Forecasting corporate credit spreads: regime-switching in LSTM

Christina Erlwein-Sayer[1], Stefanie Grimm[2], Alexander Pieper[1], Rümeysa Alsaç[2]

[1] *Department IV, Financial Mathematics, University of Applied Sciences HTW Berlin, Berlin, Germany*
[2] *Department of Financial Mathematics, Fraunhofer Institute for Industrial Mathematics ITWM, Kaiserslautern, Germany*

Abstract

Corporate credit spreads are modelled through a Hidden Markov model (HMM) which is based on a discretised Ornstein-Uhlenbeck model. We forecast the credit spreads within this HMM and filter out state-related information hidden in the observed spreads. We build a long short-term memory recurrent neural network (LSTM) which utilises the regime-switching information as a feature to predict the change of the credit spread. The performance of the LSTM is analysed and compared to the accuracy of an LSTM without the regime-switching information. Furthermore, purely utilising the HMM forecast, the prediction of the credit spread is compared to the prediction within the LSTM. The HMM-LSTM model is calibrated on corporate credit spreads from three European countries between 2004 and 2019. Our findings show that in most cases the LSTM performance can be improved when regime information is added.

Keywords: Long Short-Term Memory, Artificial Neural Networks, Hidden Markov Models, Filtering, Regime-switching model, Credit Spread, Forecasting

1 Introduction

Modelling corporate credit spreads is vital to understand and predict risks in the Fixed Income market as well as to understand macroeconomic dynamics. In recent years, studies have investigated modelling and predicting credit spreads by including macroeconomic information, see e.g. [Clark and Baccar (2018)]. Credit spreads are likely to widen when financial risk increases. Various macroeconomic factors such as inflation or a countries' GDP growth rate are taken into consideration, interconnections between credit spreads and the economy are present, see [Bernoth and Erdogan (2012)] or [Maltritz (2012)] amongst others. More recently, the influence of macroeconomic news on sovereign bond spreads has been studied, see [Erlwein-Sayer (2018)]. Furthermore, common and firm-specific factors from corporate credit spreads were investigated for predicting economic activity in [Kobayashi (2021)]. The author highlights that nonlinear models could be established in order to capture nonlinearities and switching regimes of risk factors in credit spreads. Recently, [Klein and Pliszka (2018)] found, that risk factors for corporate credit spreads change over time in European markets. In turbulent market times, systematic risk factors have a higher impact on spreads. Our approach takes these findings under consideration and investigates further, how information on regime shifts in corporate credit spread dynamics can be incorporated into statistical learning algorithms.

In stock markets, statistical learning methods haven been introduced in various studies. In stock price prediction for example, machine learning techniques are often used to forecast the direction of the change, see e.g. [Persio and Honchar (2016)] and [Roondiwala et al. (2017)]. In [Siami-Namini and Siami Namin(2018)] a Long Short-Term Memory (LSTM) model has been found to outperform a classical ARIMA model for return predictions. Also when compared to other machine learning techniques, LSTM performs well when forecasting financial time series (e.g. [Yan and Ouyang (2018)]). Statistical learning for credit market modelling also becomes increasingly important. In credit risk modelling, machine learning techniques have been applied to model credit default probabilities [Papouskova and Hajek (2019)], to calibrate credit risk models [Manzo and Qiao (2020)] or to assess corporate credit rating based on financial statement reports, e.g. [Golbayani et. al (2020)]. Artificial Neural Networks (ANN) as well as Decision Trees or Support Vector Machines (SVM) have been built to include various features, that help to predict credit spread dynamics. A recent paper by Xiong et al [Xiong et al. (2016)] investigates modelling corporate bond spreads through different types of machine learning techniques. In their study, neural network approaches, especially a LSTM model, lead to the highest accuracy in forecasting the direction of change, beating Random Forest and Bayesian Additive Regression Trees. The authors state that the forget gate of the LSTM enables the model to make accurate predictions even in changing states.

In general, when modelling financial time series, the actual state of the market, e.g. whether the market is in a bull or bear regime or whether it passes a high volatility period, has an effect on the dynamics of the time series. The state of the market is often not observable and therefore treated as a latent variable. This hidden state of the market can be modelled through a Hidden Markov model (HMM). Work on HMMs for financial time series goes back to Hamilton [Hamilton (1989)] and has evolved into a widely applied concept for modelling in asset allocation, portfolio optimisation and risk management (see [Mamon and Elliott (2007)], [Sass and Haussmann (2004)] or [Nguyen (2018)], amongst others). Furthermore, modelling of interest rates and bonds in an HMM framework has been studied by various authors, see amongst others [Landen (2000)] and [Thomas et al. (2002)]. An HMM based on an Ornstein-Uhlenbeck (OU) process has been developed by [Erlwein and Mamon (2009)] and [Grimm et al. (2020)] in a discrete-time and continuous-time model respectively. We will base our HMM approach for corporate credit spreads in this work on filters and adaptive parameter estimation derived in [Erlwein and Mamon (2009)] as follows.

In this paper, we model the credit spreads within a Hidden Markov model (HMM) based on a discretised OU process which is designed to capture hidden states of the underlying market and adapt the parameter estimates according to the state estimation. We compare the forecast of the credit spreads obtained with the OU-HMM to a one-step ahead forecast calculated by an LSTM. The forecast errors are investigated and compared. In a third step, we combine both HMM and LSTM. This is achieved by using the estimated state of the Markov chain as an additional feature in the LSTM. Furthermore we include the state of the economy of the country, where the credit spread is emitted. The country state is estimated by applying the HMM on a mean country corporate spread. A daily mean spread is calculated over all available credit spreads in a country on a given day. The estimated underlying economic state is then included in the LSTM.

The contribution of our paper is two-fold: we develop and test models to forecast corporate credit spreads under changing market conditions. Our underlying data sets are comprised of data from three different European countries, making this, to the best of our knowledge, the first time series modelling study on European corporate bond spreads. We investigate the accuracy of the one-step ahead point forecast of an HMM and LSTM and improve it by merging the two models, therefore adding state information to our neural network. Our HMM feature for LSTM gives state information to the neural network, which enhances the predictability of the corporate credit spread.

2 Modelling Credit Spreads with Regime Shifts: Model Set-Up

To capture the market dynamics in credit spreads, our model parameters are based on latent market states. The stylized facts of corporate credit spreads in European markets are similar to those of short-term interest rates. We observe mean-reverting credit spreads with changing volatilities over time. Periods of low volatility are followed by periods of high volatility. The mean-reversion level and speed can differ across market states. Regime-switching models for interest rates have been studied by [Gray (1996)] and [Ang and Bekaert(2002)], amongst others. In a review on regime changes in financial markets by [Ang and Timmermann(2012)] the authors note that regime changes in interest rates are often linked to monetary policies and depend on real rates and inflation expectations. We base our model for credit spreads on the Vasicek model, which has been widely used to model short term interest-rates. Our HMM is based on a discretised version of the Ornstein-Uhlenbeck-process, see [Erlwein and Mamon (2009)] for details.

The number of states in the HMM is assumed to be given. Typically, two to three states are sufficient to capture the business cycles in the financial markets. The number of states were investigated in studies on asset allocation, see e.g. [Guidolin and Timmermann (2007)] and for bonds and interest rates, see [Ang and Timmermann(2012)].

In the following, let N denote the number of states. Let $(\Omega, \mathcal{F}, P)$ be the underlying probability space of the homogeneous Markov chain $\mathbf{x}$ with finite N-dimensional state space in discrete time. We assume the initial distribution of $\mathbf{x}_0$ to be known. Without loss of generality, the state space of $\mathbf{x}$ can be identified with the canonical basis $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_N\}$ of $\mathbb{R}^N$. Under the real world probability measure P, the Markov chain $\mathbf{x}$ has dynamics

$$\mathbf{x}_{k+1} = \mathbf{\Pi}\mathbf{x}_k + \mathbf{v}_{k+1} \tag{1}$$

where $\mathbf{v}_{k+1}$ is a martingale increment. The transition probability matrix is $\mathbf{\Pi} = (\pi_{ij}) \in \mathbb{R}^{N\times N}$ with $\pi_{ij} = P(\mathbf{x}_{k+1} = \mathbf{e}_j | \mathbf{x}_k = \mathbf{e}_i)$.

The credit spread c follows the stochastic process

$$dc_t = a(\mathbf{x}_t)[\beta(\mathbf{x}_t) - r_t]\, dt + \xi(\mathbf{x}_t)\, dW_t \tag{2}$$

for $r_0 \geq 0$ with $a(\mathbf{x}_t) = \langle \mathbf{a}, \mathbf{x}_t \rangle$, $\beta(\mathbf{x}_t) = \langle \boldsymbol{\beta}, \mathbf{x}_t \rangle$ and $\xi(\mathbf{x}_t) = \langle \boldsymbol{\xi}, \mathbf{x}_t \rangle$, where $\langle \cdot, \cdot \rangle$ is the usual Euclidean scalar product, $W = \{W_t : 0 \leq t \leq T\}$ is a Wiener process. All three parameters are governed by a Markov chain, which ensures, that the model parameters switch between market regimes. This Ornstein-Uhlenbeck process is chosen to capture the mean-reverting dynamics observed in corporate credit spreads over time. Here, a_t denotes the rate of mean-reversion, β_t is the long-term mean of the process and ξ_t is the volatility.

Our observation process $\{\mathbf{y}_{k+1},\ 0 \leq k \leq K-1\}$ is then a discrete representation of the credit spread based on the discretised solution of Equation 2.

$$y_{k+1} = \alpha(\mathbf{x}_k) y_k + \gamma(\mathbf{x}_k) + \eta(\mathbf{x}_k) z_{k+1} \tag{3}$$

where we set $\alpha(\mathbf{x}_k) = e^{-a(\mathbf{x}_k)\Delta t_{k+1}}$, $\gamma(\mathbf{x}_k) = \beta(\mathbf{x}_k)(1 - e^{-a(\mathbf{x}_k)\Delta t_{k+1}})$, $\eta(\mathbf{x}_k) = \xi(\mathbf{x}_k)\sqrt{\frac{1-e^{-2a(\mathbf{x}_k)\Delta t_{k+1}}}{2a(\mathbf{x}_k)}}$ and $\Delta t_{k+1} = t_{k+1} - t_k$. Here, $\{z_k\}$ is a sequence of independent identically distributed (i.i.d) standard normal random variables. The model parameters α, γ and η are governed by the Markov chain. The filtrations generated by the processes are defined by $\mathcal{F}^y = \sigma(y_1, y_2, ...)$, $\mathcal{F}^{\mathbf{x}} = \sigma(\mathbf{x}_1, \mathbf{x}_2, ...)$ and $\mathcal{G} = \mathcal{F}^y \vee \mathcal{F}^{\mathbf{x}}$.

The parameter estimation and filtering of states of the Markov chain is based on a filtering techniques by Elliott [Elliott et al. (1995)]. It involves adaptive filtering of hidden state information as well as parameter estimation based on these adaptive filters. It utilizes the Baum-Welch and the EM-algorithm. The adaptive filters and optimal parameter estimates are derived in a previous work, see [Erlwein and Mamon (2009)] for details. The derivation involves a change of measure from a reference probability measure $\bar{P}$ under which observation process and hidden Markov chain are independent. The real world measure P is derived through the Radon-Nikodym derivative $dP/d\bar{P}\big|_{\mathcal{G}_t} = \Lambda_t$ with

$$\begin{aligned}
\bar{\Lambda}_l &= \prod_{k=1}^{l} \bar{\lambda}_k \\
\bar{\lambda}_l &= \exp\Bigg[-\frac{\langle \boldsymbol{\alpha}, \mathbf{x}_{l-1}\rangle y_{l-1} + \langle \boldsymbol{\gamma}, \mathbf{x}_{l-1}\rangle}{\langle \boldsymbol{\eta}, \mathbf{x}_{l-1}\rangle} \cdot \frac{y_l}{\langle \boldsymbol{\eta}, \mathbf{x}_{l-1}\rangle} \\
&\qquad - \frac{(\langle \boldsymbol{\alpha}, \mathbf{x}_{l-1}\rangle y_{l-1} + \langle \boldsymbol{\gamma}, \mathbf{x}_{l-1}\rangle)^2}{2\langle \boldsymbol{\eta}, \mathbf{x}_{l-1}\rangle^2} \Bigg]
\end{aligned}$$

2.1 Filter Equations and Parameter Estimation

For the estimation of the model parameters we utilise adaptive filters for processes of the Markov chain. We do not only need filters for the state of the Markov chain, but also filters for the number of jumps from one state to another, the occupation time in each state and filters for three additional related processes that come without an obvious interpretation.

By Bayes' theorem, and with $\bar{\mathbb{E}}$ the expectation under $\bar{P}$, for any $\{\mathcal{G}_k\}$-adapted process $\{H_k\}$, the conditional expectation of H_k given $\mathcal{Y}_k$ is given by

$$\hat{H}_k = \mathbb{E}[H_k \mid \mathcal{Y}_k] = \bar{\mathbb{E}}[\Lambda_k \, H_k \mid \mathcal{Y}_k] \Big/ \bar{\mathbb{E}}[\Lambda_k \mid \mathcal{Y}_k] \,.$$

We abbreviate the unnormalized conditional expectation $\bar{\mathbb{E}}[\Lambda_k \, H_k \mid \mathcal{Y}_k]$ by $\sigma_k(H_k)$.

To relate $\sigma_k(H_k)$ and $\sigma_k(H_k\mathbf{x}_k)$ we note that $\langle \mathbf{1}, \mathbf{x}_k\rangle = 1$. The optimal (normalized) filter $\hat{H}_k := \sigma_k(H_k)$ for H_k is then given by

$$\langle \mathbf{1}, \sigma_k(H_k\mathbf{x}_k)\rangle = \sigma_k(H_k\langle \mathbf{1}, \mathbf{x}_k\rangle) = \sigma_k(H_k). \tag{4}$$

We find a recursive relation for $\sigma_k(H_k\mathbf{x}_k)$ for a scalar $\mathcal{G}-$adapted process $H_l = H_{l-1} +$

$a_l + \langle \mathbf{b}_l, \mathbf{v}_l \rangle + g_l f(y_l)$ with a, b and g are $\mathcal{G}$-predictable and f is a scalar-valued function:

$$\begin{aligned} \sigma_k(H_k \mathbf{x}_k) &= \sum_{i=1}^{n} \Gamma^i(y_k) \big[\langle \mathbf{e}_i, \sigma_{k-1}(H_{k-1}\mathbf{x}_{k-1}) \rangle \mathbf{\Pi} \mathbf{e}_i \\ &\quad + \langle \mathbf{e}_i, \sigma_{k-1}(a_k \mathbf{x}_{k-1}) \rangle \mathbf{\Pi} \mathbf{e}_i \\ &\quad + (\mathrm{diag}(\mathbf{\Pi}\mathbf{e}_i) - \mathbf{\Pi}\mathbf{e}_i \otimes \mathbf{\Pi}\mathbf{e}_i) \sigma_{k-1}(b_k \langle \mathbf{e}_i, \mathbf{x}_{k-1} \rangle) \\ &\quad + \sigma_{k-1}(g_k \langle \mathbf{e}_i, \mathbf{x}_{k-1} \rangle) f(y_k) \mathbf{\Pi}\mathbf{e}_i \big] \end{aligned} \tag{5}$$

with

$$\Gamma^i(y_l) = \exp\left[-\frac{\alpha_i y_{l-1} + \gamma_i}{\eta_i} \cdot \frac{y_l}{\eta_i} - \frac{(\alpha_i y_{l-1} + \gamma_i)^2}{2\eta_i^2} \right] \tag{6}$$

where $\otimes$ denotes the tensor product of vectors in $\mathbb{R}^n$ and $\mathrm{diag}(B)$ is a diagonal matrix B with $(b_1, b_2, ..., b_n)'$ in the diagonal.

We obtain

$$\sigma_k(\mathbf{x}_k) = \sum_{i=1}^{n} \Gamma^i(y_k) \langle \mathbf{e}_i, \sigma_{k-1}(\mathbf{x}_{k-1}) \rangle \mathbf{\Pi} \mathbf{e}_i \quad . \tag{7}$$

Filters needed for the optimal parameter estimates are the jump process J, the occupation time process O and auxiliary processes T_y, $T_{y_k^2}$ und T_{y_{k+1}, y_k}. Let J_k^{dc} represent the number of jumps of $\mathbf{x}_k$ from state $\mathbf{e}_c$ to state $\mathbf{e}_d$ in time k. So,

$$J_k^{dc} = J_{k-1}^{dc} + \langle \mathbf{x}_{k-1}, \mathbf{e}_d \rangle \pi_{dc} + \langle \mathbf{x}_{k-1}, \mathbf{e}_c \rangle \langle \mathbf{v}_k, \mathbf{e}_d \rangle \quad . \tag{8}$$

We derive following filter for J_k^{dc}:

$$\begin{aligned} \sigma_k(J_k^{dc} \mathbf{x}_k) &= \sum_{i=1}^{n} \Gamma(y_k) \langle \sigma_{k-1}(J_{k-1}^{dc} \mathbf{x}_{k-1}), \mathbf{e}_i \rangle \mathbf{\Pi} \mathbf{e}_i \\ &\quad + \Gamma^c(y_k) \sigma_{k-1}(\langle \mathbf{x}_{k-1}, \mathbf{e}_r \rangle) \pi_{dc} \mathbf{e}_d \quad . \end{aligned} \tag{9}$$

Let O_k^c denote the occupation time of the Markov process $\mathbf{x}$ in state e_c up to time k. Then,

$$O_k^c = O_{k-1}^c + \langle \mathbf{x}_{k-1}, \mathbf{e}_c \rangle \tag{10}$$

and we obtain the filter equation

$$\begin{aligned} \sigma_k(O_k^c \mathbf{x}_k) &= \sum_{i=1}^{n} \Gamma^i(y_k) \langle \sigma_{k-1}(O_{k-1}^c \mathbf{x}_{k-1}), \mathbf{e}_i \rangle \mathbf{\Pi} \mathbf{e}_i \\ &\quad + \Gamma^c(y_k) \langle \sigma_{k-1}(\mathbf{x}_{k-1}), \mathbf{e}_c \rangle \mathbf{\Pi} \mathbf{e}_c \quad . \end{aligned} \tag{11}$$

For auxiliary processes define the process $T_k^c(f)$ as

$$\begin{aligned} T_k^c(f) &:= \sum_{l=1}^{k} \langle \mathbf{x}_{l-1}, \mathbf{e}_c \rangle f(y_l) \\ &= T_{k-1}^c(f) + \langle \mathbf{x}_{k-1}, \mathbf{e}_c \rangle f(y_k) \end{aligned} \tag{12}$$

where f is a function of the form $f(y) = y$, $f(y) = y^2$ or $f(y) = y_{l+1}y_l, 1 \leq l \leq k$. The adaptive filters are then

$$\begin{aligned} \sigma_k(T_k^c(f)\mathbf{x}_k) &= \sum_{i=1}^{n} \Gamma^i(y_k)\{\langle \sigma_{k-1}(T_{k-1}^c(f)\mathbf{x}_{k-1}), \mathbf{e}_i \rangle \mathbf{\Pi}\mathbf{e}_i \\ &\quad + \Gamma^c(y_k)\langle \sigma_{k-1}(\mathbf{x}_{k-1}), \mathbf{e}_c \rangle f(y_k)\mathbf{\Pi}\mathbf{e}_c \quad . \end{aligned} \tag{13}$$

The respective optimal normalized filters $\hat{\mathbf{x}}_k$, $\hat{O}_k^c$, $\hat{\mathcal{J}}_k^{dc}$, $\hat{T}_k^c(f(y))$ for $\mathbf{x}_k$, O_k^c, J_k^{dc}, $T_k^c(f(y))$ are obtained from equations (9) to (13) by applying equation (4).

To estimate the model parameters for the observation process $y_{k+1} = \alpha(\mathbf{x}_k)y_k + \gamma(\mathbf{x}_k) + \eta(\mathbf{x}_k)z_{k+1}$ and $\mathbf{\Pi}$, we apply the EM-algorithm and plug in the adaptive filter estimates of the processes of the Markov chain. For a derivation and further explanations see [Erlwein and Mamon (2009)].

The set of parameters ρ, which determines our model is $\rho = \{\pi_{ji}, \alpha_i, \gamma_i, \eta_i, 1 \leq i, j \leq N\}$. Initial values of these parameters are assumed to be given.

Lemma 2.1. *Let $\hat{J}$, $\hat{O}$ and $\hat{T}(f)$ denote the filtered values of the processes J, O and $T(f)$ (equations (9) to (13)). The optimal parameter estimates $\hat{\pi}_{ji}, \hat{\alpha}_i, \hat{\gamma}_i, \hat{\eta}_i$ are given by*

$$\begin{aligned} \hat{\pi}_{ji} &= \frac{\hat{J}_k^{ji}}{\hat{O}_k^i} \\ \hat{\alpha}_i &= \frac{\hat{T}_k^i(y_{k+1}, y_k) - \hat{T}_k^i(y)\hat{\gamma}_i}{\hat{T}_k^i(y^2)} \\ \hat{\gamma}_i &= \frac{\hat{T}_{k+1}^i(y) - \hat{T}_k^i(y)\hat{\alpha}_i}{\hat{O}_k^i} \\ \hat{\eta}_i &= \sqrt{\frac{\hat{T}_{k+1}^i(y^2) + \hat{\alpha}_i^2\hat{T}_k^i(y^2) + \hat{\gamma}_i^2\hat{O}_k^i - 2\hat{\alpha}_i\hat{T}_k^i(y_{k+1}, y_k)}{\hat{O}_k^i} - \frac{2\hat{\gamma}_i\hat{T}_{k+1}^i(y) + 2\hat{\alpha}_i\hat{\gamma}_i\hat{T}_k^i(y)}{\hat{O}_k^i}} \end{aligned}$$

Parameters are updated after each batch of m data points. Recursive filters are obtained for all time steps, enabling us to retrieve adaptive parameter estimates whenever new information of the time series is available. Our algorithm is therefore run batchwise through time, therefore combining batch-updates with on-line estimation. The batch size m is set taking time-series characteristics into consideration. This includes amongst others the frequency or signal to noise ratio. For details on building the batchwise online estimation see [Erlwein, Grimm et al (2020)].

A typical plot of estimated parameters from a time series in our data set is depicted in Figure 1. We filter the time series of daily credit spreads over a time of six years between 2014 and 2019. The plots depict one credit spread time series of a company in France. The credit spread is analysed by the HMM and the optimal parameter values are estimated in batches of 30 data points, therefore the estimates are updated roughly on a monthly basis. The parameter estimates are utilised to calculate a one-step ahead forecast of the corporate credit spread. The estimated Markov chain in a 2-state set-up plotted over the actual modelled time series can be seen in the last graph of the figure. The probability estimates of the Markov chain react to changes in the spread. High or low levels of the time series are mirrored in high or low probabilities of a given state.

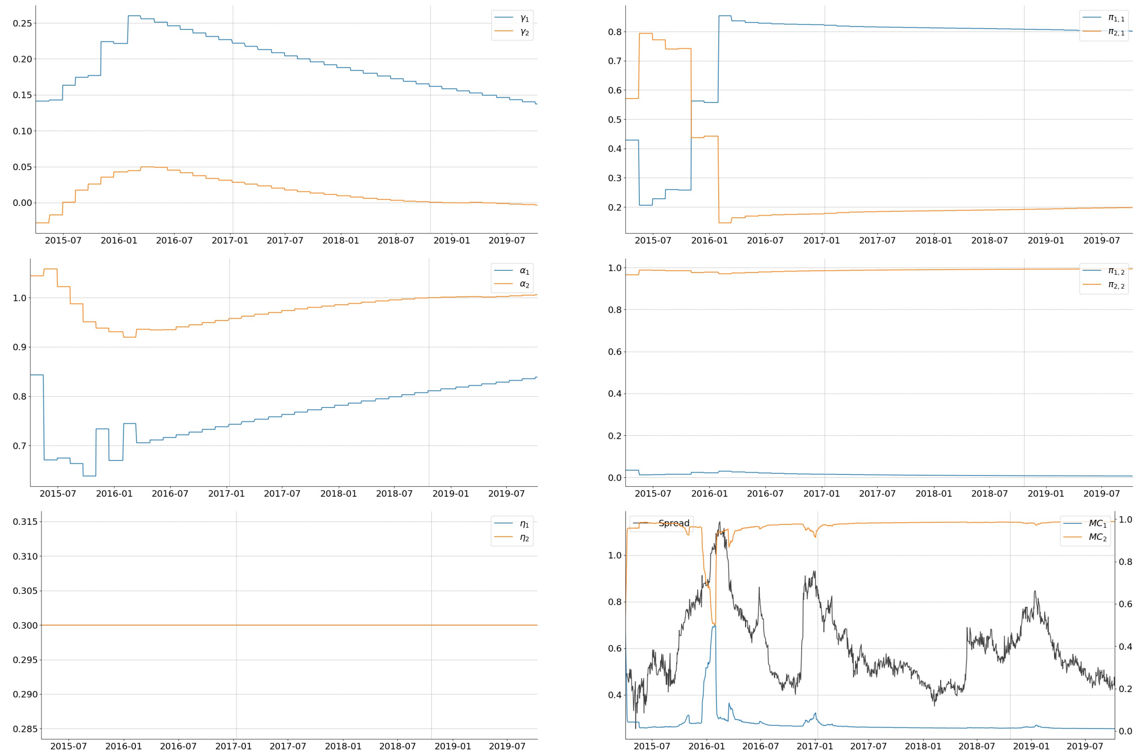

Figure 1: Adaptive parameter estimation and estimated Markov chain for a corporate credit spread in Spain

In our analysis, a 2-state HMM is here able to capture the real-world dynamics of corporate credit spreads in Europe over the last decade. One can see, that the Markov chain switches states at the beginning of 2016 when the credit spread decreases. The regime shift, which is depicted in the last graph of Figure 1 mirrors a shift in the mean of the time series. The estimated state of the Markov chain, which is updated in each time step going forward through the time series, captures the shifting market state.

The mean spread series for each country is also analysed through the HMM. In Figure 2 the credit spreads and estimated Markov chain for France, Germany and Spain are shown.

Figure 2: Estimated Markov chains for mean corporate credit spreads in France, Germany and Spain

Information which is obtained through on-line estimates of the credit spread series shall now be utilised to improve a further model for forecasting financial time series, namely a Long-Short-Term-Memory (LSTM) model.

3 Data

Our modelling approach is applied on a data set, which includes corporate bond spreads from three European countries over a period of sixteen years. We examine corporate bond spreads of France, Germany and Spain within a period between the end of 2003 and 2019. The corporate spreads are retrieved from Thomson Reuters Eikon.
The spreads in each country set are from companies which are or were at the time of retrieval constituents of the main countries' indices, namely from CAC 40, DAX and

IBEX 35. The time period of the corporate bond time series includes both bull and bear markets and covers the European sovereign credit spread crisis around 2012. Effects on the corporate spreads of this crisis are visible.
In all countries, various sectors are considered. Analysed sectors include Financial, Telecommunications, Utilities, Industrials, Energy, Services and Healthcare. We distinguish between long-and short-term corporate bonds, where bonds are classified as long-term with a duration of ten years and more.

In Table 1 the data set of credit spreads is sorted by countries and summary statistics for spreads in three time periods are depicted. We see that the second period, covering spreads between 2011 and 2015, has a higher mean and higher standard deviation than the period before and after. This period covers the European credit crisis, which lead to rising credit spreads and higher risks. Note, that this effect is most prominent in the corporate spreads of Spanish companies. The effect of economic fundamentals in this time period is captured in the high yields. The study by Afonso et al [Afonso et al. (2015)] investigates and underlines these effects on government yields in times of crisis. In addition, our dataset reflects these econometric dynamics for corporate credit spreads.

	Period 1: 2003-09-21 - 2011-01-01			**Period 2:** 2011-01-01 - 2015-01-01		
	ESP	GER	FRA	ESP	GER	FRA
number of time series	13	3	4	50	50	50
min	-0.734	0.143	0.237	0.28	-3.309	-2.122
mean	0.884	0.954	0.622	2.42	1.136	1.339
max	4.708	3.205	1.376	12.823	5.493	7.002
std	0.56	0.317	0.095	0.798	0.207	0.417

	Period 3: 2015-01-01 - 2019-10-22		
	ESP	GER	FRA
number of time series	50	50	50
min	0.105	0.003	-3.332
mean	0.891	1.188	0.879
max	5.991	5.59	5.051
std	0.266	0.329	0.242

Table 1: Summary statistics of corporate credit spreads, grouped by country and time period

The used packages are pandas, numpy, sklearn, tensorflow, math, matplotlib, time and os in Python and data.table in R for our implementation.

4 LSTM for modelling credit spreads

Beyond further development of classical statistical models, there is an expanding popularity of machine learning algorithms to model and forecast financial time series. Especially neural networks have gained in popularity. However, some characteristics of time series as for example autoregressive coherences, can hardly be covered by basic neural networks. Thus, specially designed deep learning algorithms called long short-term memory networks (LSTM), proposed in [Hochreiter and Schmidhuber (1997)], are of growing interest in recent years [Mehtab and Sen (2020)]. Those artificial recurrent neural networks make use of sequences of data instead of single data points. They have proven successful with time series problems due to their versatility and great efficiency to remember information from the time series history in the long and short term. For each time step, the LSTM cell takes three different inputs: the current input data, the short-term memory from the previous cell, and the long-term memory. The short-term memory is also known as the hidden state, and long-term memory is generally referred to as the cell state. These inputs can be seen in the bottom of Figure 3, which displays one LSTM unit with its inputs, activations, gates and outputs. For reference, a simple perceptron neural network would only have one input x_t, one activation σ and one output $h_{1,t}$.

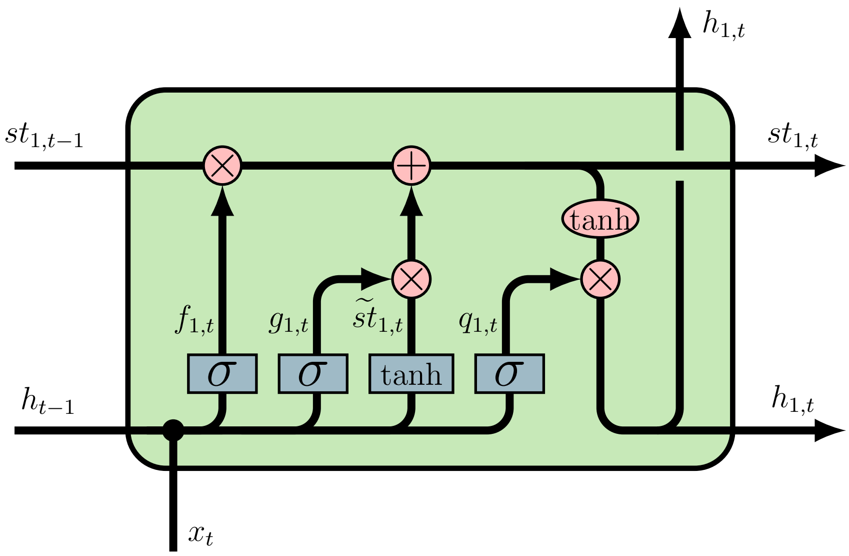

Figure 3: LSTM model with one hidden layer and N_h LSTM cells. The figure at the bottom is a magnification of the first LSTM cell with inputs, outputs, activations and gates. State's $st_{1,t-1}$ and output's h_{t-1} flow through time was only depicted for the first cell for reasons of clarity. The output layer's activation will be linear, which is why the output arrow was omitted.
This figure of the LSTM cell follows a depiction in [Olah (2015)].

An LSTM uses gates to regulate the information to be kept or discarded at each time step. These gates are known as the Input Gate, the Forget Gate, and the Output Gate and can be seen in Figure 3. Given an input vector x_t and a vector h_{t-1} with the past outputs and an existing memory or state $st_{i,t-1}$ in a specific LSTM cell i, the forget gate controls how much of this memory should be reset or forgotten. Its calculation is defined as

$$f_{i,t} = \sigma \left(b_i^f + \sum_{j=1}^{N_x} U_{i,j}^f x_{t,j} + \sum_{j=1}^{N_h} W_{i,j}^f h_{t-1,j} \right),$$

where b_i^f is the bias scalar, U_i^f the input weight vector, and W_i^f the output weight vector of the forget gate from the i^{th} LSTM unit [[Goodfellow et. al (2016)], p.405 ff]. The new state will be updated using this scalar $f_{i,t}$ as a weight. The sigmoid activation function σ leads to weights between 0 (total memory reset) and 1 (memory keeps all information). The second gate is the external input gate $g_{i,t}$. It controls how much current input should be considered in the new cell state. It is calculated similar to the forget gate

$$g_{i,t} = \sigma \left(b_i^g + \sum_{j=1}^{N_x} U_{i,j}^g x_{t,j} + \sum_{j=1}^{N_h} W_{i,j}^g h_{t-1,j} \right)$$

Note that it holds its own set of weights $W_{i,j}^g, U_{i,j}^g$ and biases b_i^g, that are calibrated in the training process independently of the others.
The next step is to calculate the candidate value for the new state $\widetilde{st}_{i,t}$ that interacts directly with the input gate. This value is again calculated with its own set of weights and biases as

$$\widetilde{st}_{i,t} = tanh \left(b_i + \sum_{j=1}^{N_x} U_{i,j} x_{t,j} + \sum_{j=1}^{N_h} W_{i,j} h_{t-1,j} \right)$$

With these two gates and the new candidate, the new cell memory state can be calculated as

$$st_{i,t} = f_{i,t} st_{i,t-1} + g_{i,t} \widetilde{st}_{i,t},$$

where the first term represents the forgetting of the old state and the second term the addition of the new state, each scaled by its gate $f_{i,t}$ or $g_{i,t}$.
The output gate $q_{i,t}$ is again a value between 0 and 1 and decides on the weight the current input should have for the output. It is calculated as

$$q_{i,t} = \sigma \left(b_i^q + \sum_{j=1}^{N_x} U_{i,j}^q x_{t,j} + \sum_{j=1}^{N_h} W_{i,j}^q h_{t-1,j} \right)$$

and holds again its own set of weights and biases.
The activation functions $\sigma()$ and $tanh()$ have the purpose of scaling some value $x \in \mathbb{R}$ into a specific range, i.e. $(0, 1)$ or $(-1, 1)$. They are defined as $\sigma(x) = \frac{1}{1+e^{-x}}$ and $\tanh(x) = \frac{2}{1+e^{-2x}} - 1$.

In our setup we are interested in one step ahead forecasts of credit spreads. To obtain them, we chose the three most recent data points of every step as the input data for the

LSTM. To fit the model, we chose MSE as our loss function and the Adam optimizing algorithm. Let Θ be the set of weights and biases from the LSTM model, as described above, y_i the true credit spread at time i and $\widehat{y_i}(\Theta)$ the output from the model for day t. Then the optimization problem can be formulated as

$$arg \min_{\Theta} MSE(\Theta), \text{where}$$
$$MSE(\Theta) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \widehat{y_i}(\Theta))^2$$

An LSTM model needs to be calibrated to the data set and optimal hyperparameter have to be found. This is done through experiments on training sets. The hyperparameters of the model used in our experimental study are shown in Table 2. For future work hyperparameters might still be improved by automatic search engines.

Parameter	Value
Hidden layers	1
Neurons per hidden layer	40
Loss function	MSE
Type of layer	LSTM
Activation output	Linear
Activation hidden layers	Sigmoid
Epochs	1000
Optimizer	Adam
Batch Size	128

Table 2: Hyperparameters for LSTM with a maximum of 1000 epochs and early stopping with patience = 120

For every credit spread time series we use the first half as training set for the LSTM and the second half as test set to evaluate the forecast quality. For the training, we set the number of epochs to 1000, although this number is hardly ever reached, due to early stopping. This causes the training to stop, if there has been no improvement with respect to the loss function in the last 120 epochs, and thus prevents overfitting.

Figure 4 shows the one-step ahead forecast obtained by an LSTM model as described above. One can see that the forecast appears like a smoothed and slightly lagged version of the actual spread.

Figure 4: LSTM one-step ahead forecast vs actual spread in test set

4.1 Linking LSTM and HMM

Even though a purely data driven forecast by a long short-term memory network might lead to improved results compared to traditional approaches, we would like to enrich our model by incorporating knowledge on movements of the data. Different ways of incorporating expert opinions have proven successful in applications of LSTMs or neural networks in financial time series (e.g. [Zhuge et al. (2017)], [Avdullai et al. (2021)]). Nevertheless, even though experts are extremely good in solving different tasks, there are some trends and patterns, which are usually hidden only in the data [Chattha et al. (2019)] and thus are likely to be visible to algorithms developed for this purpose.

Our LSTM is enhanced by including regime-switching information retrieved from the HMM filters described in Section 2. HMM-based feature engineering for machine learning algorithms has been recently introduced for credit card fraud detection by [Lucas (2020)]. They succesfully utilise HMMs to model sequences in credit card transactions and use state information in random forests. We include the filtered state estimation of the Markov chain as an additional feature into the LSTM. The goal is to give the information of current market states to the LSTM to improve the error measure of one-step ahead forecasts. The HMM filters are included in two steps: first, the estimated state of the Markov chain from the observation series, which is one particular credit spread series, shall be included. In a second step, an overall market regime feature shall be added, which adaptively measures the switching market state of a given market. Further information of HMM parameter estimates can be included in a third step. Given the additional market state information, the LSTM point forecast of the credit spread series shall be improved.

5 Simulation Study

To evaluate the performance of our HMM as well as the pure LSTM and the LSTM-HMM combination, we perform a simulation study. We simulate 1,000 credit spread paths as defined in Equation 3 using 2,500 time points each, which roughly corresponds to a ten year data set of daily data. We employ a two state HMM model with the following simulation parameters:

$$r_0 = 0.6,\ \alpha = (0.72, 0.90),\ \gamma = (0.2, 0.08),\ \eta = (0.08, 0.084),\ \Pi = \begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{pmatrix}$$

We obtain the initial parameters $\hat{\alpha}_0, \hat{\gamma}_0, \hat{\eta}_0$ for a two-state HMM out of blurred least square estimates of a single-state setting of the model described in Equation 3. The initialization window in our implementation comprises 250 data points. The initial Markov chain state probability estimate is set to be $(0.5, 0.5)$. As has been noted, data batches are processed through recursive filtering equations. We chose our batches to comprise $m = 30$ data points.

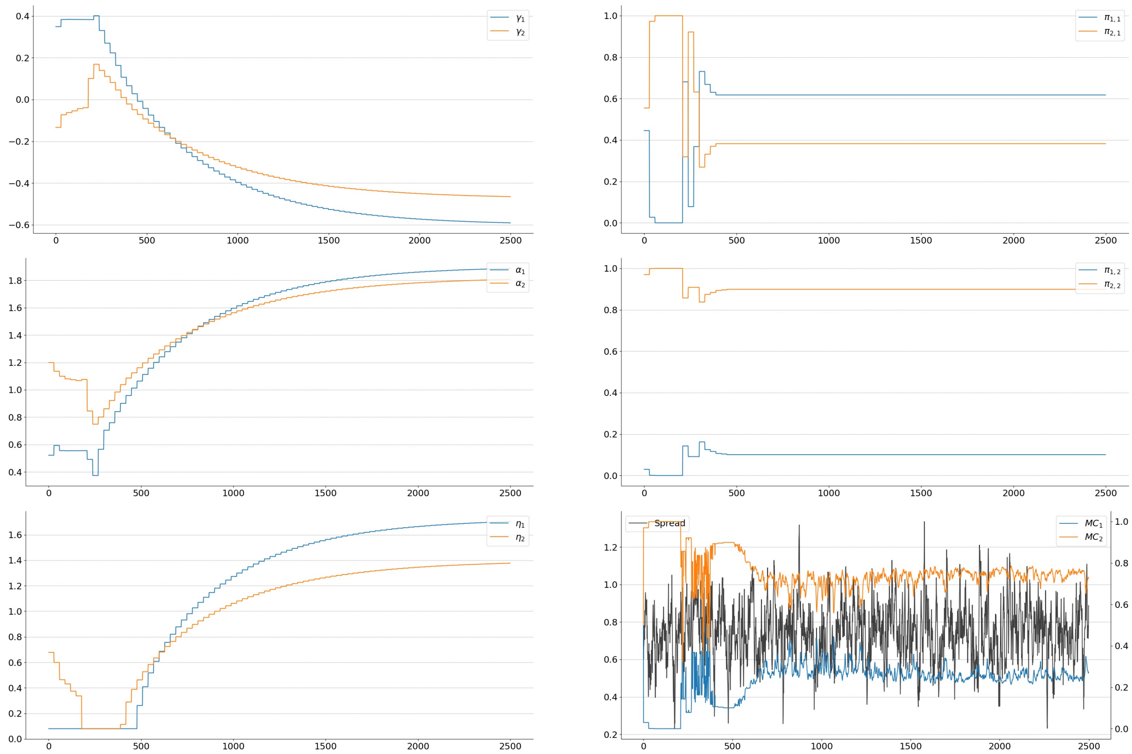

Figure 5: Simulated short-rate path and estimated parameters, single run.

Figure 5 shows a sample path of credit spreads as well as the adaptive estimation of parameters through the HMM over time. In addition, Figure 6 includes the evolution of the parameter estimates over all 1,000 simulations. It displays the median, five and

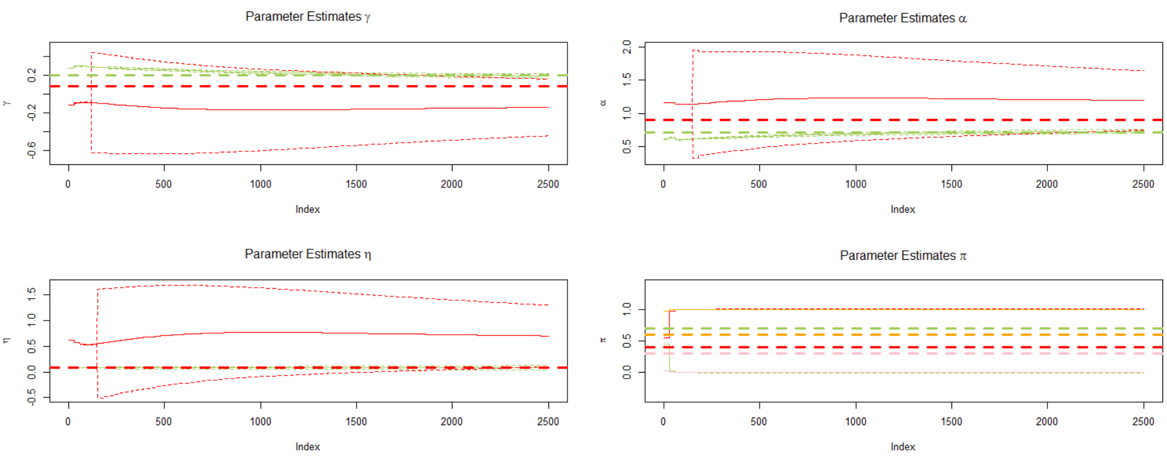

Figure 6: Confidence Bounds of Estimated Parameters - Median

	HMM	LSTM base	LSTM with MC
mean	0.019664	**0.007028**	0.007031
min	0.008769	0.006156	0.006134
25%	0.009905	0.006819	**0.006810**
50%	0.010305	**0.007007**	0.007011
75%	0.010721	**0.007238**	0.007257
max	4.662979	0.008096	0.008041

Table 3: Mean Squared Error (MSE) - 1000 simulations

ninety-five percentage quantiles and the simulation parameters. While the green colour demonstrates the first state of the estimated parameters, the red colour shows the second state of them. All parameter estimates show convergence and enclose the respective simulation parameters with the estimated confidence bounds.

To combine HMM and LSTM, the adaptively estimated states of the Markov chain from the simulated paths are included as an additional feature in the LSTM. In the following, we consider the forecasting accuracy of three models, HMM, pure LSTM and LSTM with HMM feature, and analyse the error measures Mean Squared Error (MSE) and Median Relative Absolute Error (MdRAE) of the simulated data set, shown in Table 3 and 4 respectively. Beyond the mean, we also state the median as well as the 25% and 75% of observed errors in this table due to outliers which might be misleading.

Both LSTM models without and with estimated Markov Chain give always better accuracy measure results than the pure HMM. While the LSTM performance without Markov Chain is better than the LSTM performance with Markov Chain in terms of MSE in around 54% of simulated time series, the LSTM performance with estimated Markov Chain feature is more accurate with respect to the MdRAE in just over 50% of time series. The simulation results therefore show potential within HMM features which shall be elaborated on a real data set in the following sections.

	HMM	LSTM base	LSTM with MC
mean	1.265576	**0.959159**	0.959271
min	1.098225	0.929989	0.931534
25%	1.141652	0.950594	**0.949893**
50%	1.155628	0.958166	**0.957990**
75%	1.171376	**0.966098**	0.966765
max	24.031442	1.021943	1.071003

Table 4: Median Relative Absolute Error (MdRAE) - 1000 simulations

6 Modelling of corporate bond spreads under regime shifting information

The models are now applied to our real-world data set, where the corporate credit spreads are calculated to a European base rate through Svenson yields (see [Svensson (1994)] for further details). Three European countries, Germany, France and Spain, are considered.

Our data set covers 149 corporate credit spread time series. To depict the typical characteristics of the time series, means are calculated per country over all spreads available for a given point in time since not all time series comprise an equal length. Maturities are not considered in this calculation. Figure 7 shows the credit spread means of the European countries which are considered in the dataset. The graph depicts the credit spreads between 2004 and 2020. Credit spreads in France and Spain increase significantly in the period of the European credit crisis (between 2011 and 2014) but come down to a more stable credit spread over the last years. The mean credit spread of French corporates remains higher between 2018 and 2020 compared to the other European countries in the data set. Corporate spreads of companies in Germany and France are more stable over the whole time period, often moving together in the same direction. A sharp increase in 2011 and 2012 is visible.

On the 149 spread time series, we perform a one-step ahead forecast with the developed models: (a) the OU-HMM in a 2-state setting is computed for the time series and its performance is analysed, (b) the LSTM is used to predict the credit spreads and (c) the third approach ensembles both models: a state prediction from the HMM for each credit spread series is included into the LSTM as an additional feature.

A typical plot of an estimated Markov chain through the 2-state HMM can be seen in Figure 8. The estimated state probabilities of the Markov chain are plotted above the credit spread, highlighting the interplay between rising or falling credit spreads and changing states of the Markov chain. The probabilities change more frequently in 2017 and 2018 after a sharp increase followed by a strong decrease of this credit spread. It is apparent, that changes in the market, which can be captured with the hidden Markov

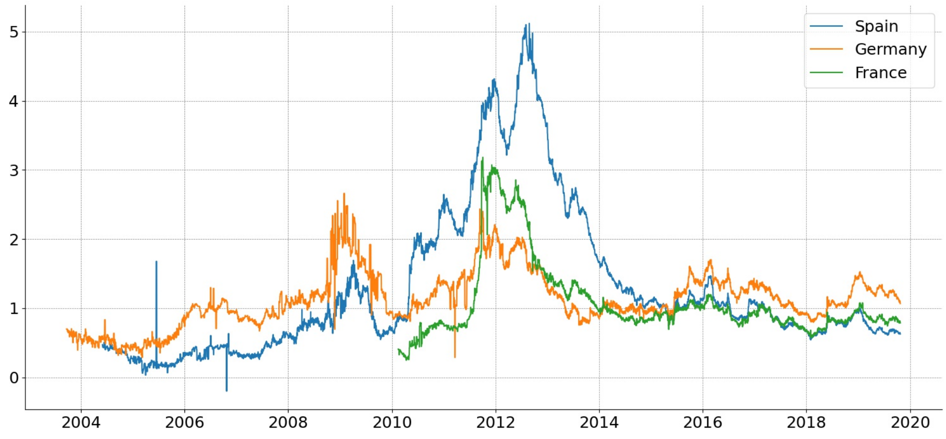

Figure 7: Mean corporate credit spreads over time in Spain, Germany and France

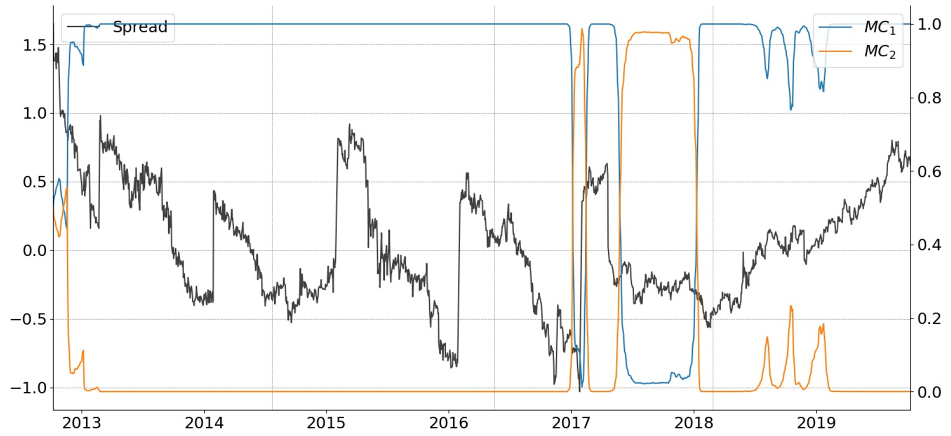

Figure 8: Credit spread and state estimation

chain, affect the credit spreads.

The filtered hidden Markov chain of each single spread in the data set is then fed into the LSTM. We examine the performance of the models in the data set by comparing prediction errors MdRAE and MSE. The performance of the ensemble HMM-LSTM model is dependent on the performance of the HMM prediction. The distribution of the prediction error MdRAE of one-step ahead predictions with HMM in the different countries is depicted in Figure 9. The spreads from companies in the Spanish index IBEX 50 have the highest prediction errors in contrast to spreads from constituents of the French index CAC and from German companies listed in the DAX. This might be caused by the fact,that the Spanish time series are more volatile and contain a higher number of outliers. To overcome this issue, one could consider a robustification of the algorithm described in Section 2 for future work, see e.g. [Erlwein-Sayer and Ruckdeschel (2014)].

Based on the accuracy, i.e. MdRAE, of the HMM prediction, we divide the data

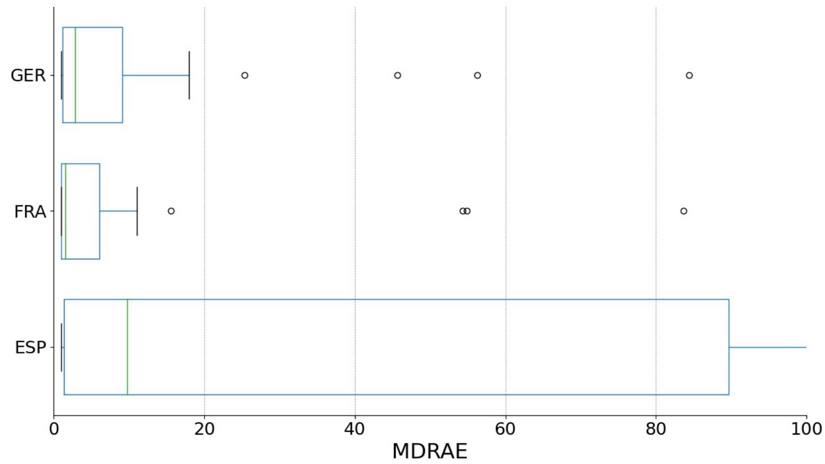

Figure 9: Distribution of MdRAE for different countries

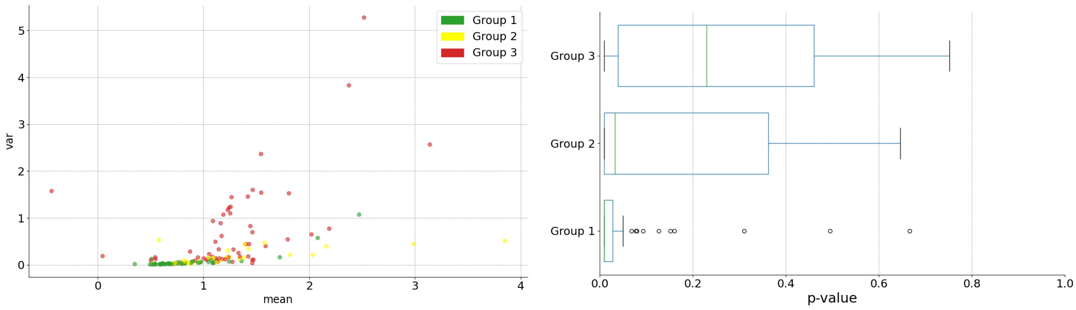

Figure 10: Distribution (mean and variance) of credit spreads grouped by the fit of HMM results and p-values of the Dickey Fuller test on those groups

set into three groups: in Group 1, the performance measure of HMM are very good, an accurate one-step ahead performance is reached (54 spreads). Group 2 incorporates spreads where the HMM performance is still good, but error measures are higher, we label them as normal forecasts (33 spreads). Group 3 contains those spreads which lead to high forecast errors when predicted with HMM (59 spreads). Here, the credit spread cannot be captured well through an HMM approach. In Figure 10, we depict the error distribution in each group. Furthermore we perform the augmented Dickey-Fuller test with added constant to test for stationarity. Spreads in Group 1 and 2 are more likely to be stationary based on the resulted p-value. That highlights the fact that the OU-HMM assumes an underlying stationary time series and thus non-stationary time series tend to lead to poorer forecast results in that model.

We now incorporate the forecasted HMM state into the LSTM prediction and analyse the performance of the ensembled model over the three groups. Overall we see that ensembling HMM and LSTM leads to an improved one-step ahead prediction. The improvements are highest for time series in Group 1, where the HMM forecast performs

	HMM	LSTM base	LSTM with MC	LSTM with means
mean	**1.114139**	1.144359	1.181757	1.292254
min	0.994600	0.846091	0.861230	0.858935
25%	1.008341	**0.913331**	0.919954	0.930733
50%	1.039138	0.960752	**0.938366**	0.978500
75%	1.214803	1.000097	**0.999914**	1.151036
max	1.487116	8.859531	7.674243	6.991015

Table 5: Performance of models: Mean of MdRAE for first set of 54 credit spreads

	HMM	LSTM base	LSTM with MC	LSTM with means
mean	2.607646	**1.701957**	1.912710	2.165460
min	1.526479	0.904732	0.916230	0.916088
25%	1.973540	**0.958066**	0.963007	0.984040
50%	2.676886	**1.036917**	1.057541	1.115894
75%	3.212400	**1.356105**	1.515903	1.875953
max	3.944889	14.450686	16.231333	16.040625

Table 6: Performance of models: Mean of MdRAE for second set of 33 credit spreads

best.
Table 5 shows the mean of the MdRAE for forecasted credit spreads in Group 1. In mean, the pure HMM forecast leads to best results, while the median and the 75% quantile error is best for the LSTM model with HMM information. In the last column, we depict forecast results, when the LSTM is combined with a forecasted state of the Markov chain over the mean credit spread of the country. This model cannot outperform the accuracy of the model with the single HMM state prediction. This might be due to the fact that the mean spread is taken for each day over all sectors and durations. Thus, the time series might not have the same hidden factors of influence and the filtered state is not valid equally for all corporate credit spreads of one country. In summary, we observe an advantage of incorporating HMM results in the LSTM in Group 1 and 3. While Group 1 provides well-fitting HMM forecasts which are thus also suitable features for the LSTM, this is not valid for Group 3. However, even though the forecasts itself are not highly accurate, the estimated Markov chain and especially the time points considered for changes in regime give still an additional insight to the time series and thus give an advantage in highly volatile cases as contained in Group 3.

We furthermore evaluate the outperformance of the LSTM model with additional HMM features over the base LSTM model in Table 8. It becomes apparent, that the LSTM with individual HMM estimation outperforms the pure LSTM in at least 58% of the cases. Whereas the mean spread HMM enhances the pure LSTM in at least 46% . This leads to the conclusion, that incorporating HMM state estimations can improve pure LSTM models for credit spreads.

	HMM	LSTM base	LSTM with MC	LSTM with means
mean	165.083743	1.628369	**1.339296**	1.501138
min	5.070772	0.893300	0.901667	0.884119
25%	9.703972	0.984138	**0.967197**	0.970316
50%	27.987564	1.140296	**1.117939**	1.100954
75%	99.560398	**1.498353**	1.581059	1.495457
max	1485.709200	16.509256	3.862876	6.671055

Table 7: Performance of models: Mean of MdRAE for third set of 59 credit spreads

LSTM MSE Outperformace, adding MC	0.664
LSTM MSE Outperformace, adding means MC	0.548
LSTM MdRAE Outperformace, adding MC	0.582
LSTM MdRAE Outperformace, adding means MC	0.466

Table 8: Improvement in error measure with added HMM feature over the set of all time series (146 credit spread series)

7 Conclusion

Corporate credit spreads exhibit regime-switching dynamics through time. When analysing the real data set we found higher spreads and volatilities in times of financial turmoil or unstable markets. To capture the regime-switching characteristics, an on-line estimated HMM is developed and further utilised. Modelling and prediction of European corporate credit spreads is performed in this work through three models: the first model is a pure HMM approach where parameters and hidden states are estimated adaptively through time. The resulting one-step ahead forecast of the spreads shows low prediction errors for most of the analysed spreads of the real data set. Secondly, we apply an LSTM model for the one-step ahead prediction and merge this model with the regime-switching HMM by including estimated state information into the LSTM. The enhanced HMM-LSTM model outperforms the pure model set-ups for more than half of the analysed corporate credit spreads. We can see that the underlying market state adds additional information to the LSTM which is utilized to increase the precision of the one-step ahead prediction. We note that for real-world corporate credit spreads which are mean-reverting the forecast could typically be improved.

Corporate credit spreads often reflect market characteristics. Therefore, future work will include market fundamentals into the models. The actual market situation which is stated through an estimated regime can then be further enhanced through economic indicators.

Acknowledgement:

The implementation and data analysis was supported by Ivo Richert, which we gratefully acknowledge.

References

[Afonso et al. (2015)] Afonso, A., Arghyrou, M. G., Bagdatoglou, G. & Kontonikas, A. (2015) On the time-varying relationship between EMU sovereign spreads and their determinants. *Economic Modelling*, 44, 363-371.

[Ang and Bekaert(2002)] Ang, A., Bekaert, G. (2002) Regime Switches in Interest Rates. *Journal of Business and Economic Statistics,* 20, 163-182.

[Ang and Timmermann(2012)] Ang, A.,Timmermann, A. (2012) Regime changes and financial markets. *Annual Review of Financial Economics*, 4(1), 313-337.

[Bernoth and Erdogan (2012)] Bernoth, K., Erdogan, B. (2012) Sovereign bond yield spreads: A time-varying coefficient approach. *Journal of International Money and Finance*, 31(3), 639-656.

[Chattha et al. (2019)] Chattha, M. A., Siddiqui, S. A., Malik, M. I., van Elst, L., Dengel, A. & Ahmed, S. (2019) KINN: Incorporating Expert Knowledge in Neural Networks. *arXiv:1902.05653.*

[Clark and Baccar (2018)] Clark, E., Baccar, S. (2018) Modelling credit spreads with time volatility, skewness and kurtosis. *Annals of Operations Research*, 262, 431–461.

[Elliott et al. (1995)] Elliott, R.J., Aggoun, L. & Moore, J.B. (1995) *Hidden Markov Models: Estimation and Control*, Springer, New York.

[Erlwein and Mamon (2009)] Erlwein, C., Mamon, R. (2009) An online estimation scheme for a Hull–White model with HMM-driven parameters. *Statistical Methods and Applications*, 18(1), 87-107.

[Erlwein-Sayer and Ruckdeschel (2014)] Erlwein-Sayer, C., Ruckdeschel, P. (2014) Robustification of an on-line EM algorithm for modelling asset prices within an HMM. In: R. Mamon and R. J. Elliott (Eds.) *Hidden Markov Models in Finance*, 1-31. Springer, Boston,MA.

[Erlwein-Sayer (2018)] Erlwein-Sayer, C. (2018) Macroeconomic News Sentiment: Enhanced Risk Assessment for Sovereign Bonds. *Risks*, 6(4), 141.

[Erlwein, Grimm et al (2020)] Erlwein-Sayer, C., Grimm, S., Ruckdeschel, P., Sass, J., & Sayer, T. (2020) Filter-based portfolio strategies in an HMM setting with varying correlation parametrizations.*Applied Stochastic Models in Business and Industry* , 36(3), 307-334.

[Fischer] Fischer, T., Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2), 654-669.

[Golbayani et. al (2020)] Golbayani, P., Wang, D. & Florescu, I. (2020) Application of Deep Neural Networks to assess corporate Credit Rating. *arXiv:1803.06386*

[Goodfellow et. al (2016)] Goodfellow, I., Bengio, Y., & Courville, A. (2016) Deep learning, MIT Press.

[Gray (1996)] Gray, S.F. (1996) Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics*, 42(1),27-62.

[Grimm et al. (2020)] Grimm, S., Erlwein-Sayer, C., & Mamon, R. (2020) Discrete-time implementation of continuous-time filters with application to regime-switching dynamics estimation. *Nonlinear Analysis: Hybrid Systems,* 35, 100814.

[Guidolin and Timmermann (2007)] Guidolin, M., Timmermann, A. (2007) Asset allocation under multivariate regime switching. *Journal of Economic Dynamics and Control*, 31(11), 3503–3544.

[Hamilton (1989)] Hamilton, J., D. (1989) A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica*, 57(2), 357–384.

[Hochreiter and Schmidhuber (1997)] Hochreiter, S., Schmidhuber, J. (1997) Long short-term memory. *Neural Computation*, 9(8), 1735–1780.

[Klein and Pliszka (2018)] Klein, A., Pliszka, K. (2018) The time-varying impact of systematic risk factors on corporate bond spreads. Deutsche Bundesbank No 14/2018.

[Kobayashi (2021)] Kobayashi, T. (2021) Common Factors in the Term Structure of Credit Spreads and Predicting the Macroeconomy in Japan. *International Journal of Financial Studies*, 9(2), 23.

[Landen (2000)] Landen, C. (2000) Bond pricing in a hidden Markov model of the short rate. *Finance and Stochastics*, 4(4), 371-389.

[Avdullai et al. (2021)] Avdullai, F., Korn, R. & Hoben, HM. (2021) Neuronale Netze und Zeitreihenansätze zur Vorhersage des auslösenden Faktors in der privaten Krankenversicherung. *ZVersWiss,* 110, 21–48.

[Lucas (2020)] Lucas, Y., Portier, P., Laporte, L., He-Guelton, L., Caelen, O. Granitzer, M. & Calabretto, S. (2020) Towards automated feature engineering for credit card fraud detection using multi-perspective HMMs. *Future Generation Computer Systems*, 102, 393-402.

[Maltritz (2012)] Maltritz, D. (2012) Determinants of sovereign yield spreads in the Eurozone: A Bayesian approach. *Journal of International Money and Finance*, 31(3), 657-672.

[Mamon and Elliott (2007)] Mamon, R. S., Elliott, R. J. (Eds.). (2007) *Hidden Markov models in finance.* Springer, New York.

[Manzo and Qiao (2020)] Manzo, G., Qiao, X., (2020) Deep Learning Credit Risk Modeling. Available at SSRN: http://dx.doi.org/10.2139/ssrn.3651379

[Mehtab and Sen (2020)] Mehtab, S., Sen, J., (2020) Analysis and Forecasting of Financial Time Series Using CNN and LSTM-Based Deep Learning Models. *2nd International Conference on Advances in Distributed Computing and Machine Learning.*

[Nguyen (2018)] Nguyen, N. (2018) Hidden Markov model for stock trading. *International Journal of Financial Studies*, 6(2), 36.

[Olah (2015)] Olah, C. (2015). Understanding lstm networks. https://colah.github.io/posts/2015-08-Understanding-LSTMs/, last visited: 24.09.2021

[Persio and Honchar (2016)] Di Persio, L., Honchar, O. (2016) Artificial neural networks architectures for stock price prediction: Comparisons and applications. *International Journal of circuits, systems and signal processing*, 10, 403-413.

[Papouskova and Hajek (2019)] Papouskova M., Hajek P. (2019) Two-stage consumer credit risk modelling using heterogeneous ensemble learning. *Decision Support Systems,* 118 , 33-45.

[Roondiwala et al. (2017)] Roondiwala, M., Patel, H., & Varma, S. (2017) Predicting stock prices using LSTM. *International Journal of Science and Research (IJSR)*, 6(4), 1754-1756.

[Sass and Haussmann (2004)] Sass, J., Haussmann, U.G. (2004) Optimizing the Terminal Wealth under Partial Information: The Drift Process as A Continuous Time Markov Chain. *Finance and Stochastic* 8, 553–577.

[Siami-Namini and Siami Namin(2018)] Siami-Namini, S., Siami Namin, A. (2018) Forecasting Economics and Financial Time Series: ARIMA vs. LSTM. *arXiv preprint arXiv:1803.06386*.

[Yan and Ouyang (2018)] Yan H., Ouyang H. (2018) Financial Time Series Prediction Based on Deep Learning. *Wireless Personal Communications,* 102, 1-18.

[Svensson (1994)] Svensson, L.E. (1994) Estimating and Interpreting Forward Interest Rates: Sweden 1992-1994.*National Bureau of Economic Research,* No. w4871.

[Thomas et al. (2002)] Thomas, L. C., Allen, D. E., & Morkel-Kingsbury, N. (2002) A hidden Markov chain model for the term structure of bond credit risk spreads. *International Review of Financial Analysis*, 11(3), 311-329.

[Xiong et al. (2016)] Xiong, R. Y., Cai, H., Diego-Guerra, I., Lu, Y., Xu, X., & Yin, Y. (2016) Forecasting Credit Spreads: A Machine Learning Approach. *International Association for Quantitative Finance*, available at: https://www.iaqf.org/resources/Documents/2019 Student Competition/UMich_Submartingale_Web.pdf

[Zhuge et al. (2017)] Zhuge, Q., Xu, L., & Zhang, G. (2017) LSTM neural network with emotional analysis for prediction of stock price. *Engineering Letters,* 25(2), 167-175.