig. 1 shows the FSHMM.

The marginal probability of $\mathbf{z}$ is:

$$p(\mathbf{z}|\Lambda) = \prod_{l=1}^{L} \rho_l^{z_l} (1-\rho_l)^{1-z_l} \tag{2}$$

The joint probability distribution of y_t and $\mathbf{z}$ given $\mathbf{x}$ is:

$$p(y_t, \mathbf{z}|x_t = i, \Lambda) = \prod_{l=1}^{L} [\rho_l r(y_{lt}|\mu_{il}, \sigma_{il}^2)]^{z_l} [(1-\rho_l) q(y_{lt}|\epsilon_l, \tau_l^2)]^{1-z_l} \tag{3}$$

The complete likelihood for the FSHMM is given by:

$$p(\mathbf{x}, \mathbf{y}, \mathbf{z}|\Lambda) = \pi_{x_0} p(y_0, \mathbf{z}|x_0, \Lambda) \prod_{t=1}^{L} a_{x_{t-1}, x_t} p(y_t, \mathbf{z}|x_t, \Lambda) \tag{4}$$

The MAP estimation of the FSHMM is similar to the HMM using EM but the Q function incorporates the hidden variables associated with feature saliency and can be written as:

$$\begin{aligned} Q(\Lambda, \Lambda') = \; & E[\log p(\mathbf{x}, \mathbf{y}, \mathbf{z}|\Lambda)|\mathbf{y}, \Lambda'] \\ & = \sum_{\mathbf{x}, \mathbf{z}} \log p(\mathbf{x}, \mathbf{y}, \mathbf{z}|\Lambda) P(\mathbf{x}, \mathbf{z}|\mathbf{y}, \Lambda') \end{aligned} \tag{5}$$

The update steps of the EM algorithm are shown in Appendix B and the pseudocode for the MAP FSHMM formulation is given in Algorithm 1[1]. A detailed description of the equation derivations and the steps of the algorithm can be found in Adams (2015).

Algorithm 1: MAP FSHMM Algorithm

1: Select initial values for $\pi_i, a_{ij}, \mu_{il}, \sigma_{il}, \epsilon_l, \tau_l$ and ρ_l for $i = 1 \dots I,\ j = 1 \dots I$, and $l = 1 \dots L$
2: Select initial values for the prior hyperparameters $\bar{p}_i, \bar{a}_{ij}, m_{il}, s_{il}, \zeta_{il}, \eta_{il}, b_l, c_l, \nu_l, \psi_l$ and k_l for $i = 1 \dots I,\ j = 1 \dots I$, and $l = 1 \dots L$ following B.8 to B.14
3: Select stopping threshold δ and maximum number of iterations M
4: Set absolute percentage change in the posterior probability between current iteration and previous iteration $\Delta\mathscr{L}$ to ∞ and the number of iterations it to 1
5: **while** $\Delta\mathscr{L} > \delta$ and $it < M$
6: E-step: calculate probabilities $\gamma_t(i), \xi(i,j), e_{ilt}, h_{ilt}, g_{ilt}, u_{ilt}, v_{ilt}$ following B.1 to B.7
7: M-step: update parameters $\pi_i, a_{ij}, \mu_{il}, \sigma_{il}^2, \epsilon_l, \tau_l^2, \rho_l$ following B.15 to B.21
8: $\Delta\mathscr{L}$
9: $it = it + 1$
10: **end while**
11: Perform feature selection based on ρ_l and construct reduced models

As well as the parameters estimated through EM, the model also has several hyperparameters to set in advance. The most relevant is the weight parameter k_l that can be used as an informative exponential prior on ρ. Setting higher values of k_l for a feature translates into assigning a higher cost to it (for example, making explicit that collecting the feature is more expensive), so in order for the algorithm to select the feature, it requires more evidence of its relevance. This can be used either to reduce the number of selected features or as a proxy for the cost of selecting a feature in the optimization process. The heuristic to select a reasonable value of k_l is to scale it with the number of observations as $T/4$ with T the number of observations.

3.2. Smart beta investing

Smart beta is a systematic, low cost implementation of factor investing, where securities are selected based on their exposure to an attribute that has been associated with a persistent higher return in the past, called a factor. Factors can be fundamental characteristics of the economy (macroeconomic factors) or of companies (style factors). Macroeconomic factors can be thought of as capturing the broad risks and returns across assets classes, whilst style factors can be thought of as aiming to explain returns and risks for securities within asset classes.

This paper considers style factors in the equity market; within these style factors, dozens of indicators have been identified. The majority can be grouped into families, with style factors within a family measuring similar characteristics and often highly correlated (Ang, 2014; Fama et al., 2015; Dimson, Marsh, & Staunton, 2017). An example of this is momentum, which includes factors measuring returns over different periods (3-months, 6-months, 12-months etc). While there is no universal definition of these families or the factors that belong in each family, there are common themes. Typically, families comprise: value, growth, momentum, quality, size and some sort of volatility/risk/beta measure. There may be variations on this, for example Dividend Yield is sometimes viewed as a factor family in its own right or sometimes

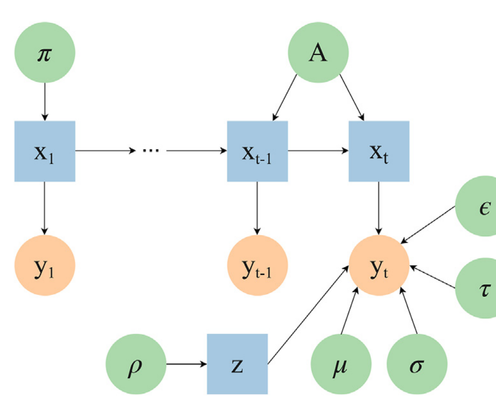

Fig. 1. FSHMM: squares with x_t represent latent variables, circles with y_t are observations and circles with $\{\pi, A, \rho, \mu, \sigma, \epsilon, \tau\}$ represent model parameters.

[1] The implementation of the algorithm can be downloaded from: https://github.com/elifons/FeatureSaliencyHMM

it is viewed as a member of the Value family; sometimes the Value family can be split into *Value* and *Deep Value*.

4. Data

Below is the description of the two datasets used in this work; Table 1 summarises their main characteristics.

4.1. Dataset-1: daily factor data from S&P500 index

Dataset-1 is a set of style factors which are constructed based on the S&P 500 universe of US stocks. The style factor for each individual stock is determined, the universe is ranked and a portfolio is constructed with the top 20% of stocks and short positions (negative weights) in the bottom 20% of stocks. This is repeated each month. The resulting style factor portfolio will have a strong exposure to the factor and no exposure to the overall market (as the negative holdings offset the positive weights) – Table 2 shows these. The data is supplied by a broker and consists of 25 style factors covering a time period from 1988 to 2016. This dataset is used throughout the analysis.

4.2. Dataset-2: daily MSCI USA enhanced indices

Dataset-2 is supplied by MSCI and consists of a range of indices which they publish. As with Dataset-1, the individual style factors are calculated using underlying stocks and their style factor exposures. These individual style factor indices are then grouped into six style factor families, and it is these indices that are used in this paper. We use the six MSCI USA enhanced style indices:: value, low size, momentum, quality, low volatility and dividend yield Bender, Briand, Melas, and Subramanian, 2013. These have different inception dates, with the most recent beginning in 1999, which limits the period for which we can use this dataset to 1999–2016. Fig. 2 shows the cumulative return of the MSCI indices for the testing period, net of market.

The advantage of using a published set of indices (such as MSCI indices) is that they can be packaged into an easy to purchase product, such as an ETF, by a separate investment company. As an example, an investor who wants to buy US value stocks can buy an MSCI US enhanced Value ETF, which would involve buying one security (the ETF) rather than the underlying stocks. By removing the need to analyse and purchase the underlying companies, the complexity and cost of implementing a smart beta strategy can be reduced. This allows us to test our DAA system using real world assets.

Table 1
Description of datasets.

Dataset	Date	Nr of features	Frequency
1: Factor data	Jan-1988 to Feb-2016	25	Daily
2: MSCI Enhanced	Jan-1999 to Feb-2016	6	Daily

5. DAA system

Investment on single factor strategies has been shown to have significant returns over the long term; however, it is not straightforward to build multi-factor strategies and rotate factors according to market conditions. Factor indices are time series data, hence we take advantage of the capacity of HMMs to identify underlying regimes in sequences of observations and build a DAA system. We will firstly determine the optimal number of hidden states to model market regimes and then, in order to avoid excessive transactions costs through frequent rebalancing, we optimize the rebalancing signal.

5.1. DAA system architecture

We design a dynamic trading framework with daily evaluations and monthly re-adjustments as shown in Fig. 3. Each day a new vector of returns is added to the training set with an expanding window, and the state is predicted. Returns are lagged by one day in order to avoid look-ahead bias. Because this prediction is noisy, we determine an optimal window of consecutive days in the new state before the portfolio is rebalanced. Once a change of state has been accepted, the vector of means and covariance matrix from the new state are retrieved and the portfolio weights optimized using the different portfolio construction techniques, with transaction costs calculated after rebalance. Once a full month

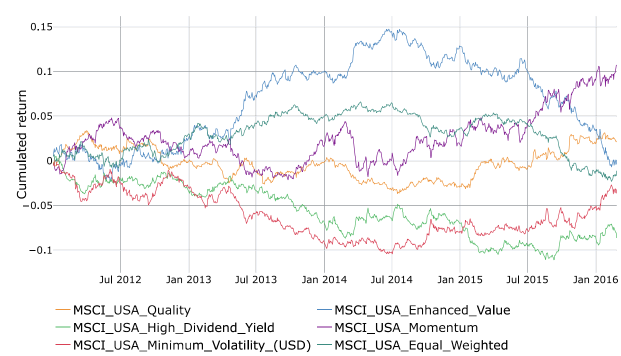

Fig. 2. Cumulative returns of MSCI USA enhanced factors. Returns are in excess of the market in USD, for the date range Jan 2012–Feb 2016.

Table 2
Representative factor indices used for building regime switching frameworks.

#	Factor	Family	#	Factor	Family
1	Book Value Yield	Value	14	Operating Margin Growth-1Yr	Quality
2	1 Yr Fwd Earnings Yield	Value	15	Operating Margin Growth-3Yr	Quality
3	Free Cash Flow Yield	Value	16	Historical Free Cash Flow Growth-1Yr	Growth
4	Sales Yield	Value	17	Historical Free Cash Flow Growth-3Yr	Growth
5	Dividend Yield	Value	18	Historical DPS Growth-1Yr	Growth
6	Historical ROE	Quality	19	Historical DPS Growth-3Yr	Growth
7	Operating (EBIT) Margin	Quality	20	6 Month Price Momentum	Momentum
8	AltmanZ	Quality	21	12 Month Price Momentum	Momentum
9	ROA	Quality	22	3 Month Avg Mean EPS	Quality
10	Piotroski	Quality	23	Size	Risk
11	Earnings Growth FY1 to FY2	Growth	24	EPSCV	Quality
12	Historical Sales Growth-1Yr	Growth	25	Beta	Risk
13	Historical Sales Growth-3Yr	Growth			

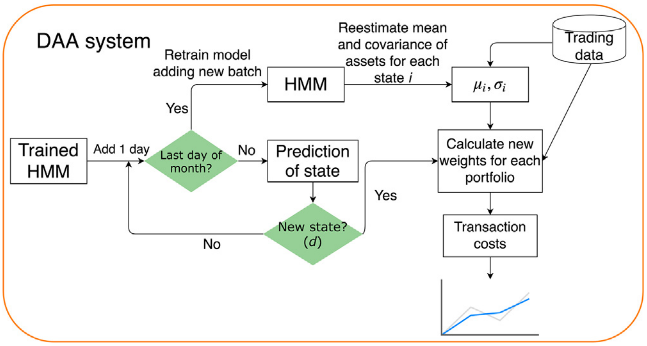

Fig. 3. Dynamic Asset Allocation system diagram.

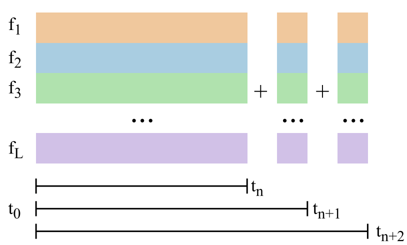

Fig. 4. Data scheme.

has passed, we add this new batch of data to the training set with an expanding window and retrain the model. Fig. 4 shows how data is added daily with an expanding window. While this will not produce immediate changes in the model parameters (transition matrix and emission distributions), in time they should change slightly to accommodate the new information. Therefore, we can capture changes on the dynamics of the system over time.

5.1.1. Model selection

The number of latent states in a HMM has to be set before training. An option is to use the Bayesian Information criterion (BIC), a penalized log-likelihood function that can be used for model selection (Schwarz, 1978). BIC is defined by:

$$BIC = -2 \log p\left(D|\hat{\theta}\right) + d \log(N)$$

where d is the number of free parameters in the model and N is the number of samples. Thus, calculating the score over a range of K states, we can select the model with the lowest value. Another option is to follow a greedy approach, calculating performance of the portfolios built with a different number of regimes and selecting the model with highest performance.

In the financial HMM literature (Guidolin & Timmermann, 2008), regime switching models normally range between 2 and 4 states so we selected random combinations of 5 assets each (where each asset belongs to a factor family as described in Table 2) and and used these combinations to train an HMM with 2, 3, 4, 5 and 6 hidden states respectively. Keeping the number of states low allows better interpretability, and with more hidden states (5 or 6) we observe that some states only occur for a few days so is more likely to be caused by over-fitting, and the covariance calculation is inaccurate for these states. We draw 200 combinations of assets to build portfolios in order to estimate the optimal number of states (more than 15% of the total number of possible combinations). From each HMM information we built different types of portfolios – explained in Section 6.1. The performance of each portfolio was calculated using the Information Ratio (IR – the ratio between excess return and standard deviation of excess returns, annualized

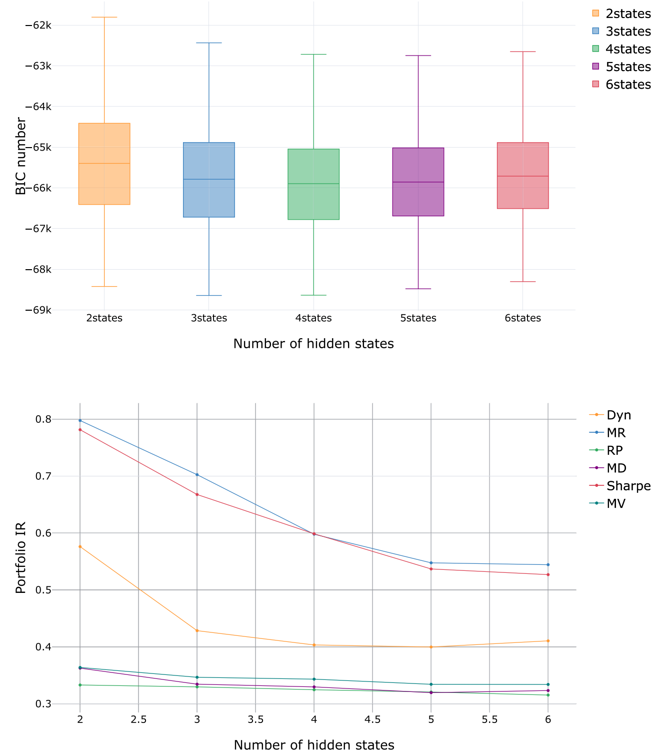

Fig. 5. The top plot shows the boxplot of BIC number for different number of states for 200 combinations of assets: a two state model has a higher BIC but there is no distinction between three, four and five; the bottom plot shows performance of portfolios as a function of number of hidden states. The two state model yields better performance for the majority of portfolios.

Bacon (2012)); the plots of BIC and performance as a function of number of states are shown in Fig. 5. The BIC score is quite similar for states 3 to 6 (4 being the lowest) and is slightly higher for 2 states. While this suggests use of a 4-regime model, performance of portfolios for 3 and 4 states is significantly lower than for 2 states, so we have selected a two-state model. Two-state models can be interpreted as expansion–contraction.

5.1.2. System calibration

The DAA system requires a trained HMM to model regime changes and the selection of an optimal time window to decide when a change of state has taken place and the portfolio has to be rebalanced.

For the first part of the work – where we test if the proposed DAA system adds value to multi-factor strategies – we use multiple combinations of factors, and calibrate the system for each combination. From a pool of 25 factor indices we select n assets randomly and use their returns to train a HMM. As factors can be grouped into five families (following Table 2), we randomly select one factor from each group so all families are represented. This yields a total of 1260 combinations. We then use the same factors to build the portfolios.

We divide the data set into three subsets: training set (15 years), validation set (9 years) and test set (4 years). In order to avoid getting stuck in a local maximum we do random initialization with initial parameters calculated from the training data and select the model with highest score. Fig. 6 shows the process of training, validation and test using the DAA system.

The regime prediction is done by passing the whole series of returns up to the previous day to decode the most probable sequence of hidden states, and keep the last value as the state prediction. This daily prediction is noisier than it would be if a whole

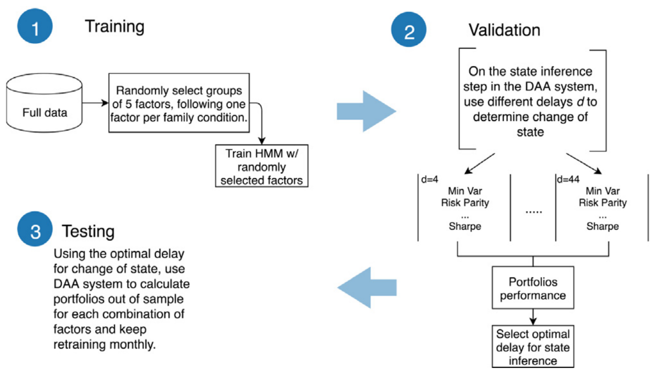

Fig. 6. Full schematic of calibration and usage of the DAA system for smart beta investing.

month of returns was passed together, and we cannot re-balance a portfolio each time a change of state is flagged without incurring large transaction costs, as quite often this would mean a daily re-balance. Instead, in the validation set, we look for a window of *d* consecutive days in the same new state and then we flag a change of regime and re-balance the portfolio accordingly. Fig. 7 shows the performance of a selection of portfolios as a function of the time window *d*. While certain combinations of assets perform consistently better than others with larger windows, smaller windows have the worst performance in all cases. The main reason is that performance of portfolios is adjusted for transaction costs, so smaller windows mean higher portfolio turnover and therefore, higher costs. We use the training set to identify the optimal window for each combination of assets.

5.2. DAA system with feature saliency: FS-DAA

So far, we have proposed a DAA system where the time series to train the HMM were known in advance, which can be a limitation. To address this, we propose a novel DAA system that incorporates an embedded feature selection method during training, by using a FSHMM, described in Section 3.1. This new method, termed FS-DAA, allows selection of features that contribute to the regime identification, called regime dependent, and rejects features that do not depend on the regimes.

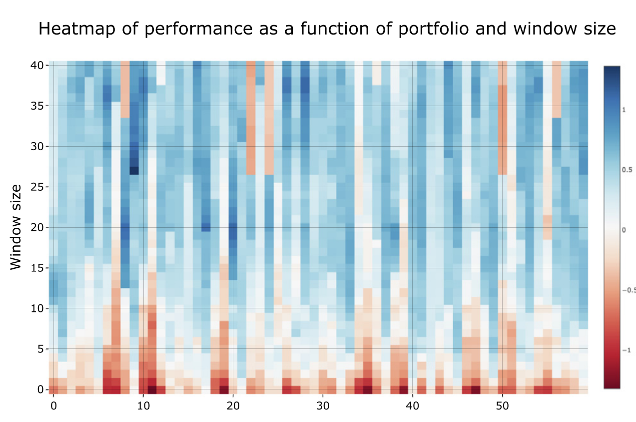

Fig. 7. A subset of the 1260 portfolios is plotted. The colormap corresponds to the performance measured by IR (adjusted for transaction costs) as a function of window size. In the majority of cases performance is low for smaller windows due to frequent re-balance; performance tends to improve with window size, 15. However, if the window is too large, performance may decrease again as it fails to take advantage of more frequent regime changes.

Fig. 8 shows the different stages for training, validation and test using FS-DAA. FS-DAA takes multiple time series data and fits an FSHMM, that assigns a saliency to each factor time series. Higher saliency means that the feature is more relevant and therefore is selected. Because FSHMM proposes that features are conditionally independent, the fitted model has diagonal covariance matrices. We therefore take the selected relevant features and used them to train a HMM with full covariance matrices.

As a first step to assess whether FSHMM can distinguish between relevant features and noise, we generated irrelevant features of random noise and added them to our daily factor data set. We tested this using different numbers of relevant and irrelevant features (following Adams et al. (2016)), number of observations and values of k_l. For each case, k_l was the same for all features, both relevant and noise. Results are summarized in Tables 3–6. In all cases, the algorithm assigned low values of saliency for the irrelevant features and high values for the relevant ones.

Secondly, we train a DAA system using all 25 features from the factor dataset, and we train an FS-DAA system that takes the 25 features, selects the relevant ones and then trains a HMM only with those factors and compares the regimes obtained. Finally, using both systems, we build a strategy using an MSCI USA enhanced family of factor indices. Both models are trained using 16 years of data (1990–2006) and then retrained every month until 2016. We use 7.5 years of trading data to estimate mean and covariance of the MSCI indices for each regime, (Jan 1999–June 2006), to obtain a robust estimation of the covariance matrix for both regimes. We then use a validation set of 6 years to select the optimal time window to set a change of state, and a test set of 4 years.

An advantage of the proposed DAA system is that it allows the data used to train the HMM to detect regimes to be decoupled from the data used for allocation. This is useful for factor investing as we can build factors with a long history (as the factor dataset) and then use real life, investable assets that have a shorter history (MSCI enhanced data) to build portfolios.

6. Results and analysis

We compare the DAA system performance with baseline strategies on the large factor dataset. Then, the implementation of the FSHMM algorithm is discussed. Lastly, we test the proposed FS-DAA system with real life assets using the MSCI indices dataset.

6.1. Trading strategies and benchmarks

Instead of constructing one kind of portfolio we build several: Risk Parity, Maximum diversification, Minimum Variance,

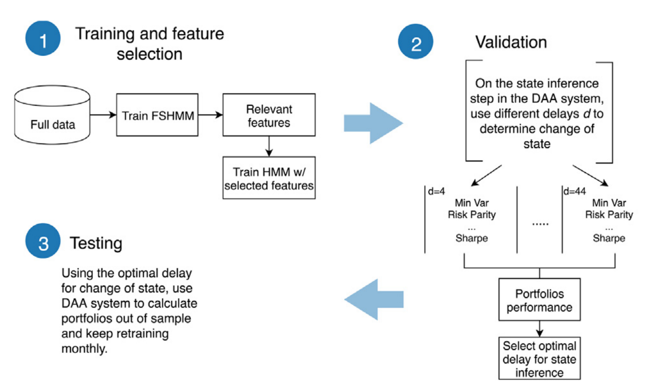

Fig. 8. Full schematic of calibration and usage of the FS-DAA system with embedded feature selection for smart beta investing.

Table 3
Average performance of portfolios built using HMMs and their benchmarks. Top portfolios that are more aggressive have a higher risk adjusted return (measured through IC and Sortino ratios) than their unconditional counterpart and the equally weighted portfolio. Bottom portfolios that are more defensive (only the covariance matrix is taken into account in the construction process) perform worse than their benchmark counterparts and the EQ portfolio.

	Ann ret	Ann vol	IR	Skw	kurt	D. risk	Sortino	DD	DD days
EQ	0.77	2.88	0.26	−0.14	0.81	2.05	0.37	379	318
Dyn HMM	1.67	4.73	0.34	−0.19	1.35	3.37	0.48	32	291
Dyn Bench	−0.60	3.98	−0.14	−0.40	1.68	2.96	−0.19	1136	682
Sharpe HMM	2.31	4.66	0.53	−0.19	1.16	3.29	0.75	429	253
Sharpe Bench	−3.14	4.89	−0.64	−0.79	4.49	3.80	−0.82	1375	873
MR HMM	3.19	7.03	0.46	−0.19	1.34	4.98	0.65	35	264
MR Bench	−5.03	7.20	−0.69	−0.78	3.71	5.63	−0.88	>4000	1001
MV HMM	0.61	2.41	0.24	−0.14	0.96	1.72	0.35	662	309
MV Bench	−0.12	2.24	−0.07	−0.11	0.83	1.61	−0.09	520	511
MD HMM	0.69	2.54	0.26	−0.14	1.01	1.80	0.37	340	306
MD Bench	0.01	2.39	−0.02	−0.12	0.84	1.71	−0.02	454	447
RP HMM	0.63	2.58	0.24	−0.13	1.04	1.84	0.34	212	302
RP Bench	0.20	2.40	0.07	−0.13	1.04	1.72	0.10	475	416

Table 4
Metrics for portfolios built using FSHMM, all assets (HMM), their benchmark and the individual MSCI indices used to build the portfolios. The metrics covered the period Jan 2012–Feb 2016.

	Ann ret	Ann vol	IR	Skw	kurt	D. risk	Sortino	DD	DD days
Sharpe FSHMM	0.06	0.50	0.12	−0.71	2.85	0.37	0.16	−94	387
Sharpe HMM	−0.11	0.65	−0.16	−0.70	3.84	0.49	−0.22	−164	522
Sharpe Bench	−1.62	0.92	−1.76	−2.75	15.0	0.82	−1.98	19825	1452
Dyn FSHMM	0.39	0.65	0.61	−0.41	0.84	0.47	0.84	−52	141
Dyn HMM	−0.02	0.60	−0.03	−1.12	9.03	0.45	−0.04	−175	566
Dyn Bench	−1.10	1.03	−1.07	−2.76	16.2	0.88	−1.24	−1508	1123
MR FSHMM	2.02	3.20	0.63	−0.39	1.83	2.30	0.88	−82	62
MR HMM	1.85	3.19	0.58	−0.39	1.84	2.29	0.80	−92	62
MR Bench	−3.46	3.78	−0.91	−2.71	20.5	3.17	−1.09	−4032	1250
MSCI Quality	0.50	2.76	0.18	0.20	2.02	1.90	0.26	−208	837
MSCI Enhanced Value	0.03	3.97	0.01	0.03	0.86	2.83	0.01	−105	599
MSCI High Dividend Yield	−2.16	3.22	−0.67	0.38	0.85	2.24	−0.96	−2374	1317
MSCI Momentum	2.48	4.35	0.57	−0.35	1.42	3.11	0.80	−144	475
MSCI Minimum Volatility	−0.89	3.58	−0.25	0.10	0.69	2.52	−0.35	−38371	906
MSCI Equal Weighted	−0.27	2.94	−0.09	−0.05	0.74	2.09	−0.13	−135	675

Table 5
Feature saliency of five factor returns time series (ρ_1 to ρ_5) and three irrelevant series of random noise (ρ_6 to $\rho 8$), all calculated with $k = 50$. All irrelevant features have saliency below 0.25, and most of the financial series have saliency close to one, except ρ_3 that has a small saliency in most of the cases.

Case		ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	$\boldsymbol{\rho}_{bf6}$	$\boldsymbol{\rho_7}$	$\boldsymbol{\rho_8}$
500 points	2 states	0.99	0.97	0.31	0.98	0.97	0.14	0.04	0.05
500 points	3 states	0.99	0.99	0.26	0.98	0.99	0.17	0.04	0.07
2000 points	2 states	0.99	0.99	0.19	0.99	0.99	0.02	0.01	0.02
2000 points	3 states	1.00	1.00	0.12	1.00	1.00	0.07	0.20	0.03

Table 6
Feature saliency of ten factor returns time series (ρ_1 to ρ_{10}) and five irrelevant series of random noise (ρ_{11} to ρ_{15}). With a small value of k all irrelevant features are discarded and all relevant features have high saliency. With a larger k, noise features are discarded, but also financial features start being selected. All series have 3800 observations.

Case		ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6	ρ_7	ρ_8	ρ_9	ρ_{10}	$\boldsymbol{\rho_{11}}$	$\boldsymbol{\rho_{12}}$	$\boldsymbol{\rho_{13}}$	$\boldsymbol{\rho_{14}}$	$\boldsymbol{\rho_{15}}$
$k = 100$	2 states	0.99	0.99	0.56	0.99	0.91	1.00	0.99	0.95	0.99	0.97	0.11	0.11	0.04	0.26	0.07
$k = 100$	3 states	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.24	0.09	0.40	0.10	0.11
$k = 380$	2 states	0.75	0.03	0.13	0.98	0.44	0.99	0.99	0.17	0.98	0.14	0.05	0.02	0.02	0.01	0.04
$k = 380$	3 states	1.00	0.37	0.08	0.99	0.55	1.00	1.00	0.13	0.99	0.22	0.04	0.17	0.04	0.05	0.03

Max return, Max Sharpe and a modified max return, where all portfolios are long only, i.e. the weights are always positive. In most cases, portfolio construction is an optimization problem, where the weights of the portfolio are optimized to maximize/minimize a desired utility function, as described below.

- **Max return:** Given an estimated vector of means, it maximizes the return given a constraint that no asset can have a weight greater than 80%.
- **Dyn:** If all estimated mean asset returns are positive, it weights the assets proportional to their mean, otherwise it equally weights them.

- **Sharpe:** A classic mean–variance portfolio that maximizes return given a set level of risk.
- **Risk parity:** Focuses on allocation of risk: each asset in the portfolio contributes the same risk as defined by:

$$\frac{w_j(Vw)_j}{\sqrt{wVw\prime}}$$

  where V is the covariance matrix.
- **Max diversification** Maximizes the diversification ratio defined as:

$$\frac{w\prime\Sigma}{\sqrt[2]{w\prime Vw}}$$

  where Σ is a vector of all asset volatility and V is the covariance matrix.
- **Min Var:** finds the portfolio with minimum variance, defined by:

$$w\prime Vw$$

  where V is the covariance matrix.

Risk Parity (RP), Maximum diversification (MD) and Minimum Variance (MV) are constructed taking into account only the covariance matrix, so they can be considered more risk aware. Max return (MR), Max Sharpe (Sharpe) and modified max return (Dyn) all consider the mean of the return during the construction, so they tend to be more aggressive.

For comparison, we built an equally weighted portfolio and a benchmark for each asset combination. Each benchmark is constructed using the same optimization method as its DAA system counterpart, but are rebalanced monthly and the covariance matrix is estimated using "single regime" past returns. The DAA system instead has two covariance matrices, one for each regime. All portfolios and their benchmarks are constructed considering transaction costs. Costs are calculated by multiplying portfolio turnover (how much a portfolio is rebalanced) with a transaction cost of 50bps (0.5%), for each selling and buying.

6.2. DAA system compared to baseline

We first evaluated our DAA system by using 1260 combinations of randomly selected assets to train the HMM and for the allocation, and compared it with their benchmarks.

Fig. 9 shows the performance measured through the Sortino ratio of all portfolios calculated using the DAA system, and their benchmarks. The Sortino ratio is the annualized return divided by the downside risk, therefore it differentiates harmful volatility from total overall volatility in contrast with IR (no risk free asset is considered). We see that all portfolios constructed using regime information perform better than their counterpart. Using the mean returns in the optimization steps, the more return-oriented portfolios show great improvement relative to their benchmark. More risk-focused portfolios show an improvement with respect to their single-regime counterparts but show a similar performance to equally weighted portfolios.

The highest performing portfolio is Sharpe, that considers both mean and covariance in the construction process. Fig. 10-Top shows the annualized return as a function of annualized volatility for the Sharpe portfolios and their benchmarks. Portfolios built using HMMs show a higher return and less volatility than their unconditional counterpart, and higher return and volatility than the EQ portfolios. Fig. 10-Bottom shows the Sortino ratio for the same portfolios. We see that the HMM portfolios yield better performance than their benchmarks.

Table 3 shows different performance metrics averaged for each type of portfolio. In most cases, HMM-portfolios show better performance than their single-regime benchmarks on all metrics, and more return-oriented portfolios perform better than equally weighted ones. Performance improvement comes both from higher returns and risk reduction in return-oriented portfolios. Additionally, skewness and kurtosis are lower than benchmark returns and maximum drawdown is lower (and for a shorter period of time) in most cases.

6.3. DAA system with FSHMM

We then used the FSHMM algorithm to detect relevant features in our data set of 25 factor indices. To ensure the algorithm is

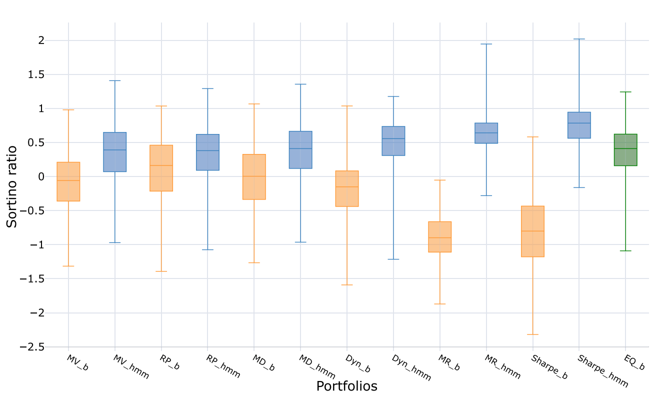

Fig. 9. Boxplots corresponding to the Sortino ratio for all portfolios calculated using a HMM (blue) and their benchmarks (orange) and an equally weighted portfolio (green).

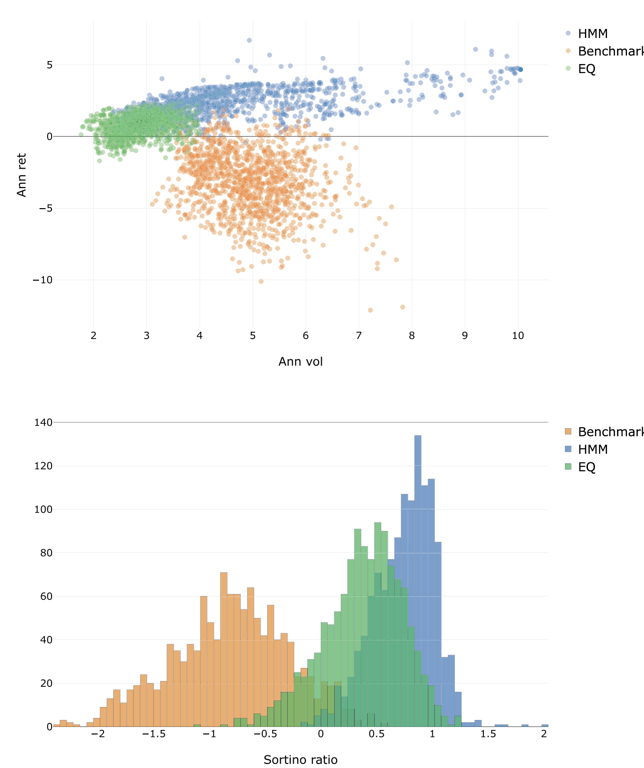

Fig. 10. Top plot shows annualized return as a function of annualized volatility for Sharpe portfolios built using HMM information (blue), Sharpe portfolios rebalanced monthly (orange) and EQ portfolios (green). Bottom plot corresponds to the Sortino distribution of the plots. All plots correspond to the test set (are out of sample).

indeed differentiating between relevant and irrelevant features, we tested it on feature vectors that consist of factor returns (relevant features) and random noise (irrelevant features). Tables 5 and 6 in C show the saliency of all features for 2 and 3-state models, for different lengths of the time series and two values of k. In all cases, irrelevant features are discarded (saliency values are close to zero) and when k is small, saliency of the relevant features is close to one.

Fig. 11 shows the feature saliencies of all factor return series for different values of k. As the training set has about 3800 observations, we chose values of k closer to a quarter of that number following the heuristic proposed in Adams et al. (2016). The selected features are: Book Value Yield, 1 Yr Fwd Earnings Yield, Sales Yield, 6 Month Price Momentum, 12 Month Price Momentum, EPSCV, Beta. This is of interest as the selected factors represent four of the six factor families mentioned in Section 3.2.

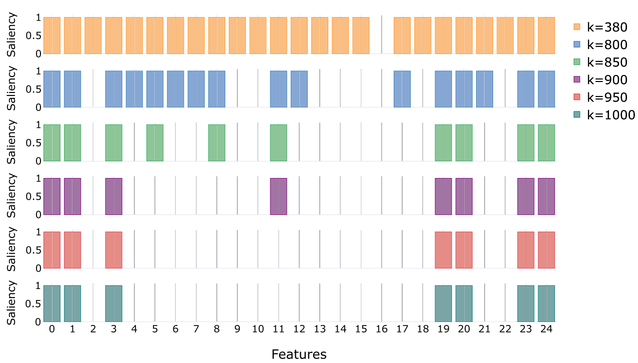

Fig. 11. Selected features in the training set ($T = 3800$ observations) of the 25 factor return series with different values of k. With small values of k all features are accepted. With $k \geqslant T/4$ the algorithm selects a relevant subset of features.

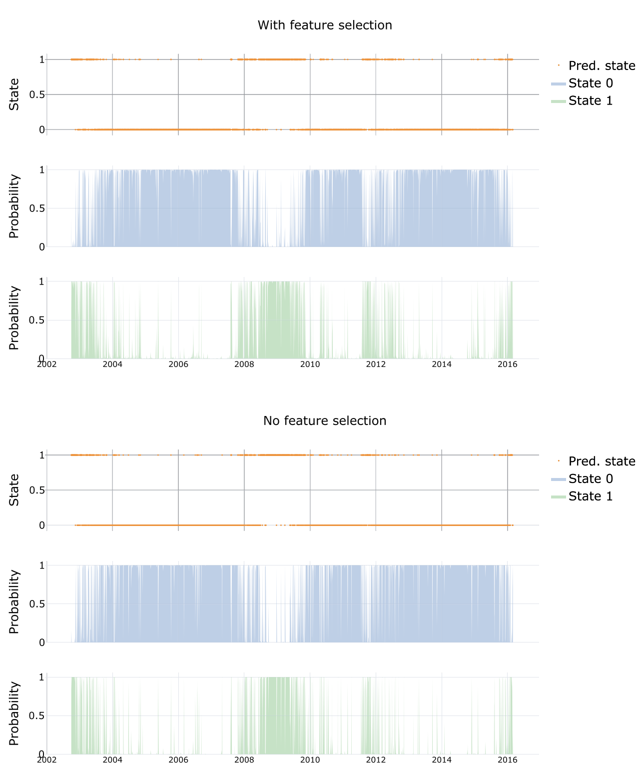

Fig. 12. Top plot corresponds to predicted state and state probabilities for the model trained with relevant features. Bottom plot corresponds to the HMM trained with all 25 features.

For comparison, we trained a HMM using all 25 features and a model trained with the selected assets. Fig. 12 shows the predicted state and estimated probabilities for the model after training; we identify state 1 as a "good state", and state 0 as a "bad state". The plots clearly identify the 2008 economic crisis – the first steps developed in August and September of 2007 with some episodes between January and May 2008 before the big crash in September 2008. Both models identify spikes of state 0 in the second half of 2007 and transition fully to state zero during 2008. The model trained with relevant features tends to be more sensitive to the distress state - it spends 24% of the time in this state versus 20% of the model trained with the full set of features. The average duration of state 0 is 3.8 days vs average length of 3.2 days of the full model. No smoothing was applied to the predicted probabilities to calculate these values.

6.4. FS-DAA system with MSCI indices

In this section we evaluate performance of the FS-DAA system using a subset of factors from the daily factor dataset after feature selection, and MSCI enhanced factors for allocation, and compare it with the DAA system without feature selection, that trains the HMM with all 25 factors from the dataset.

For simplicity we calculated only Sharpe, MR and Dyn portfolios, as they showed a significantly better performance when using a regime switching model in their construction than risk-focused portfolios and their benchmarks. Fig. 13 shows the cumulative return of these portfolios with a full feature HMM, FSHMM and

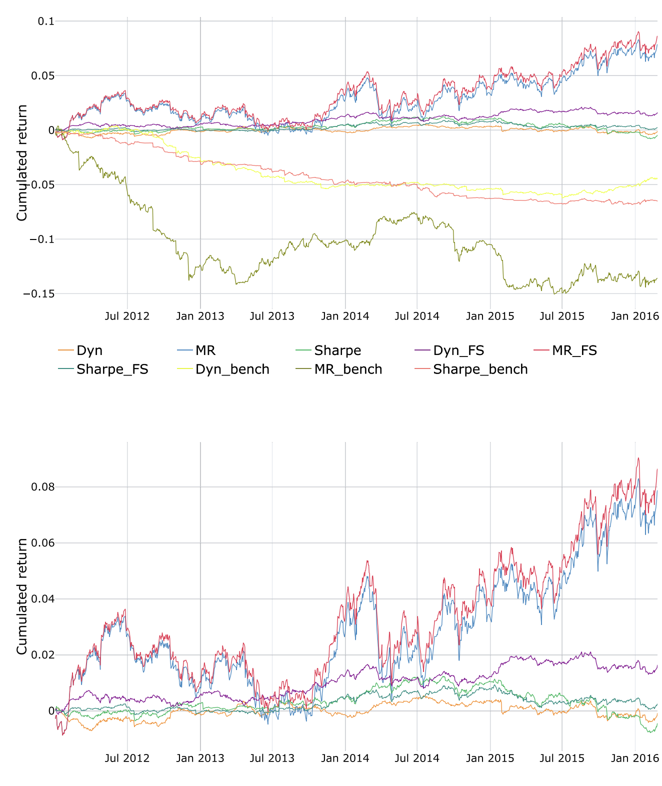

Fig. 13. Top plot corresponds to portfolios built using information from an HMM with feature saliency, portfolios built using information from an HMM with full features and their benchmarks. Both HMM portfolios accumulate higher returns than the benchmarks. Bottom plot shows that cumulative returns of FSHMM and fullHMM portfolios built using FS have a better performance. Returns are in excess of the market in USD, for the period Jan 2012–Feb 2016.

the benchmarks constructed without regime information. Both HMM portfolios perform better than their benchmarks (top plot) and portfolios constructed using an HMM with feature selection perform slightly better than portfolios built with a full feature HMM (bottom plot).

Metrics performance for all portfolios and for the MSCI enhanced indices net of market are shown in Table 4. All metrics are annualized and are out-of-sample, covering the period Jan-2012–Feb-2016. The results obtained using DAA and FS-DAA show a robust improvement with respect to their benchmarks. We see that only three MSCI indices have a positive IR in the period, and two of the three FSHMM portfolios show the highest IR in all cases. Reduction of downside risk is achieved in most cases that use either a full-feature HMM or a FSHMM with respect to their benchmarks and the MSCI indices.

7. Conclusions and future work

The main focus of the paper is improvement of smart beta strategies through the use of regime switching models. The main contributions from this work are:

1. We have shown that constructing a portfolio using information from a HMM with two latent states trained with the same assets that will be used for allocation, improves performance with respect to the same portfolio built with a single regime approach.
   We have tested this by calculating different types of portfolios, ranging from more risk focused to more aggressive. The improvement is more significant for return-oriented and balanced portfolios where return or risk-adjusted return is optimized achieving on average an information ratio of 50% annually in excess of market, and is less evident in risk-focused portfolios (Risk Parity, Minimum Variance and Maximum diversification) with an improvement on IR of 25% on average annually.
2. We have developed a systematic framework for asset allocation using an embedded feature selection algorithm to identify features of relevance to the model. This improves the model's accuracy and allows for a more objective approach to portfolio construction in the sense that it should help to prevent biases in the feature selection process which is normally done by a financial expert.
   We used a FSHMM algorithm to select relevant features from a pool of well known factor indices and compared it with a HMM trained with the whole set of assets. Both models showed agreement on regime identification, with the model trained using only relevant features being more sensitive to periods of economic distress.
3. We have tested both models using real, investable assets through MSCI USA enhanced factor indices. Portfolios constructed using information from the FSHMM trained with relevant features show a higher performance than the same portfolios constructed using a HMM trained with a full set of features.

An extension of the work to select relevant economic series could be to include macroeconomic series in the HMM, where the embedded feature selection could allow for a more precise identification of economic cycles. This could be of interest for other asset classes such as fixed income.

In this paper, the evaluation and verification of the proposed approach is mainly based on empirical performance of the smart beta investment return, as this is the ultimate goal and verification of any smart beta investment techniques. An interesting further research problem is to analyze and compute the standard errors, confidence intervals, and statistical properties of estimated FSHMM parameters by the proposed learning method, and therefore provide further reliability and robustness testing and verification of the SFHMM approach for smart beta investing.

A drawback of using HMMs is that the number of latent states has to be known in advance, or selected through BIC, which is not always effective, or to use a greedy approach to choose the model with higher performance. This could be addressed using an infinite HMM (Beal, Ghahramani, & Rasmussen, 2002).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Elizabeth Fons: Investigation, Formal analysis, Writing - original draft. **Paula Dawson:** Resources, Writing - review & editing, Supervision. **Jeffrey Yau:** Supervision. **Xiao-jun Zeng:** Writing - review & editing, Supervision, Funding acquisition. **John Keane:** Writing - review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors are grateful to Sahil Kahn, David Hutchins and Andrew Chin for their valuable feedback on early results of this work, and Erik Vynckier for the initial conceptualization of the project. This work was supported by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie Grant Agreement No. 675044 (http://bigdatafinance.eu/), Training for Big Data in Financial Research and Risk Management.

Appendix A. Hidden Markov Models (HMMs)

HMMs are sequential models that assume an underlying hidden process modeled by a Markov chain and a sequence of observed data as a noisy manifestation of this latent process (Murphy, 2012).

Given $o = \{\mathbf{y_1}, \ldots, \mathbf{y_T}\}$ the sequence of observed data where each $\mathbf{y_t} \in \mathbb{R}^L$ with L the dimension of observations and $\mathbf{x} = x_1, \ldots, x_T$ the latent sequence of states where $x_t \in \{1, \ldots, K\}$ with K the number of latent states. The Gaussian HMM model parameters are $\Lambda = (\pi, A, \mu, \sigma)$ where π and $\mathbf{A}$ correspond to the initial probability vector and transition probability matrix, and μ and σ are the mean vector and covariance matrix of the state dependent L-dimensional Gaussian feature distribution (generally called emission probabilities, symbolized here by b_{x_t}), the graphical model of the HMM can be seen in Fig. 14 where blue squares represent latent variables, orange circles are observations and green circles represent model parameters. The complete likelihood can be written as:

$$p(\mathbf{x}, \mathbf{y}|\Lambda) = \pi(x_0) b_{x_0}(y_0) \prod_{t=1}^{T} \mathbf{A}(x_{t-1}, x_t) b_{x_t}(y_t) \tag{A.1}$$

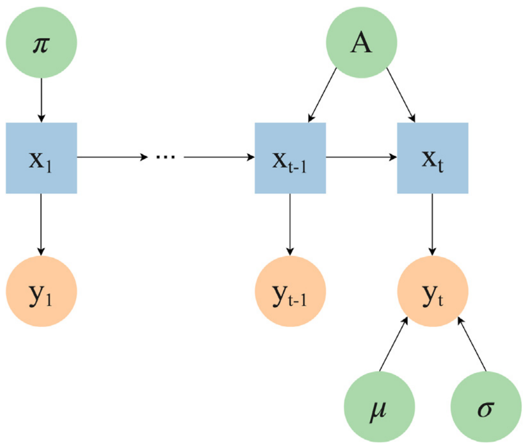

Fig. 14. HMM: squares with x_t represent latent variables, circles with y_t are observations and circles with $\{\pi, A, \mu, \sigma\}$ represent model parameters.

In this work the sequence of noisy observations are factor indices returns and the underlying hidden process is the state of the market that generates them. We assume that the emission probabilities are Gaussian. While normal distributions are a poor fit to financial returns, the mixture of normal distributions provide a much better fit capturing stylize behaviors including fat tails and skewness (Nystrup et al., 2015; Ang & Timmermann, 2012).

The training of HMMs is done by the Baum-Welch algorithm, a type of Expectation–Maximization (EM) algorithm (Rabiner, 1989). The E-step calculates the expected value of the log-likelihood with respect to the probability laws of the state, given the data and current model parameters and the M-step maximizes the expectation computed in the previous step to update the model parameters. The algorithm iterates between these two steps until convergence. The expectation of the complete log-likelihood function is given by:

$$Q(\Lambda, \Lambda\prime) = E[\log p(\mathbf{x}, \mathbf{y}|\Lambda)|\mathbf{y}, \Lambda\prime] \tag{A.2}$$

where Λ are the parameters for the current iteration and Λ' is the set of parameters from the previous iteration.

Following Adams et al. (2016), we place priors on the parameters and calculate the MAP estimate, so the Q function is modified by adding the prior on the model parameters, $G(\Lambda)$:

$$Q(\Lambda, \Lambda\prime) + \log G(\Lambda) \tag{A.3}$$

The EM algorithm is as follows, the Q function in A.2 is calculated in the E-step and the Eq. A.3 is maximized in the M-step.

Appendix B. Feature saliency HMM algorithm

The FSHMM algorithm as developed by Adams, Beiling and Cogill has the following EM update steps (for simplicity we follow their notation):

B.1. E-Step

$$\gamma_t(i) = P(x_t = i|y,\ \Lambda\prime) \tag{B.1}$$

$$\xi(i,j) = P(x_{t-1} = i,\ x_t = j|y,\ \Lambda\prime) \tag{B.2}$$

With $\gamma_t(i)$ and $\xi(i,j)$ calculated with the forward–backward algorithm. The additional updates are:

$$e_{ilt} = \rho_l r(y_{lt}|\mu_{il}, \sigma_{il}^2) \tag{B.3}$$

$$h_{ilt} = (1-\rho_l) q(y_{lt}|\epsilon_l, \tau_l^2) \tag{B.4}$$

$$g_{ilt} = e_{ilt} + h_{ilt} \tag{B.5}$$

$$u_{ilt} = \frac{\gamma_{it} e_{ilt}}{g_{ilt}} \tag{B.6}$$

$$v_{ilt} = \gamma_{it} - u_{ilt} \tag{B.7}$$

B.1.1. MAP M-step:

For the M-step, the following priors are used, where Dir corresponds to the Dirichlet distribution, $\mathcal{N}$ is the Gaussian distribution and IG is the inverse gamma distribution:

$$\pi \sim \mathrm{Dir}(\pi|\bar{\mathbf{p}}) \tag{B.8}$$

$$A_i \sim \mathrm{Dir}(A_i|\bar{\mathbf{a}}_i) \tag{B.9}$$

$$\mu_{il} \sim \mathcal{N}(\mu_{il}|m_{il}, s_{il}^2) \tag{B.10}$$

$$\sigma_{il}^2 \sim \mathrm{IG}(\sigma_{il}^2|\zeta_{il}, \eta_{il}) \tag{B.11}$$

$$\epsilon_l \sim \mathcal{N}(\epsilon_l|b_l, c_l^2) \tag{B.12}$$

$$\tau_l^2 \sim \mathrm{IG}(\tau_l|v_l, \psi_l) \tag{B.13}$$

$$\rho_l \sim e^{k_l \rho_l} \tag{B.14}$$

The parameter update equations are listed below:

$$\pi_i = \frac{\gamma_0(i) + \beta_i - 1}{\sum_{i=1}^{I}(\gamma_0(i) + \beta_i - 1)} \tag{B.15}$$

$$a_{ij} = \frac{\sum_{t=1}^{T}\xi_t(i,j) + \alpha_{ij} - 1}{\sum_{j=1}^{I}\left(\sum_{t=1}^{T}\xi_t(i,j) + \alpha_{i,j-1}\right)} \tag{B.16}$$

$$\mu_{il} = \frac{s_{il}^2 \sum_{t=0}^{T} u_{ilt} y_{lt} + \sigma_{il}^2 m_{il}}{s_{il}^2 \sum_{t=0}^{T} u_{ilt} + \sigma_{il}^2} \tag{B.17}$$

$$\sigma_{il}^2 = \frac{\sum_{t=0}^{T} u_{ilt}(y_{lt} - \mu_{il})^2 + 2\eta_{il}}{\sum_{t=0}^{T} u_{ilt} + 2(\zeta_{il} + 1)} \tag{B.18}$$

$$\epsilon_l = \frac{c_l^2 \sum_{t=0}^{T}\left(\sum_{i=1}^{I} v_{ilt}\right) y_{ilt} + \tau_l^2 b_l}{c_l^2 \sum_{t=0}^{T}\left(\sum_{i=1}^{I} v_{ilt}\right) + \tau_l^2} \tag{B.19}$$

$$\tau_l^2 = \frac{\sigma_{t=0}^{T}\left(\sum_{i=1}^{I} v_{ilt}\right)(y_{lt} - \epsilon_l)^2 + s\psi_l}{\sigma_{t=0}^{T}\left(\sum_{i=1}^{I} v_{ilt}\right) + 2(v_l + 1)} \tag{B.20}$$

$$\rho_l = \frac{\widehat{T} - \sqrt[2]{\widehat{T}^2 - 4k_l\left(\sum_{t=0}^{T}\sum_{i=1}^{I} u_{ilt}\right)}}{2k_l} \tag{B.21}$$

where $\widehat{T} = T + 1 + k_l$.

Appendix C. FSHMM with real and noise features

Table 5 shows feature saliency of 5 relevant features and three irrelevant features generated with $\mathcal{N}(0,1)$ with different number of observations and number of hidden states. Table 6 shows the same but with 10 relevant features and 5 added series of noise, for different states and values of k parameter.

References

Adams, S. (2015). Simultaneous Feature Selection and Parameter Estimation for Hidden Markov Models. Ph.D. thesis School of Engineering and Applied Science, University of Virginia.

Adams, S. & Beling, P. A. (2017). A survey of feature selection methods for gaussian mixture models and hidden markov models. Artificial Intelligence Review, Online First..

Adams, S., Beling, P. A., & Cogill, R. (2016). Feature selection for hidden markov models and hidden semi-markov models. *IEEE Access, 4*, 1642–1657.

Agather, R. & Gunthorp, P. (2017). Smart beta: 2017 global survey findings from asset owners. Technical Report FTSE Russell..

Ang, A. (2014). *Asset management: A systematic approach to factor investing*. Oxford University Press.

Ang, A., & Bekaert, G. (2003). *How do regimes affect asset allocation?. Working paper 10080*. National Bureau of Economic Research.

Ang, A., & Sorensen, M. (2012). Risks, returns, and optimal holdings of private equity: A survey of existing approaches. *Quarterly Journal of Finance (QJF), 02*, 1–27.

Ang, A. & Timmermann, A. (2012). Regime changes and financial markets. Annual Review of Financial Economics, 4..

Arnott, R., Beck, N., Kalesnik, V. & West, J. (2016). How Can "Smart Beta" Go Horribly Wrong?. Technical Report Research Affiliates..

Asness, C. S. (2016). The Siren Song of factor timing. *The Journal of Portfolio Management.*.

Axelsson, O. (2017). *Efficient risk factor allocation with regime based models*. Student Paper.

Bacon, C. (2012). *Practical risk-adjusted performance measurement. The Wiley finance series*. Wiley.

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*, 450–458.

Baum, L. E., Petrie, T., Soules, G., & Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *Annals of Mathematical Statistics, 41*, 164–171.

Beal, M. J., Ghahramani, Z. & Rasmussen, C. E. (2002). The infinite hidden markov model. In Machine learning (pp. 29–245). MIT Press..

Bender, J., Briand, R., Melas, D. & Subramanian, R. A. (2013). Foundations of factor investing..

Bulla, J., & Bulla, I. (2006). Stylized facts of financial time series and hidden semi-markov models. *Computational Statistics & Data Analysis, 51*, 2192–2209.

Bulla, J., Mergner, S., Bulla, I., Sesboue, A., & Chesneau, C. (2011). Markov-switching asset allocation: Do profitable strategies exist? *Journal of Asset Management, 12*, 310–321.

Dabrowski, J. J., Beyers, C., & de Villiers, J. P. (2016). Systemic banking crisis early warning systems using dynamic bayesian networks. *Expert Systems with Applications, 62*, 225–242.

Dias, J., & Ramos, S. (2014). Dynamic clustering of energy markets: An extended hidden markov approach. *Expert Systems with Applications, 41*, 7722–7729.

Dimson, E., Marsh, P., & Staunton, M. (2017). Factor-based investing: The long-term evidence. *The Journal of Portfolio Management, 43*, 15–37.

Elliott, R. J., Siu, T. K., & Fung, E. S. (2014). A double hmm approach to altman z-scores and credit ratings. *Expert Systems with Applications, 41*, 1553–1560.

Erlwein, C., Mamon, R., & Davison, M. (2011). An examination of hmm-based investment strategies for asset allocation. *Applied Stochastic Models in Business and Industry, 27*, 204–221.

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics, 116*, 1–22.

Guidolin, M. (2012). *Markov switching models in empirical finance*. Working Paper.

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

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica, 57*, 357–384.

Hamilton, J. D. (2010). Regime switching models. In L. E. Durlauf, N. Steven & Blume (Eds.), Macroeconometrics and time series analysis (pp. 202–209). London: Palgrave Macmillan UK..

Kim, C. -J. & Nelson, C. (1999). State-space models with regime switching: Classical and Gibbs-sampling approaches with applications (Vol. 1). (1st ed.). The MIT Press..

Liu, P. P., Xu, K., & Zhao, Y. (2011). Market regimes, sectorial investments, and time-varying risk premiums. *International Journal of Managerial Finance, 7*, 107–133.

Ma, Y., Maclean, L., Xu, K., & Zhao, Y. (2011). A portfolio optimization model with regime-switching risk factors for sector exchange traded funds. Pacific Journal of Optimization, 7..

Mulvey, J. M., & Liu, H. (2016). Identifying economic regimes: Reducing downside risks for university endowments and foundations. *The Journal of Portfolio Management, 43*, 100–108.

Murphy, K. P. (2012). Machine learning: A probabilistic perspective. The MIT Press..

Nguyen, N. (2017). An analysis and implementation of the hidden Markov model to technology stock prediction. *Risks, 5*, 62–78.

Nystrup, P., Hansen, B. W., Larsen, H. O., Madsen, H., & Lindstrom, E. (2017). Dynamic allocation or diversification: A regime-based approach to multiple assets. *The Journal of Portfolio Management, 44*, 62–73.

Nystrup, P., Hansen, B. W., Madsen, H., & Lindstrom, E. (2016). Detecting change points in vix and s&p 500: A new approach to dynamic asset allocation. *Journal of Asset Management, 17*, 361–374.

Nystrup, P., Madsen, H., & Lindstrom, E. (2015). Stylised facts of financial time series and hidden Markov models in continuous time. *Quantitative Finance, 15*, 1531–1541.

Nystrup, P., Madsen, H., & Lindstrom, E. (2017a). Dynamic portfolio optimization across hidden market regimes. *Quantitative Finance, 18*, 83–95.

Nystrup, P., Madsen, H., & Lindstrom, E. (2017b). Long memory of financial time series and hidden markov models with time-varying parameters. *Journal of Forecasting, 36*, 989–1002.

Petropoulos, A., Chatzis, S. P., & Xanthopoulos, S. (2016). A novel corporate credit rating system based on student's-t hidden markov models. *Expert Systems with Applications, 53*, 87–105.

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

Reus, L., & Mulvey, J. M. (2016). Dynamic allocations for currency futures under switching regimes signals. *European Journal of Operational Research, 253*, 85–93.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*, 461–464.

Zhou, N., & Mamon, R. (2012). An accessible implementation of interest rate models with markov-switching. *Expert Systems with Applications, 39*, 4679–4689.