The application of hidden Markov models for economic state prediction and the calculation of economic state transition probabilities

Master Thesis Financial Engineering & Management

University of Twente

T.W. Rikken
Amsterdam, August 15, 2022

This is a dissertation submitted in fulfillment of the degree Master of Science in Industrial Engineering & Management.

Document properties	
Author	T.W. Rikken
University	University of Twente Drienerlolaan 5 7522 NB Enschede
Programme	Master Industrial Engineering & Management
Specialization	Financial Engineering & Management
Company	Ernst & Young Antonio Vivaldistraat 150 1083 HP Amsterdam
Department	Technology Consulting
Team	Digital & Emerging Technology
Examination Committee	
First supervisor	Dr. Ir. W.J.A. van Heeswijk University of Twente
Second supervisor	Dr. R.A.M.G. Joosten University of Twente
External supervisor	C. Meunier Charlette M.Sc. Quantitative Advisory Services

Management summary

The International Financial Reporting Standards (IFRS) requires financial institutions (or other companies with financial assets like loans) to estimate potential credit losses with a forward-looking view. Most financial institutions (77%) are taking a scenario-based approach to include forward-looking macro-economic impact in their estimations of potential credit risk losses. The weights of the three scenarios used by most financial institutions are often quite basic where the most likely scenario (baseline) accounts for 50% and the remaining two (upside and downside) share the other 50% equally. These weights are currently not determined by any quantitative method, hence, this research aims to determine these weights based on a quantitative method. To meet this goal the method should first be able to predict and identify the economic states (the economic scenarios) correctly. We used several hidden Markov models for this purpose.

To validate if the hidden Markov models can predict the economic states, we compare the recession state predicted by the hidden Markov models (HMM) with the historical recession data of the countries United States, United Kingdom, and Japan. We validate the performance using the performance metrics accuracy, recall, precision, and the F-score. We benchmark the performance of the HMM against three methods:

1. Classifying all data points as no-recessions. This result in a high accuracy (average of 0.77) with unbalanced data, however for the other performance metrics this is not the case (average of 0.00 for all other performance metrics).

2. Classifying all data points with negative growth as a recession. A negative growth rate is a common indicator of a recession. The average accuracy for this method is 0.78 and the average F-score is 0.51.

3. The martingale method. In this research this results in using the current state x_t as a prediction for the next state x_{t+1}. The assumption with the martingale is that the current state with a probability of 1 is the next state, which is not in line with the requirements of the IFRS9, which states that it is not allowed to have only one scenario unless there are adjustments made to compensate for the non-linearity in the expected credit risk losses. The martingale method does have high scores for the performance metrics, with an average accuracy of 0.97 and an average F-score of 0.93.

Furthermore, we calculated the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) using the likelihood calculation of the hidden Markov model, to compare the relative performance of the different types of hidden Markov models with each other. We experimented with the number of hidden states, the amount of historical input-data used, the initialization of the hidden Markov model, cutoff points for extreme outliers, and the features used to predict the economic states, to evaluate the best performance of the hidden Markov models.

In short, we can conclude that the two-state, and especially the three-state higher-order hidden Markov models can predict economic states. The three-state model outperforms all previously discussed methods with an average accuracy of 0.85 and an average F-score of 0.68. These scores are higher than the scores of the GDP classification method, which has an average accuracy of 0.78 and an average F-score of 0.51. The no recession method has an average accuracy of 0.77 and an average F-score of 0.00, which are also lower than the scores of the three-state HMM. Only the martingale method has a better performance with an average accuracy of 0.97 and an average F-score of 0.93. However, as previously discussed a single forward-looking scenario (i.e the most likely scenario) does not meet the requirements of IFRS9 unless there is an adoption of an adjustment to reflect non-linearity in the credit loss distribution for alternative scenarios.

We also calculated the correlations between the state sequences (output of the HMM) of the countries for all models, and compared them to the correlations of the historical data, resulting in an additional method of validation. For the three-state HMM these correlations were quite close. The correlation between the US-UK is 0.5 for the three-state HMM and the historical correlation is 0.7. For the US-JP this is 0.2 for the three-state HMM and 0.2 for the historical correlations. Lastly, the correlation for the UK-JP is 0.1 for the three-state HMM and 0.2 for the historical correlation. The closeness of the correlation values for the economic state sequences is another indication that the three-state HMM is working properly.

We calculated the AIC and BIC for the base model, the two-state model, and the three-state model. The two-state model had the lowest (indicating the best performance for AIC and BIC) value. There are several reasons for the two-state model outperforming the three-state model:

1. The AIC and BIC add a complexity term, which goes up when the number of states increases, which results in a higher score for the AIC and BIC.
2. It is easier to divide the economic periods into two states than three states. Therefore the paths are easier to predict, which results in more likely paths and a higher log-likelihood value.

Lastly, the economic state transition probabilities (weights) calculated by the HMM are not in line with the weight distribution currently used by most financial institutions (0.5, 0.25, 0.25). The state transition probabilities calculated by the HMM are not close to the weight distribution currently used for the United States, United Kingdom, and Japan. This deviation indicates that the current method is not optimal and that it is worth to research the impact of using the weights calculated by the HMM for the expected credit loss calculations.

Preface

This master thesis *"The application of hidden Markov models for economic state prediction"* finalizes my master Financial Engineering & Management at the University of Twente and my days as a student. I really enjoyed the master's program, as well as my internship.

Several people have helped me during this research, and I would like to thank them. First, I would like to thank the EY Digitial & Emerging Technology Consulting team for the opportunity to write my master thesis, work on a project, and follow the Tech Consulting training sessions. Furthermore, I would like to thank Christophe Meunier Charette, my supervisor from EY. His excellent insight into the quantitative aspects of this research has helped me a lot.

Moreover, I would like to express my gratitude to Wouter van Heeswijk and Reinoud Joosten, my supervisors from the university, for providing me with useful feedback during this research.

Finally, I would like to thank my family and friends for their support.

Thom Rikken
August 15, 2022

Glossary

- **Correlation matrix** A correlation matrix depicts the correlation between all the possible pairs of variables.
- **Exponential moving average** The exponential moving average is a type of weighted moving average that gives more weighting or importance to recent price data.
- **Gross Domestic Product** Gross domestic product is the total monetary or market to recent price data. value of all the finished goods and services produced within a country's borders in a specific time period.
- **Hidden Markov model** The hidden Markov model is a Markov chain whose internal state cannot be observed directly but only through some probabilistic function.
- **Inflation** Inflation is the decline of purchasing power of a given currency over time.
- **Interest rate** The interest rate is the amount a lender charges a borrower and is a percentage of the principal—the amount loaned.
- **Machine learning** The study of computer algorithms that can improve automatically through experience and by the use of data.
- **macroeconomics** macroeconomics is the study of behavior of the economy as a whole.
- **Market index** A market index is a hypothetical portfolio of investment holdings that represents a segment of the financial market.
- **Markov process** A Markov chain or Markov process is a stochastic model describing a sequence of possible events in which the probability of each event depends only on the state attained in the previous event.
- **Observation sequence** An observation sequence is assumed to be produced by a series of hidden states.
- **Observation probability** The observation probabilities define the probability of seeing certain observed variable given a certain value for the hidden variables/states.
- **Simple moving average** A simple moving average is an arithmetic moving average calculated by adding recent values and then dividing the the summed value by the number of time periods in the calculation average.
- **Standard deviation** The standard deviation is a measure of the amount of variation or dispersion of a set of values.

- **State sequence** The state sequence specifies the sequence of states that the hidden Markov model visits.
- **Transition probability** The probabilities that explain the transition to/from hidden states are transition probabilities.
- **Unemployment rate** The unemployment rate is the percentage of the labor force without a job.

Acronyms and abbreviations

- CA Canada.
- CN China.
- CPB Central Planning Bureau.
- CPI Inflation.
- D&ET Digital & Emerging Technology.
- DE Germany.
- EM Expectation-Maximization.
- EMA Exponential moving average.
- ES Spain.
- EY Ernst & Young.
- FR France.
- GDP Gross Domestic Product.
- HK Hong Kong.
- HMM Hidden Markov model.
- IT Italy.
- Itr Interest rate.
- JP Japan.
- KR Korea.
- MX Mexico.
- NBER National Bureau of Economic Research.
- NL Netherlands.
- RU Russia.
- SG Singapore.

- SMA Simple moving average.
- TW Taiwan.
- UK United Kingdom.
- UNR Unemployment rate.
- US United States.

Mathematical notations used

- $\alpha_{t-1}(i)$ The previous forward path probability from the previous time step.
- $\alpha_t(j)$ Probability of being in state j after seeing the first t observations, given λ. Also called the forward term or forward variable.
- a_{ij} The transition probability from previous state x_i to current state x_j.
- A The state transition probability distribution.
- $b_j(O_t)$ The state observation likelihood of the observation symbol O_t given the current state j.
- B The observation probability distribution.
- $\beta_t(j)$ Probability of being in state j at time t given everything that comes after t, given λ. Also called the backward term or backward variable.
- EMA_t Exponential moving average at time t.
- $\gamma_t(j)$ Probability of q_j at time t given an observation sequence $\mathcal{O}$ and the model λ.
- λ Compact notation for the HMM, $\lambda = (A, B, \pi)$.
- μ_{GDP} Mean of a country's GDP over the period used by the HMM.
- M The number of distinct observation symbols.
- N Total number of unobservable states.
- O_t Observable signal at time t.
- $\mathcal{O}$ Sequence of observations $O_1 ... O_T$.
- q_i State i in the state space $\mathcal{S}$.
- π The initial state distribution.
- σ_{GDP} Standard deviation of a country's GDP over the period used by the HMM.
- s Smoothing factor used for the exponential moving average.
- $\mathcal{S}$ State space giving the distinct states of the HMM.
- SMA_t Simple moving average at time t.

- t Clock time.
- T Length of the observation and state sequence $1 \leq t \leq T$.
- $\mathcal{V}$ Observable signal space giving the set of possible observations.
- $\xi_t(i, j)$ Probability of state q_i at time t and state q_j at time $t+1$.
- x_t State of the HMM at time t.
- $\mathcal{X}$ Sequence of states $x_1 ... x_T$.
- y Multiplying factor used to determine the cutoff points for the GDP feature.

List of Figures

1 Percentage of countries in a recession between 1871 and 2021. Note that the percentage of countries in a recession is the highest during the COVID-crisis (Buysse & Essers, 2020). 2
2 Approach used for evaluating the range of possible outcomes in the expected credit loss amount (EBA, 2021). Note that most financial institutions use a scenario based approach with 3 scenarios 4
3 Schematic representation of the scenario weights currently used by most financial institutions to include macro-economic impact in their expected credit loss calculations. The numbers above the arrows represent the probabilities to move to the particular scenario the arrow points to. 5
4 Chronological steps taken during this research. 8
5 Machine Learning Algorithm Mind Map (Kumar, 2020). 10
6 Visualization of the classification type of supervised learning. Note that the different classes are separated by a decision boundary (ODSC, 2020). 11
7 Visualization of the regression type of supervised learning. Note that the regression type identifies a pattern in the data (ODSC, 2020). 12
8 Schematic representation training and test set (Joy, 2022). 12
9 Schematic representation of the HMM weather guessing game. The start indicates with which probability the sequence starts in which state. So there is a 0.6 probability that the first state is 'Rainy' and 0.4 that it is 'Sunny'. The black lines between the states 'Rainy' and 'Sunny' indicate the state transition probabilities. The circles with 'Walk', 'Shop', and 'Clean' indicate the observable processes and the dotted lines indicate the emission probabilities (Wikipedia, 2022). 18
10 Illustration hidden Markov model (Stamp, 2004). 20
11 Illustration model topologies of the hidden Markov model. 21
12 Example of a Gaussian mixture probability density. The density consists of three normal distributions ($\mu_1 = 5, \mu_2 = 10, \mu_3 = 15$) with equal weights. 21
13 Visualization of combining the α and β to find the state sequence. We find the probability of the most likely state sequence by dividing the most likely state sequence path, by all possible paths (Patterson, 2020b). 23
14 Schematic visualization of the example of the robot vacuum cleaner. The black arrows indicate the transitions the robot can make between the possible states q_i. The red arrow indicates an impossible transition (transition probability is equal to zero between q_5 to q_1) that can be returned with the method discussed in this section. . . 24
15 Schematic representation the two-state annual HMM for the US. The black arrows beginning at 'start' represent the initial state distribution, the black arrows originating from 'Recession' and 'No Recession' represent the state transition distribution, and the dotted lines the emission probabilities. 28
16 Schematic representation of the confusion matrix. 31
17 Schematic representation of the validation process for the 3-state hidden Markov model. The first column indicates the output of the model. After this, the performance metrics are calculated for all three possible recession state scenarios; In case 0, 1, or 2 is the recession state. Finally, it must be determined based on the performance metrics and graphs what the recession state is. 33

18 Schematic representation of the scenario weights currently used by most financial institutions to include macro-economic impact in their expected credit loss calculations. The numbers above the arrows represent the probabilities of moving to the particular scenario the arrow points to. 34
19 Schematic representation of the scenario weights used by the Martingale method. The numbers above the arrows represent the probabilities of moving to the particular scenario the arrow points to. 35
20 Correlation plot of the historical recession periods for the countries the United States, United Kingdom, and Japan. 36
21 US: Prediction of the model versus historical recession data. The blue line in the graph shows when the historical recessions took place, if the blue line is 1, there was a recession and if it is 0 there was no recession. The black line shows when the model predicts a recession. The recession state is 1 in this case, if the black line is 0 there was no recession. 40
22 US: Mean of the features in state i (=0, 1) versus the actual values of the features. The blue line here is the actual value of the feature and the green line is the average value of the feature given the hidden state it is in. 41
23 UK: Prediction of the model versus historical recession data. The blue line in the graph shows when the historical recessions took place, if the blue line is 1, there was a recession and if it is 0 there was no recession. The black line shows when the model predicts a recession. The recession state is 0 in this case, if the black line is 1 there was no recession. 42
24 UK: Mean of the features in state i (=0, 1) versus the actual values of the features. The blue line here is the actual value of the feature and the green line is the average value of the feature given the hidden state it is in. 42
25 Prediction of the JP versus against historical recession data and probability of being in state i (=0, 1) according to the model. The blue line in the top graph shows when the historical recessions were. The black line shows when the model predicts a recession. At best, these two graphs would overlap exactly, or be exactly mirrored. This is because the model sometimes gives the recession state a 0 and the other time a 1. In this case, the recession state is 0. The other graphs in the figures indicate the probability according to the model that a country is in a particular hidden state at that moment. 43
26 JP: Mean of the features in state i (=0, 1) versus the actual values of the features. The blue line here is the actual value of the feature and the green line is the average value of the feature given the hidden state it is in. 44
27 Correlation plot countries state sequences created by the 2 state base model. 45
28 Correlation plot of the historical recession periods for the countries the United States, United Kingdom, and Japan. 45
29 US: Prediction of the model versus historical recession data. The blue line in the graph shows when the historical recessions took place, if the blue line is 1, there was a recession and if it is 0 there was no recession. The black line shows when the model predicts a recession. The recession state is 1 in this case, if the black line is 0 there was no recession. 49
30 US: Mean of the features in state i (=0, 1) versus the actual values of the features for the model. The blue line here is the actual value of the feature and the green line is the average value of the feature given the hidden state it is in. 50

31 UK: Prediction of the model versus historical recession data. The blue line in the graph shows when the historical recessions took place, if the blue line is 1, there was a recession and if it is 0 there was no recession. The black line shows when the model predicts a recession. The recession state is 1 in this case, if the black line is 0 there was no recession. 51
32 UK: Mean of the features in state i (= 0, 1) versus the actual values of the features for the model. The blue line here is the actual value of the feature and the green line is the average value of the feature given the hidden state it is in. 51
33 Prediction of the JP versus historical recession data and probability of being in state i (=0, 1) according to the model. The blue line in the top graph shows when the historical recessions were. The black line shows when the model predicts a recession. The other graphs in Figure indicate the probability according to the model that a country is in the particular hidden state at that moment. 52
34 JP: Mean of the features in state i (= 0, 1) versus the actual values of the features for the model. The blue line here is the actual value of the feature and the green line is the average value of the feature given the hidden state it is in. 53
35 Correlation plot of the historical recession periods for the countries the United States, United Kingdom, and Japan. 53
36 Correlation plot countries state sequences for the improved model. 54
37 Transition probabilities calculated by the HMM adjusted for yearly transitions between economic states. It is noticeable that the probabilities calculated by the HMM are not in line with the weight distribution currently used (0.5, 0.25, 0.25) by most financial institutions. 58
38 US: Prediction of the model versus historical recession data. The blue line in the graph shows when the historical recessions took place, if the blue line is 1, there was a recession and if it is 0 there was no recession. The black line shows when the model predicts a recession. The recession state is 0 in this case, if the black line is 1 or 2 there was no recession. 59
39 US: Mean of the features in state i (= 0 to 2) versus the actual values of the features for the model. The blue line here is the actual value of the feature and the green line is the average value of the feature given the hidden state it is in. 59
40 UK: Prediction of the model versus historical recession data. The blue line in the graph shows when the historical recessions took place, if the blue line is 1, there was a recession and if it is 0 there was no recession. The black line shows when the model predicts a recession. The recession state is 2 in this case, if the black line is 0 or 1 there was no recession. 60
41 UK: Mean of the features in state i (=0 to 2) versus the actual values of the features for the model. The blue line here is the actual value of the feature and the green line is the average value of the feature given the hidden state it is in. 60
42 JP: Prediction of the model versus historical recession data. The blue line in the graph shows when the historical recessions took place, if the blue line is 1, there was a recession and if it is 0 there was no recession. The black line shows when the model predicts a recession. The recession state is 2 in this case, if the black line is 0 or 1 there was no recession. 61
43 JP: Mean of the features in state i (= 0 to 2) versus the actual values of the features for the model. The blue line here is the actual value of the feature and the green line is the average value of the feature given the hidden state it is in. 62

44 Correlation plot of the countries state sequences of the three-state hidden Markov model. 63
45 Correlation plot of the historical recession periods for the countries the United States, United Kingdom, and Japan. 63
46 Correlation matrix of the state sequences produced by the 5 state hidden Markov model. 65
47 Correlation plot of the historical recession periods for the countries the United States, United Kingdom, and Japan. 65
48 Example of the calculation of the likelihood for one particular sequence of observations and states. The first black arrow indicates with what probability the sequence starts in the "Rainy" state (π_1), the other black arrows give the state transition probabilities (a_{12} and a_{22}) and the dotted arrows give the emission probabilities ($b_1(O_1)$, $b_2(O_2)$, $b_2(O_3)$). 77
49 For the convenience of the reader we repeat Figure 9. 78
50 Example of the growth of multiplications with the number of states N and length of the sequence T. There can be seen that a lot of the sequences follow largely the same path, and thus can be calculated more efficiently with the forward algorithm (Patterson, 2020a). 78
51 To calculate the value of the backward variable at time t, you only need the values of the forward variables for all states at time $t+1$. In this figure q are the values of the distinct states, a_{ij} is the probability of transitioning from state i to j, and β is the backward variable. 80
52 Visualization of the calculation of ξ. 82
53 Prediction of the US versus against historical recession data and probability of being in state i (= 0 to 1) according to the model. The blue line in the top graph shows when the historical recessions were. The black line shows when the model predicts a recession. At best, these two graphs would overlap exactly, or be exactly mirrored. This is because the model sometimes gives the recession state a 0 and the other time a 1. The other graphs in the Figure indicate the probability according to the model that a country is in a particular hidden state at that moment. 83
54 Prediction of the UK versus against historical recession data and probability of being in state i (= 0 to 1) according to the model. The blue line in the top graph shows when the historical recessions were. The black line shows when the model predicts a recession. At best, these two graphs would overlap exactly, or be exactly mirrored. This is because the model sometimes gives the recession state a 0 and the other time a 1. The other graphs in the Figure indicate the probability according to the model that a country is in a particular hidden state at that moment. 84
55 Prediction of the US versus against historical recession data and probability of being in state i (= 0 to 1) according to the model. The blue line in the top graph shows when the historical recessions were. The black line shows when the model predicts a recession. At best, these two graphs would overlap exactly, or be exactly mirrored. This is because the model sometimes gives the recession state a 0 and the other time a 1. The other graphs in Figure indicate the probability according to the model that a country is in a particular hidden state at that moment. 89

56 Prediction of the UK versus against historical recession data and probability of being in state i (= 0 to 1) according to the model. The blue line in the top graph shows when the historical recessions were. The black line shows when the model predicts a recession. At best, these two graphs would overlap exactly, or be exactly mirrored. This is because the model sometimes gives the recession state a 0 and the other time a 1. The other graphs in Figure indicate the probability according to the model that a country is in a particular hidden state at that moment. 90
57 Prediction of the US versus historical recession data and probability of being in state i (= 0 to 2) according to the model. The blue line in the top graph shows when the historical recessions were. The black line shows when the model predicts a recession. The other graphs in Figure indicate the probability according to the model that a country is in a particular hidden state at that moment. 96
58 Prediction of the UK versus historical recession data and probability of being in state i (= 0 to 2) according to the model. The blue line in the top graph shows when the historical recessions were. The black line shows when the model predicts a recession. The other graphs in Figure indicate the probability according to the model that a country is in the particular hidden state at that moment. 97
59 Prediction of the JP versus historical recession data and probability of being in state i (= 0 to 2) according to the model. The blue line in the top graph shows when the historical recessions were. The black line shows when the model predicts a recession. The other graphs in Figure indicate the probability according to the model that a country is in the particular hidden state at that moment. 98

List of Tables

1 Overview of the treatment of sub-questions per subject and chapter. 7
2 Matrices HMM weather example. 18
3 Length of T per country. 38
4 Results of the base model compared to classifying all data points as no recession, classifying all data points with negative growth rate as recession, and classifying the next state as the current state. 39
5 Results of the improved 2 state HMM compared to classifying all data points as no recession, classifying all data points with negative growth rate as recession, classifying the next state as the current state, and the 2 state base model. The run defines the experiment run, this makes it easier to find the corresponding parameters of the results in Section 7.4. 48
6 Results of the 3 state HMM compared to the other methods and models. The run defines the experiment run, this makes it easier to find the corresponding parameters of the results in Section 7.5.2. 56
7 Results of the AIC and BIC for the base model, the two-state model, and the three-state model. 66
8 Results performance metrics for the experiments of the United States, United Kingdom, and Japan for the two-state hidden Markov model. 85
9 Parameters used for all experiments of the 2 state hidden Markov model. 88
10 Results of all experiments for the 3 state hidden Markov model. 90
11 Parameters used for all experiments of the 3 state hidden Markov model. 94

Contents

Management summary **iii**

Preface **iv**

Glossary **v**

Acronyms and abbreviations **vi**

Mathematical notations used **vii**

1 Introduction **1**
1.1 Introduction to the company 1
1.2 Project context 1
1.3 Background about recessions 2
1.3.1 What is a recession? 3
1.3.2 What causes a recession? 3
1.4 Research Purpose and relevance 3
1.5 Problem definition 5
1.6 Research Questions 5
1.6.1 Main question 6
1.6.2 Sub-questions 6
1.7 Workflow 7

2 Theoretical framework: machine learning **9**
2.1 Machine Learning 9
2.2 Supervised Learning 9
2.2.1 Classification 11
2.2.2 Regression 11
2.2.3 Training and test sets 12
2.3 Unsupervised Learning 12
2.3.1 Association rule learning 13
2.3.2 Dimensionality Reduction 14
2.3.3 Clustering methods 14
2.3.4 Hidden Markov Model 14
2.4 Algorithm selection 15

3 Theoretical framework: hidden Markov models **17**
3.1 The weather example 17
3.2 A general Hidden Markov model 18
3.2.1 Assumptions of the hidden Markov model 18
3.2.2 The hidden Markov model with a discrete observable distribution 19
3.2.3 Illustration of the hidden Markov model and types of model topologies 20
3.2.4 The hidden Markov model with a continuous observation distribution 20
3.3 The three fundamental problems of the hidden Markov model 22
3.3.1 Solution to the decoding problem 22
3.4 Conclusion 26

4 Methodology **27**
4.1 Data preparation . . . 27
4.1.1 Observable processes used . . . 27
4.1.2 Missing data . . . 28
4.2 The application of the three fundamental problems that can be solved with the hidden Markov model . . . 29
4.3 Validation method . . . 30
4.3.1 Benchmarks . . . 33
4.4 Correlation between the economic state of countries . . . 35
4.5 Summary . . . 36

5 Results **37**
5.1 The two-state hidden Markov base model . . . 37
5.1.1 Initialization of the two-state base model . . . 37
5.1.2 Results of the two-state base model . . . 39
5.1.3 Conclusions of the two-state base model . . . 46
5.2 Optimizing the two-state hidden Markov model . . . 46
5.2.1 Cutoff points, simple moving average, and exponential moving average . . . 47
5.2.2 Results of the improved two-state model . . . 48
5.2.3 Conclusions of the improved two-state model . . . 54
5.3 The three-state hidden Markov model . . . 56
5.3.1 Results of the three-state model . . . 56
5.3.2 Conclusions of the three-state model . . . 63
5.4 The five-state hidden Markov model . . . 64
5.5 AIC and BIC of the hidden Markov models . . . 65
5.6 Summary . . . 66

6 Conclusion, discussion, and further research **68**
6.1 Conclusion . . . 68
6.2 Discussion . . . 72
6.3 Contributions to theory and practice . . . 72
6.4 Recommendations for further research . . . 73

7 Appendix **77**
7.1 Solution to the likelihood problem . . . 77
7.2 Solution to the learning problem . . . 80
7.3 The two-state hidden Markov base model . . . 83
7.3.1 Visual results . . . 83
7.4 The improved two-state hidden Markov model . . . 85
7.4.1 Quantitative results of all experiments . . . 85
7.4.2 Parameters used for all experiments . . . 87
7.4.3 Visual results of best experiments . . . 89
7.5 The three-state hidden Markov model . . . 90
7.5.1 Quantitative results of all experiments . . . 90
7.5.2 Parameters used for all experiments . . . 94
7.5.3 Visual results of best experiments . . . 96

1 Introduction

In this chapter we provide background information on the company and department where this research is conducted, we explain the impact of macroeconomics, and we discuss the definition and causes of recessions. In addition, the aim and relevance of the research are discussed. Based on the aim and relevance of the research, the main research question and the sub-questions are formulated. Finally, the workflow of this research is discussed.

1.1 Introduction to the company

EY (Ernst & Young) is an internationally operating service company active in the field of accountancy, tax advice, and consultancy. Until June 30, 2013, the company was known as Ernst & Young. The company has approximately 312,000 employees worldwide and is located in about 150 countries. Together with PwC, KPMG, and Deloitte, EY is also referred to as one of the Big Four. In the Netherlands, EY consists of more than 5000 employees of which around 900 are active in Consulting. My research has been carried out within Technology Consulting, and specifically within the team 'Digital & Emerging Technologies' (D&ET).

D&ET provides services focused on the use of both new and existing technologies in the financial services sector. The Digital and Emerging Technology team mainly does projects in ServiceNow, a strategic partner of EY, financial crime, and cloud services. ServiceNow is an American software company based in Santa Clara, California that develops a cloud computing platform to help businesses manage digital workflows for business operations. EY's own IT department is one of ServiceNow's largest customers. Furthermore, EY uses ServiceNow to deliver creative solutions to customer problems, making the relationship beneficial for both parties. Considering the work in financial crime; there are strict rules for banks to prevent money laundering. Banks often have great difficulty meeting all these requirements. For instance, the Dutch bank ING was fined EUR 775 million in September 2018 for negligence in combating money laundering (ANP, 2020). EY helps banks with projects to combat money laundering and financial crime. Next to these examples, there is also a lot of collaboration with the other Technology Consulting teams such as Cyber Security, Data and Analytics, and Technology Transformation, and the teams within Business Consulting.

1.2 Project context

In the past years, several events have had a significant impact on the economy. The first example is the coronavirus. The COVID-19 crisis is expected to lead to the deepest global recession since World War II, and the most synchronized ever, with a record number of countries posting negative GDP per capita growth in 2020 (Buysse & Essers, 2020), Figure 1 depicts this information.

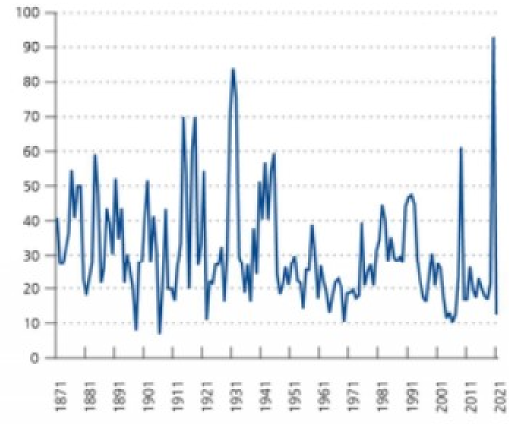

Figure 1: Percentage of countries in a recession between 1871 and 2021. Note that the percentage of countries in a recession is the highest during the COVID-crisis (Buysse & Essers, 2020).

Another example of an event that had a significant effect on the economy in Europe, is the United Kingdom which left the European Union on January 31, 2020 (the so-called Brexit) (NU.nl, 2022a). After 47 years, the UK's membership in the EU and its predecessor, the European Communities, came to an end. The Brexit had a significant impact on the UK itself, which lost 7400 jobs in the financial sector in Londen after leaving the European Union (EY, 2021). In addition, the war in Ukraine has had a major impact on the economy. The war in Ukraine not only leads to a higher price for energy, but also a decline in world trade and investment (NU.nl, 2022b).

We analyze the state of the economy with various economic indicators that tell us about the overall health of the economy. macroeconomics is important for consumers, firms as well as governments, to name a few examples (Hall, 2021):

- Consumers want to know how easy it will be to find work, how much it will cost to buy goods and services in the market, or how much it may cost to borrow money.
- Businesses use macro-economic analysis to determine whether expanding production will be welcomed by the market. Will consumers have enough money to buy the products, or will the products sit on shelves and collect dust?
- Governments turn to macroeconomics budgeting spending, creating taxes, deciding on interest rates, and making policy decisions.

Machine learning can add value when predicting macro-economic outcomes. Despite the growing interest in machine learning, little progress has been made in understanding the properties of machine learning models and procedures when they are applied to predict macro-economic outcomes (Coulombe et al., 2020).

1.3 Background about recessions

Here we provide background information on recessions and their causes. We also make a choice on which definition of a recession is used in this research.

1.3.1 What is a recession?

This section defines a recession as used in this thesis. A recession is a macro-economic term that refers to a significant decline in general economic activity in a designated region. It had been typically recognized as two consecutive quarters of economic decline, as reflected by GDP in conjunction with monthly indicators such as a rise in unemployment (Anderson, 2022). However, the National Bureau of Economic Research (NBER), which officially declares recessions, says the two consecutive quarters of decline in real GDP is not how it is defined anymore. The NBER defines a recession as a significant decline in economic activity spread across the economy, lasting more than a few months, visible in real GDP most of the time, income, employment, and industrial production (NBER, 2008). As can be seen, determining a recession remains quite subjective. In this research, the definition of the NBER is used. First, this is the agency that officially declared recessions. Second, this includes more relevant information than just GDP.

1.3.2 What causes a recession?

There are several views and theories on what causes a recession. Some economists believe that real changes and structural shifts in industries best explain when and how economic recessions occur, like a sudden, sustained spike in oil prices due to a geopolitical crisis might trigger a widespread recession (Anderson, 2022). A good example of this is the war between Russia and Ukraine mentioned earlier in this section. Another example of the type of economic shock is the spread of the COVID-19 epidemic and the resulting public health lock-downs in the economy in 2020. A recession could also be triggered by a revolutionary new technology that rapidly makes industries obsolete (Anderson, 2022). Some theories explain recessions as dependent on financial factors. These theories focus on either the over-expansion of financial risk during the good economic times preceding the recession (Anderson, 2022).

1.4 Research Purpose and relevance

The research relevance consists of the practical relevance for the company, in this case, EY, and the theoretical relevance. We first discuss the practical research relevance for EY, which has been determined in collaboration with EY's quantitative advisory team, after which the theoretical relevance of the thesis is discussed.

The International Financial Reporting Standards (IFRS) standard requires financial institutions (or other companies with financial assets like loans) to estimate potential credit losses with a forward-looking view. Although the standard does not prescribe any specific ways of doing so, most financial institutions (77%) are taking a scenario-based approach to include forward-looking macro-economic impact (EBA, 2021). The weights of such scenarios are often quite basic where the most likely scenario (baseline) accounts for 50% and the remaining two (upside and downside) share the other 50% equally. Some banks mentioned using only a single forward-looking scenario (i.e the most likely scenario) which does not meet the requirements of IFRS9 unless there is an adoption of an adjustment to reflect non-linearity in the credit loss distribution for alternative scenarios (EBA, 2021). Figure 2 depicts the approaches used by financial institutions for evaluating the possible outcomes in the expected credit loss amount.

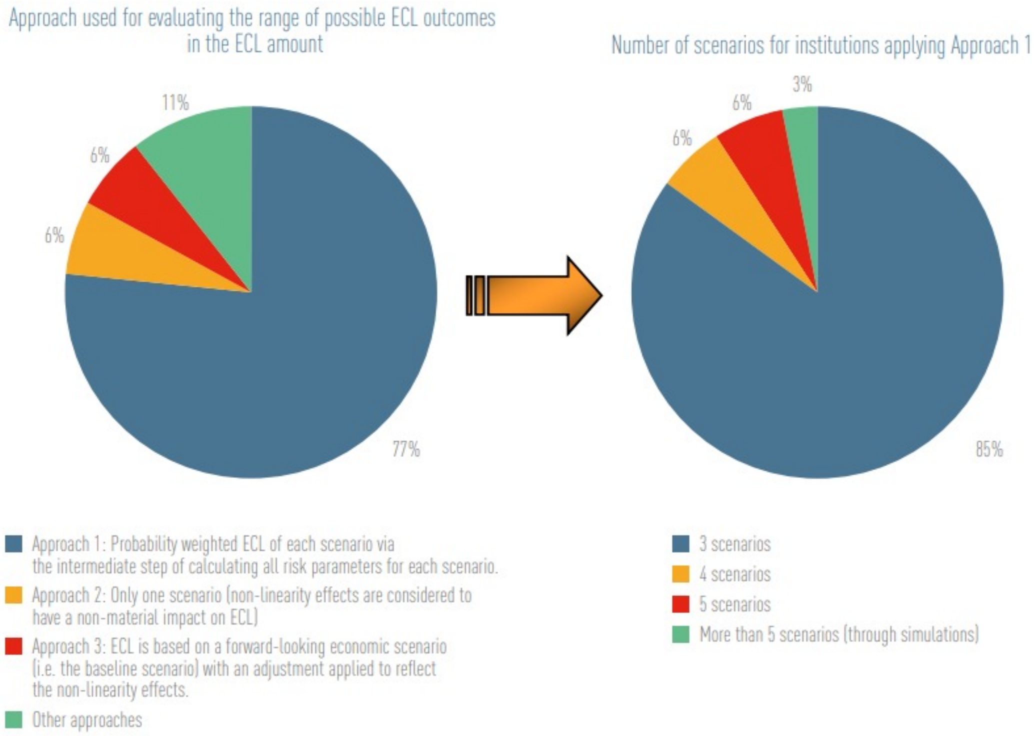

Figure 2: Approach used for evaluating the range of possible outcomes in the expected credit loss amount (EBA, 2021). Note that most financial institutions use a scenario based approach with 3 scenarios

Figure 3 shows schematically how most financial institutions take the macro-economic impact on expected credit losses into account. The numbers are the aforementioned weights, which are not determined with a quantitative method. Hence, in this research, we look for a quantitative method that can determine these probabilities.

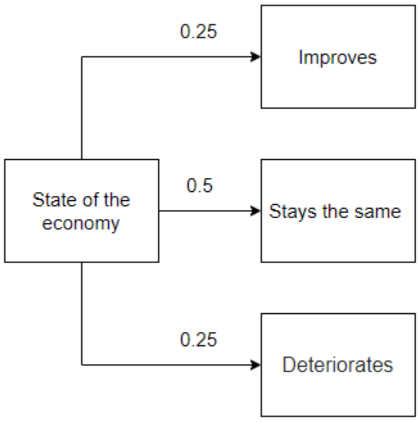

Figure 3: Schematic representation of the scenario weights currently used by most financial institutions to include macro-economic impact in their expected credit loss calculations. The numbers above the arrows represent the probabilities to move to the particular scenario the arrow points to.

We now discuss the theoretical relevance; right now mostly econometric models are used to calculate macro-economic predictions. For example, the Central Planning Bureau (CPB) in the Netherlands uses a combination of models together called the Sapphire 3.0 which is used for macro-economic predictions and scenario analysis (CPB, 2021). The model has a twofold purpose. First, it is used for projections of key economic indicators in the short and medium run (one to five years). Secondly, the model is used for simulation and evaluation of macro-economic effects of fiscal policy, like in the assessment of election platforms of Dutch political parties. For both purposes, CPB applies several models. Detailed results from government budget models as well as budget models and micro-econometric simulation for the labor market are used as input for the macro model (CPB, 2021). After contacting CPB it turned out that they do not use machine learning in the calculations made with the sapphire 3.0 model. However, they have just started using machine learning models for the unemployment estimate. In addition, there is little to no literature on predicting economic states through machine learning. For these reasons, the research is interesting both practically and theoretically.

1.5 Problem definition

As mentioned in the Section 1.4, estimating the forward-looking macro-economic impact for credit loss estimation is very difficult. Most financial institutions (including EY) use a scenario-based approach and weights that have not been established using quantitative models. The aim of this research is therefore to find a quantitative method to predict the economic states and to determine the probability (weight) of the transition into an economic state, will remain in an economic state, using the economic variables inflation, unemployment rate, GDP, market index and interest rate, which are chosen in consultation with EY.

1.6 Research Questions

The main goal of this research is to answer to which extent machine learning can be applied in predicting the economic state of countries. We formulate the main research question here:

1.6.1 Main question

In what way and to what extent can machine learning be applied using the economic features inflation, GDP, unemployment rate, market index, and interest rate for the prediction of the economic state of countries over a monthly time frame defining the performance of the model by the performance metrics accuracy, precision, recall, and the F-score?

The parts of the main question are be briefly explained here. 'In what way' in the main research question is a reference to the machine learning model which is deemed most suitable for this research. Furthermore, we want to take the constraint of the definition of a recession into account. We have defined a recession with the definition of the NBER, as described in Section 1.3. To what extent is related to the performance of the model, we want to address what has required for a good performance of the model. For instance, what are the requirements for the use of data and parameters. The economic features have been chosen in consultation with EY, EY would like to see whether the economic states of countries can be predicted using these economic features. The economic state mainly refers to the recession state. We develop models using 2, 3, and 5 states for the hidden Markov model. This research is mainly concerned with identifying the recession state(s), since this is the only economic state that can be validated because the other economic states do not have historical data available.

1.6.2 Sub-questions

The first step is to find out which machine learning algorithms are best suited for the research purpose. Based on literature research, the choice is be made. Here, the limitations of the research, mainly related to the availability of data, must be taken into account. For example, for most countries, there are no historical recession data in line with the definition of NBER. In addition, for most countries no data on the economic features are available and for some countries only data with a limited time frame. We also want our machine learning algorithm to be able to indicate with what probability it will end up in the same or a different state, for example, what is the probability that a country will be in recession next month given the state of the country at the moment. These considerations will determine the choice for the algorithm. Furthermore, it is interesting to see how correlated the economic states of countries are. For example, is it true that countries within a continent are more correlated than countries that are not in the same continent?

Then we come to the use of data. In consultation with the EY, the economic features mentioned in the main question were chosen as inputs. However, this data have to be acquired from somewhere. In addition, we should consider how much historical data should be used in order to acquire the best performance of the model. Missing data must also be handled correctly. It is therefore important that we define a way to validate the model. In this way, we can gain insight into the performance of the model and experiment with parameters and the inclusion of the amount of past data to see if we can improve the performance.

Furthermore, we define the performance using the performance metrics accuracy, precision, recall, and the F-score. These performance metrics are discussed in Section 4.3. An experiment is to vary the number of states. These are varied between 2, 3, and 5. It is interesting to see what the influence of the number of states is on the performance of the model. We expect that increasing the number of hidden states will increase the performance of the model, but will make the labeling of the states more complex. The expectation is therefore that the increase in performance will be at the expense of interpretability and explainability.

Finally, we would like to advise on possible further research that can improve the performance of the model. Table 1.6.2 provides an overview of the chapters in which these sub-questions are treated.

1. *Which machine learning algorithms can predict recessions according to the definition of the NBER and can calculate the probability of transitioning between economic states?*
2. *What is the mathematical concept of the hidden Markov model?*
3. *How can we address missing data and use data pre-processing methods to increase the performance of the model?*
4. *Which performance metrics can be used to validate our model?*
5. *What are the correlations between the economic state sequences generated by the hidden Markov model of the countries and how do they differ between the number of economic states (2,3 and 5) that are used in this research?*
6. *How do the results defined by the performance metrics accuracy, precision, recall, and the F-score differ with different number of states used for the hidden Markov model?*
7. *What further research can be performed to increase the performance of the hidden Markov model?*

Table 1: Overview of the treatment of sub-questions per subject and chapter.

Sub-question	**Chapters**	**Subject**
1	Chapter 2	Machine learning
2	Chapter 3	Hidden Markov models
3	Chapter 4	Data preparation
4	Chapter 4	Validation
5	Chapter 5	Results
6	Chapter 6	Conclusion
7	Chapter 6	Suggestions for further research

1.7 Workflow

In this section, we discuss the steps that are taken during this research. Figure 4 visualizes these steps. The first step is to see which machine learning methods may be suitable for answering the main research question. When the most suitable method is chosen, the theoretical framework of this method is discussed in detail, as well as its applications and assumptions.

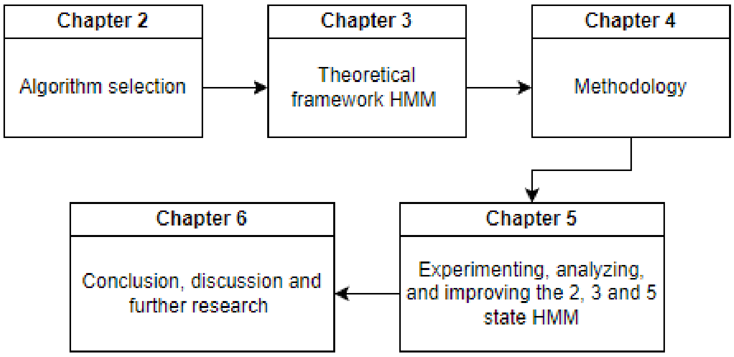

Figure 4: Chronological steps taken during this research.

.

The hidden Markov model seems to be the most suitable for this research since it meets all requirements for this research, this is explained in detail in Chapter 2. Therefore the next step is explaining the theory, assumptions, types, and problems the hidden Markov model can solve in the theoretical framework. The theoretical framework is provided in Chapter 3.

The methodology is explained in Chapter 4. We start by discussing how data is collected and how missing data is dealt with. In addition, we explain which parameters are used and how the model is validated. Lastly, we discuss how the correlation is calculated between the (economic) state sequences, that are generated by the model for different countries.

In Chapter 5 we experiment with the hidden Markov model. We then start with the two-state hidden Markov model as the base model. We start with this for the following reasons:

- The states of the model are easier to decipher than with a hidden Markov model with more states. In this case, you most likely have a recession state and a no recession state.
- By analyzing the base model, possible areas of improvement can be found. These areas of improvement can be experimented with to see if the performance increases.

After the results of the base model have been analyzed, it is possible to experiment with possible improvements. After this has happened, the model can be extended to the three-state and five-state models. Finally, in Chapter 6.1 we conclude, discuss limitations, and provide recommendations for further research.

2 Theoretical framework: machine learning

This chapter we want to achieve two objectives:

1. To select a machine learning algorithm that meets the goals, limitations, and restrictions of the research. In Section 1.7 is already mentioned that we will use the hidden Markov models during this research, in this chapter we provide the reasoning for choosing the HMM.
2. To familiarize the reader with the concept of machine learning, the different methods (supervised and unsupervised), and the algorithms that could potentially be used for this research.

First of all, a small recap of the aim of the research, the restrictions, and limitations. Our research aims to accurately determine economic states. While selecting the most suitable algorithm, we also have to take into account several limitations and restrictions. The most important of these is that the NBER does not have historical recession data available for many countries. Secondly, there are also only a limited number of countries with historical data on the economic features of inflation, unemployment rate, GDP, market index, and interest rate. In addition to these limitations, we also want the model to be able to calculate transition probabilities from moving between the economic states and staying in the current state.

In this chapter, we first explain what machine learning entails. Then we discuss the algorithms that have the potential for our research. Finally, we choose the most suitable method. This chosen method is explained in detail in the next Chapter.

2.1 Machine Learning

The most suitable algorithms are discussed in this chapter. Before the algorithms are explained we elaborate on some of the basic concepts of machine learning, to give a bit of background.

Artificial intelligence (AI) is the ability of a computer program or machine to exhibit or mimic human-like behavior (for example, visual senses, speech recognition, decision-making, natural language understanding, and so on) (Microsoft, 2022). However, this does not mean that all artificial intelligence methods aim to replicate human behavior. Machine learning is a subfield of AI and is concerned with extracting knowledge from data. It is a research field at the intersection of statistics, artificial intelligence, and computer science and is also known as predictive analytics or statistical learning (Müller & Guido, 2016). Figure 5 gives an overview of the most common algorithms used in machine learning. The supervised and unsupervised learning subdivisions are particularly interesting for this research and therefore be further discussed in this chapter.

2.2 Supervised Learning

Supervised learning is the most commonly used type of machine learning (Müller & Guido, 2016). Machine learning algorithms that learn from input/output pairs are called supervised learning algorithms because a 'teacher' provides supervision to the algorithms in the form of the desired outputs for each example that they learn from (Müller & Guido, 2016). For example, if we want to predict the eye color of a group of people, and we have a data set with their actual eye colors, we can verify the performance of our model. This is an example of what is called supervised machine learning. While creating a data set of inputs and outputs can be a time-consuming process, supervised learning algorithms are well understood and their performance is easy to measure. If your application can be formulated as a supervised learning problem, and there is a data set available that includes

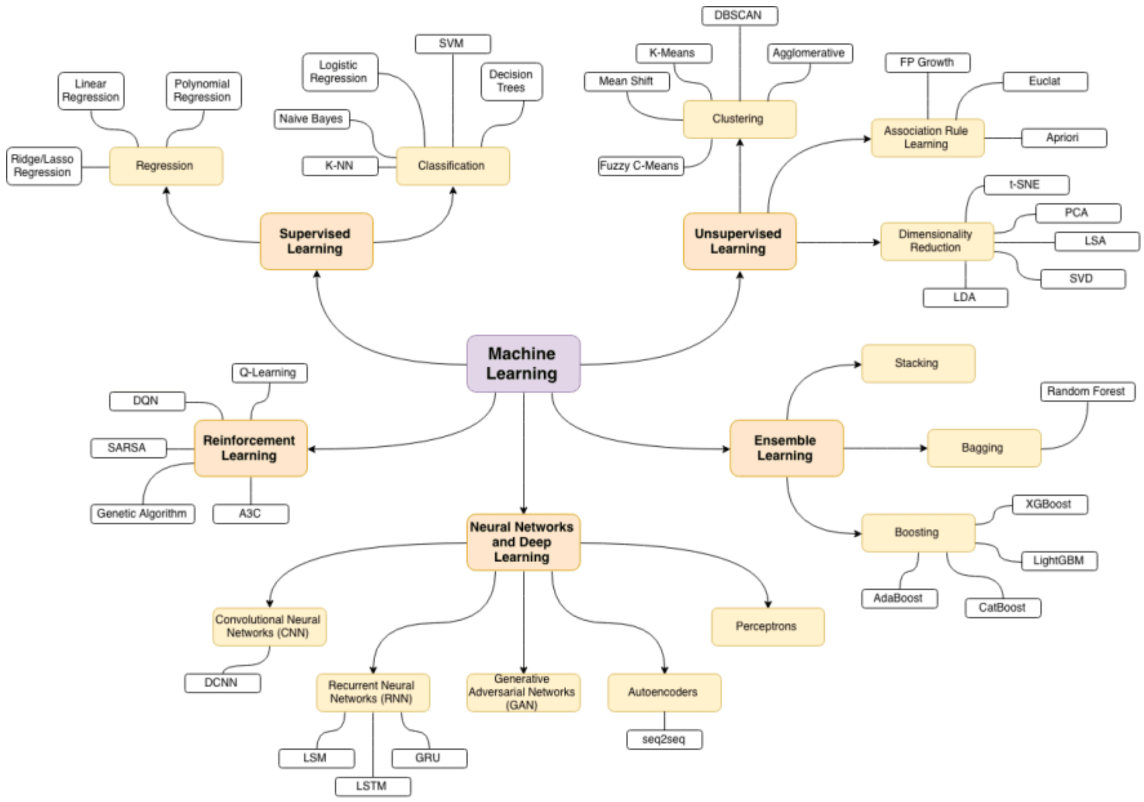

Figure 5: Machine Learning Algorithm Mind Map (Kumar, 2020).

the desired outcome, machine learning will likely be able to solve your problem (Müller & Guido, 2016). Some advantages of supervised learning are (Joy (2022), ODSC (2020)):

- Supervised learning is relatively easy to understand.
- If it is possible to be very specific about the definition of the classes, you can train the classifier in a way that has a perfect decision boundary to distinguish different classes accurately.
- You do not necessarily have to store the training data. In some cases, you can just use the decision boundary as a mathematical formula.
- Supervised learning can be very helpful in classification problems and regression problems (predicting a numerical target value from some given data and labels).

There are of course also some disadvantages to supervised machine learning (Joy (2022), ODSC (2020)):

- Supervised learning is limited in the sense that it is not able to handle some complex tasks in machine learning. Suppose that what we want to predict does not have a clear label or number, this problem cannot be solved by a supervised machine learning algorithm.
- Supervised learning can not obtain unknown information from the training data as unsupervised learning can.

- In the case of classification, supervised learning will not be able to classify an unknown class correctly. For instance, if we trained an image classifier on cats and dogs data, the classifier will classify other animals as a cat or dogs, which is incorrect.
- When training a classifier, a lot of data are needed from each class to get good results for your model.
- Often training needs a lot of computation time.

There are two types of supervised machine learning methods, called classification and regression, these methods are discussed in Sections 2.2.1 and 2.2.2.

2.2.1 Classification

In classification, the goal is to predict a class label, which is a choice from a predefined list of possibilities (Müller & Guido, 2016). For example, suppose we want to predict whether a company will go bankrupt. Then the outcome would be bankruptcy or no bankruptcy, with bankruptcy and no bankruptcy as class labels. Classification is sometimes separated into binary classification, which is the special case of distinguishing between exactly two classes, and multi-class classification, which is a classification between more than two classes (Müller & Guido, 2016). As such, predicting the bankruptcy of companies is an example of binary classification. Suppose we want to predict whether the type of fruit is strawberry, lime, or apple, this would be an example of multi-class classification. We now discuss some commonly used classification methods.

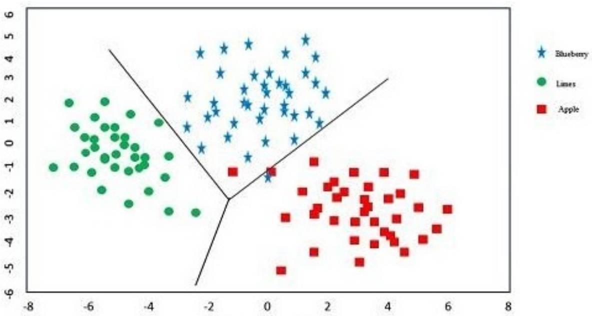

Figure 6: Visualization of the classification type of supervised learning. Note that the different classes are separated by a decision boundary (ODSC, 2020).

2.2.2 Regression

For regression tasks, the goal is to predict a continuous number (Müller & Guido, 2016). For example, if we want to predict the value of several commodities such as oil, gold, and grain, we are engaged in regression.

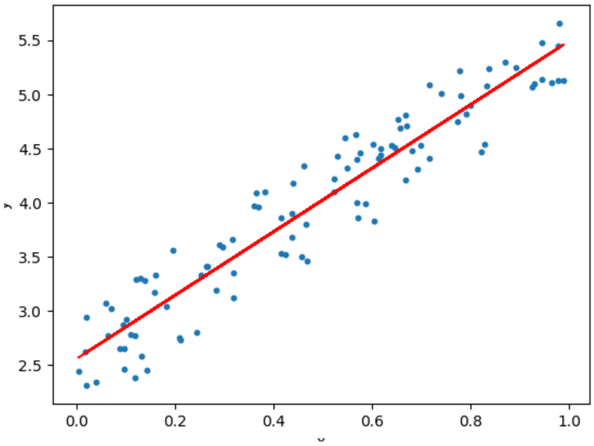

Figure 7: Visualization of the regression type of supervised learning. Note that the regression type identifies a pattern in the data (ODSC, 2020).

2.2.3 Training and test sets

Often, training and test sets are used for validating a supervised machine learning model. With training and test sets, one part of the data is used to build the machine learning model, and it is called the training data or training set. The other part of the data is used to assess how well the model works; this is called the test data or test set. Most of the time the training set contains 75 percent of the data and the test set contains the remaining 25 percent (Müller & Guido, 2016). Before making the split, the data set should be shuffled using a pseudo-random number generator, otherwise, the data points are often already clustered by features. Furthermore, we need to ensure that we will get the same output if we run the same algorithm several times, therefore we should provide the pseudo-random number generator with a fixed seed value.

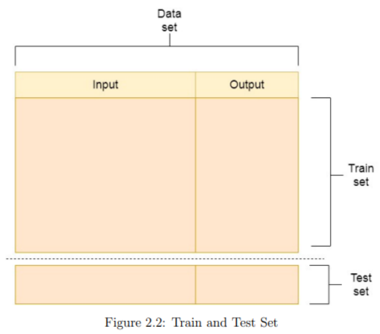

Figure 8: Schematic representation training and test set (Joy, 2022).

2.3 Unsupervised Learning

The main goal of unsupervised learning is to discover hidden and interesting patterns in unlabeled data. Unlike supervised learning, unsupervised learning methods cannot be directly applied to regression or a classification problem as one has no idea what the values for the output might be (Škoda & Adam, 2020). Some advantages of unsupervised learning are (Gareth et al. (2021), Joy, (2022):

- It can see what humans can not identify. It can be used to find hidden patterns in data that supervised machine learning methods can not.
- There is often less data pre-processing to be done since it is not required (or possible) to label the data. Labeling data often demands a lot of work and time.
- The labels can be added after the data have been classified which is often easier.
- It is easier to obtain unlabeled data most of the time.

Some disadvantages of unsupervised learning are (Joy (2022), Müller & Guido (2016), Gareth et al. (2021):

- It is often more time-consuming than supervised machine learning as it might require human intervention to understand the patterns and correlate them with the domain knowledge.
- It is not always certain that the obtained results will be useful since there is no label or output measure to confirm their usefulness.
- It is often more complex to understand how the model's output is generated and how to validate it. There is no universally accepted mechanism for performing cross-validation or validating results on an independent data set.
- The results often have less accuracy.

There are three main types of unsupervised machine learning methods: association rule learning, dimensionality reduction, and clustering.

2.3.1 Association rule learning

Association rule learning is a rule-based machine learning method for discovering interesting relations between variables in large databases. It is intended to identify strong rules discovered in databases using some measures of how interesting the relationship is (Piatetsky-Shapiro, 1991). Association rule learning is one often used in market basket analysis. To name an example, in a store all vegetables are placed in the same aisle, all dairy items are placed together and cosmetics form another set. However, the rules do not extract an individual's preference but find relationships between a set of elements of every distinct transaction. With this machine learning method relationships between products can be found and the store can then act upon those relationships. These rules may find that some products are often bought together and some are not. There are various metrics in place to indicate the strength of association between these two (Garg, 2018):

$$Support = \frac{\text{Transactions containing both X and Y}}{\text{Total number of transactions}}, \tag{1}$$

$$Confidence = \frac{\text{Transactions containing both X and Y}}{\text{Transactions containing X}}, \tag{2}$$

$$Lift = \frac{\text{Confidence}}{\text{Fraction of transactions containing Y}}. \tag{3}$$

2.3.2 Dimensionality Reduction

Dimensionality reduction refers to techniques for reducing the number of input variables in training data. Fewer input dimensions often mean fewer parameters or a simpler structure in the machine learning model. A model with a complex structure is more likely to overfit the training data set and therefore may not perform well on new data. Dimensionality reduction yields a more compact representation of the target concept, focusing the user's attention on the most relevant variables (Witten et al., 2016) . For instance, images can include thousands of pixels, not all of which matter to your analysis (Castañón, 2019), in this case, dimensionality reduction algorithms are a good solution to make the data set manageable. The most commonly used dimensionality reduction algorithms are t-distributed stochastic Neighbor embedding (t-SNE), principal component analysis (PCA), latent semantic analysis (LSA), singular value decomposition (SVD), and linear discriminant analysis (LDA).

2.3.3 Clustering methods

Clustering refers to a very broad set of techniques for finding subgroups, or clusters, in a data set (Gareth et al., 2021) . When we cluster observations from a data set, we want to do this in distinct groups so that the observations within each group are quite similar, but the observations from other groups are very different. For this method to work, we must define what it means for two or more observations to be similar or different (Gareth et al., 2021) . An application for clustering is marketing. We may have access to a large number of features containing information about a great number of people such as median household income, occupation, location, and so forth. For a marketing company, it is important to target the right customers, so we want to perform market segmentation by identifying subgroups of people who might be more receptive to a particular form of advertising.

2.3.4 Hidden Markov Model

Before we explain the Hidden Markov Model, some background knowledge is given. We start with the Markov Chain. A Markov Chain or Markov model is a special type of discrete stochastic process in which the probability of an event occurring only depends on the immediately previous event. This means, that if we know the present state of a system or feature we do not need any past information to try to predict the future state or values. Markov chains are defined by a set of states and the transition probabilities between each state.

A hidden Markov model (HMM) is a statistical Markov model in which the system is assumed to be a Markov process, with unobservable (hidden) states. HMM requires that there is an observable process whose outcomes are influenced by the outcomes of the unobservable process in a known way. Our goal is to learn the unobservable process using the observable process. In this research, the economic states (recession and no recession for the two-state model) are the unobservable processes. These unobservable processes, according to the definition of the NBER, cannot be determined with an exact value or class label. However, with our observable processes (inflation, unemployment rate, GDP, market index, and interest rate) we can try to learn the unobservable process. For a detailed explanation of the operation, assumptions, and problems that can be solved using the hidden Markov model, we refer to Chapter 3, in that chapter the hidden Markov model is described in detail.

2.4 Algorithm selection

In this section, we discuss the choice of which algorithm is used for this research. The research aims to predict economic states based on economic variables and calculate economic state transition probabilities which can be used as weights for the scenarios. Since the recession state is the only economic state with historical data available we validate by comparing the historical recession data with the output of the chosen model.

The definition of a recession is important for the choice of the algorithm, we therefore recap it: a significant decline in economic activity spread across the economy, lasting more than a few months, visible in real GDP most of the time, income, employment, and industrial production (NBER, 2008). Because the definition of the NBER is not an exact number, we can not regression to identify the recession periods. If the definition of a recession was only dependent on a numerical value this would have been possible. Furthermore, the definition is not a type of exact class. In addition, the historical recession data that is in line with the definition of the NBER is only available for a few countries. Therefore the supervised learning algorithms can not be used directly. However, it is possible to create economic labels. This could theoretically be possible, however, since the definition is not exact, creating the economic labels would be a difficult process and subjective process. To illustrate this, it will be difficult and subjective to define a significant decline in economic activity, GDP, and income. There is no theory, regarding creating recession labels, and therefore many assumptions would have to be made. Also, a recession in Japan can have other values for economic variables than a recession in the US. So, you would also have to take a country-specific approach to create the labels. Take into account that all these considerations have to be done for all three economic states and not just the recession state. It is, therefore, a lot more convenient if there are methods available that can divide periods into economic states without having to label them.

We, therefore, have to examine the unsupervised learning methods. Association rule learning is not applicable for research since it is not able to identify the economic periods as well as give transition probabilities between the economic periods. It could only be used to analyse the association between the country's state sequences and economic features. Dimensionality reduction and clustering methods are probably able to identify the economic states by clustering them. With clustering, it would even be able to find the state transition probabilities by creating a time series and inferring probabilities from them. However, this is one of the main problems that can be solved with a hidden Markov model. Because of this, it is more convenient to use the hidden Markov model since we want to use the three-state HMM for one of the fundamental problems which it can solve.

Hidden Markov models appear to be the best solution to the problem because they meet all requirements for this research:

- The hidden Markov model can divide periods into different states (even without a label), and should therefore be able to recognize the recession state(s) with the chosen features.
- The hidden Markov model can provide information about the transition probabilities. In other words, the hidden Markov model can assign a probability to the change or preservation of a state. It can therefore say how likely it is that a country will go into another state next month (or another time frame) based on which state it is in right now. This is the true essence in which the hidden Markov model distinguishes itself from the other unsupervised machine learning algorithms.
- The hidden Markov model can generate state sequences (the economic states) as output based on the observation sequences (the features). The state sequences of the countries can then

be used to create correlation matrices. In this way, it is possible to calculate the correlations between the state sequences of countries at the same point in time. These correlations can provide additional inside into the macro-economic dependencies between countries and continents. Furthermore, with the use of machine learning, the missing values for the features could be supplemented with the values of countries that have a high correlation in the economic state sequence with the country concerned.

3 Theoretical framework: hidden Markov models

To provide a background on hidden Markov models, we discuss several topics. First, We explain hidden Markov models with an example about the weather. When we have provided intuition of how the hidden Markov model works, we work out the mathematical concept of the hidden Markov model, the assumptions on which the model is based, and the different types of model typologies. After which we discuss the elements of the hidden Markov model with a discrete observation probability distribution and the hidden Markov model with a continuous observation probability distribution. Then it is time to elaborate on the problems that can be solved with the hidden Markov model and how this is possible. Finally, we conclude the chapter by how the applications of the hidden Markov models are used during this research.

3.1 The weather example

To familiarize the reader with the problem, we start by explaining the HMMs further with a weather example. Consider two friends, Bob and Alice, Bob lives in Enschede and Alice in Amsterdam. They talk over the phone daily to discuss what they did that day. Bob is interested in the following activities: walking in the park, shopping, and cleaning his apartment. The choice of what to do is determined exclusively by the weather on a given day. Alice does not have information about the weather in Enschede but based on what Bob did that day she tries to guess what the weather must have been like. Alice believes that the weather operates as a discrete Markov chain, with two states: 'Rainy' and 'Sunny', but she cannot observe them directly (they are hidden from her) (Wikipedia, 2022).

On each day, there is a certain probability that Bob will perform one of the aforementioned activities based on the weather, these activities are the observations. The probabilities of going from one hidden state to another are called the transition probabilities, in our example, these are the probabilities of going from state 'Sunny' to 'Rainy' and the other way around, and the probability of staying at a stage. The probabilities from the observations to states are called the emission probabilities, the probability that the observations are emitted. The entire system is called the hidden Markov model. Alice knows the general weather trends in the area and what Bob likes to do on average. Figure 9 depicts the results.

A hidden Markov model consists of two stochastic processes. The first stochastic process is that of the unobservable (hidden) states. In our weather example, this is the weather in another city. If we live in Amsterdam, we can not observe the weather in Enschede. The other stochastic process is the observable process, which can be observed directly, in our example these are the activities that are undertaken: walking, shopping, or cleaning.

The transition probabilities can be summarized with the state transition matrix. The probabilistic relation between the observations and states is given by the observation matrix. Lastly, we have the initial state distribution which resembles the start. The matrices corresponding to our example are given in chronological order in the Table 2. In the matrices given in Table 2, the 'S' stands for 'Sunny' and the 'R' for 'Rainy'. Furthermore, the 'W' stands for 'Walking', 'Sh' for 'Shopping', and the 'C' for 'Cleaning'. The matrices are row stochastic, meaning that each element is a probability and the elements of each row sum to 1, that is, each row is a probability distribution (Stamp, 2004). Now that the operation of the HMM has become clearer based on the weather example, we will now discuss the technical aspects of the HMM.

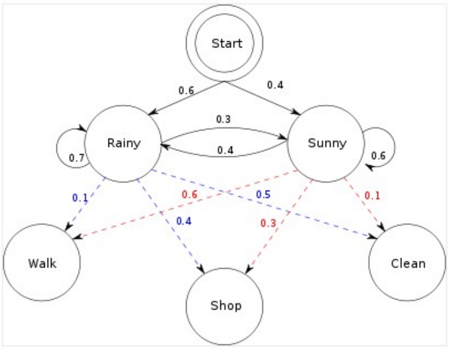

Figure 9: Schematic representation of the HMM weather guessing game. The start indicates with which probability the sequence starts in which state. So there is a 0.6 probability that the first state is 'Rainy' and 0.4 that it is 'Sunny'. The black lines between the states 'Rainy' and 'Sunny' indicate the state transition probabilities. The circles with 'Walk', 'Shop', and 'Clean' indicate the observable processes and the dotted lines indicate the emission probabilities (Wikipedia, 2022).

Table 2: Matrices HMM weather example.

A =

	S	**R**
S	0.6	0.4
R	0.3	0.7

B =

	W	**Sh**	**C**
S	0.6	0.3	0.1
R	.0.1	0.4	0.5

π =

S	**R**
0.6	0.4

3.2 A general Hidden Markov model

In this section, we describe the basic elements, notations, and assumptions of a general HMM. Since our example from the previous section is that of a discrete observation process, we first define the notation and structure of the discrete HMM. After this, we give an extension to the hidden Markov model with continuous observable signals.

3.2.1 Assumptions of the hidden Markov model

An HMM makes two independence assumptions (Jurafsky & Martin, (2021), Sutton & McCallum, (2012):

1. **Markov Assumption / limited history hypothesis:** Each state depends only on its predecessor. Formally: $P(x_t \mid x_{t-1}, ..., x_0) = P(x_t \mid x_{t-1})$, where x_t is the state at time t, x_{t-1} is the state at the preceding state and x_0 is the first state in the state sequence.

2. **Output Independence / stationary hypothesis:** Each observation variable depends only on the current state and not on any other states or any other observations (Xiao, Liu, & Wang, 2005). Furthermore, the state transition function is irrelative to the time when a state transition occurs.
Formally: $P(O_t \mid x_0, ..., x_t, O_0, ..., O_t) = P(O_t \mid x_t)$, where O_t is the observation at time t, x_t is the state at time t.

Stationarity means that the statistical properties of a process generating a time series do not change over time (Palachy, 2019).

3.2.2 The hidden Markov model with a discrete observable distribution

Below the mathematical notations of the discrete hidden Markov model are given, these are the mathematical notations used in the field of HMM and its applications. To make the reader more familiar with these notations, they are explained using the weather example. A basic, discrete HMM is characterized by the following elements (Stamp, 2004):

- T = length of observation and state sequence.
- $\mathcal{O}$ = $(O_0, O_1, ..., O_{T-1})$ = the observation sequence.
- $\mathcal{X}$ = $(x_0, x_1, ..., x_{T-1})$ = the state sequence.
- N = a finite number of unobservable states.
- M = the number of distinct observation symbols.
- $\mathcal{S}$ = $\{q_0, q_1, ..., q_{N-1}\}$ = distinct states of the Markov process.
- $\mathcal{V}$ = $\{v_0, q_1, ..., v_{M-1}\}$ = distinct observable signals of the Markov process.
- A = the state transition distribution.
- B = the observation probability distribution.
- π = $p(x_i = i)$ for $i = 1...N$ = the initial state distribution.

An HMM models a sequence of observations $\mathcal{O} = (O_0, O_1, ..., O_{T-1})$ by assuming that there is an underlying sequence of states $\mathcal{X} = (x_0, x_1, ..., x_{T-1})$. In our example about the weather, the observation sequences are the activities that have been undertaken, i.e. walking, shopping, or cleaning. The state sequences are the weather conditions, so sunny or rainy. The individual states are elements of the state space $\mathcal{S}$. A state space is the set of all possible configurations of a system (Nykamp, 2019). A sample path is a particular realisation of the process. We can define different sample paths based on the sample space. So in our weather example, the state space would be $\mathcal{S} = \{\text{Sunny, Rainy}\}$. Note that $\mathcal{S}$ and $\mathcal{X}$ are not the same, x_0 and x_1 can be the same value since it represents the state that we are in given the clock time t. $\mathcal{S}$ on the other hand represents the possible states that the HMM can be in, therefore q_0 and q_1 can not be the same. $\mathcal{V}$ is the set of possible observations, in our weather example, this results in set $\mathcal{V} = \{\text{Walking, Shopping, Cleaning}\}$. An example of what the observation sequence and state sequence (sample paths) could be is given below:

$$\mathcal{O} = (O_0, O_1, O_2, O_3, O_4) = (\text{Walking, Shopping, Shopping, Cleaning, Walking}) \tag{4}$$

$$\mathcal{X} = (x_0, x_1, x_2, x_3, x_4) = (\text{Sunny, Sunny, Sunny, Rainy, Sunny}) \tag{5}$$

The length of or observation sequence and state sequence is 5 as can be seen from Equations 4 and 5 ($T = 5$). Equations 4 and 5 are sample paths, a sample path of a stochastic process. The order and length of these sample paths can differ. A sample path is any set of possible values to which the appropriate random variables might map a given point in the sample space. The number of unobservable states in our weather example is 2 ($N = 2$) because the weather in Enschede can be ‘Sunny’ or ‘Rainy’ in our example. The number of distinct observation symbols in our weather example is 3 ($M = 3$), namely ‘Walking’, ‘Shopping’ or ‘Cleaning’.

The state transition distribution, the observation probability distribution, and the initial state distribution are shown in Table 2. The state transition matrix A has a length of $N * N$ where the elements in the matrix are given by:

$$a_{ij} = P(x_{t+1} = q_j \mid x_t = q_i), \quad 1 \leq i, \quad j \leq N. \tag{6}$$

The observation probability distribution B for the discrete HMM is $N * M$, where the elements in the matrix are given by:

$$b_j(k) = P(O_t = v_k \mid x_t = q_j), \quad 1 \leq j \leq N, \quad 1 \leq k \leq M. \tag{7}$$

Furthermore, both matrices are row stochastic which means that $\sum_{j=1}^{N} a_{ij} = 1$ and $\sum_{j=1}^{N} b_j(k) = 1$ and the probabilities a_{ij} and $b_j(k)$ are independent of t. This means that the HMM assumes that all probabilities hold at all points in time, this is also called the time-homogeneous Markov process. In summarizing, we note that under the aforementioned assumptions the HMM requires three probability distributions (Sutton & McCallum, 2012):

1. π: The distribution over the initial states.
2. A: The transition distribution.
3. B: The observation distribution.

Concluding, An HMM is defined by A, B, and π (and implicitly, by the dimensions N and M) (Stamp, 2004). The HMM is denoted by: $\lambda = (A, B, \pi)$.

3.2.3 Illustration of the hidden Markov model and types of model topologies

Figure 10 illustrates a hidden Markov model, where X_t represents the hidden state sequence and all other notation is given above. The Markov process, which is hidden behind the dashed line, is determined by the current state and transition probabilities A. As can be seen, we can only observe $\mathcal{O}$, which are related to the (hidden) states of the Markov process by the observation probabilities B. There are three main types of model topologies. The first one is ergodic, in this model topology,

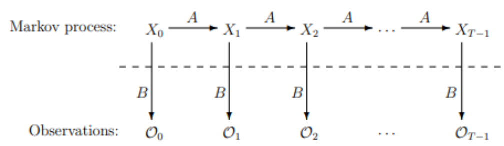

Figure 10: Illustration hidden Markov model (Stamp, 2004).

we allow transitions to any state at any time. The second one is the most commonly used type of HMM for sequential modeling. The third one is the linear model topology. In this model topology, it is not permitted to jump states. Figure 11 shows a visual representation of the three different model topologies. The one that applies to our weather example is the ergodic one depicted in Figure 11, this is also the one that is used during our research. There are other ergodic model typologies but this is the one that is used in our research since it is possible to transition between any economic state at any time.

3.2.4 The hidden Markov model with a continuous observation distribution

Earlier in this chapter, the hidden Markov model with a discrete observation distribution was discussed, but there are also applications of the hidden Markov model with continuous observation distributions. In this research, we also use a hidden Markov model with a continuous observation distribution for the emission probabilities.

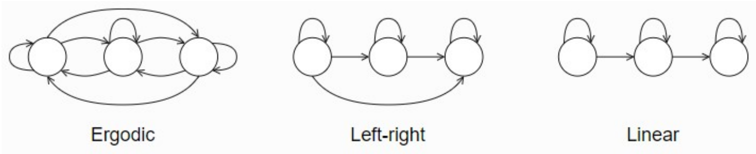

Figure 11: Illustration model topologies of the hidden Markov model.

In a continuous hidden Markov model the observation probabilities $b_j(k)$ are replaced by a continuous density $b_j(x)(1 \leq j \leq N)$, one for every state in the Markov chain), where x is some observation vector that is being modeled. The most general probability density function for the continuous HMM is defined as a finite Gaussian mixture distribution of the form:

$$b_j(x) = \sum_{m=1}^{M_j} c_{jm} \mathcal{N}(x \mid \mu_{jm}, \sum_{jm}). \tag{8}$$

Here c_{jm} is the mixture coefficient for the mth mixture in state j, $\mathcal{N}$ is the Gaussian density, with mean vector μ_{jm} and covariance matrix $\sum_{jm}$, for the mth mixture in state j. The mixture weights satisfy the following conditions:

$$\sum_{m=1}^{M} c_{jm} = 1, \quad 1 \leq j \leq N, \quad c_{jm} \geq 0, \quad 1 \leq m \leq M. \tag{9}$$

So that the probability density function is properly normalized, so that:

$$\int_{-\infty}^{\infty} b_j(x) dx = 1, \quad 1 \leq j \leq N. \tag{10}$$

The Gaussian mixture distribution given in Equation 8, can be used to approximate any finite continuous density function. This makes it an excellent method to be applied to many problems. Figure 12 depicts an example of a Gaussian mixture density.

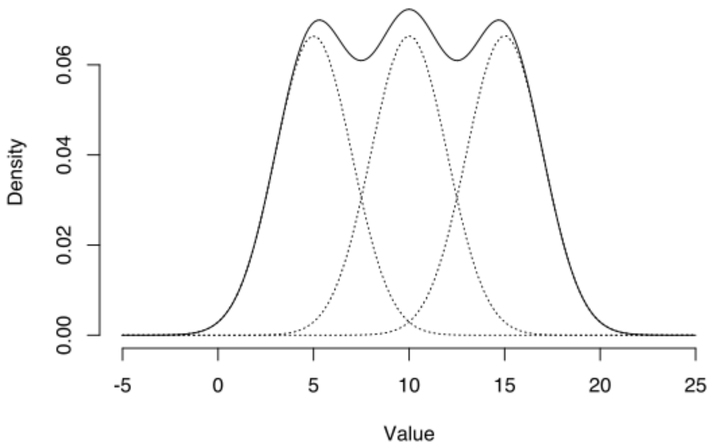

Figure 12: Example of a Gaussian mixture probability density. The density consists of three normal distributions ($\mu_1 = 5, \mu_2 = 10, \mu_3 = 15$) with equal weights.

3.3 The three fundamental problems of the hidden Markov model

Three fundamental problems can be solved using HMMs (Stamp, (2004), Jurafsky & Martin, (2021)). We first briefly describe the three problems, after which we give the efficient algorithms for their solution. The three fundamental problems are:

1. **Likelihood:** Given $\lambda = (A, B, \pi)$ and a sequence of observations $\mathcal{O}$, find $P(\mathcal{O} \mid \lambda)$. Here we want to determine a score for the observed sequence $\mathcal{O}$ concerning the given model λ. Problem 1 is also called the evaluation problem. It answers the following question: Given a model and a sequence of observations, how do we compute the probability that the observed sequence was produced by the model? An application for which the likelihood problem could be used is to understand whether a particular sentence was written by an author. The likelihood problem allows you to choose the best match among competing models. With the likelihood problem, we can adjust the parameters of the hidden Markov model (think of the initial state distribution, state transition probabilities, observation distribution and the number of states), and compare in which model the observations are most likely.

2. **Decoding**: Given $\lambda = (A, B, \pi)$ and an observation sequence $\mathcal{O}$, find an optimal state sequence for the underlying Markov process. So, what sequence of states best explains a sequence of observations? In other words, we want to uncover the hidden part of the Hidden Markov Model. The decoding problem is used in our weather example. Given the observations sequence given in Equation 4, transition distribution A, observation distribution B, and initial state distribution π, the model generates a state sequence that it considers most probable, for example, this could be the state sequence given in Equation 5. The decoding problem is a big part of this research. We use the economic features, i.e. the observation sequences inflation, unemployment rate, GDP, market index, and interest rate to see what the hidden Markov model considers the most probable economic state sequences. Normally there is no correct solution for the decoding problem. Since the actual results are normally not known when applying the hidden Markov model. Think of the weather example where we cannot know the weather in the other city. However, we looked up the historical recession periods for the countries the United States, the United Kingdom, and Japan to be able to compare these with the state sequence output of the hidden Markov model, this way the performance of the HMM can be validated.

3. **Learning:** Given an observation sequence $\mathcal{O}$ and the dimensions N and M, find the model λ = (A, B, π) that maximizes the probability of $\mathcal{O}$. So given a set of observation sequences how do we learn the model probabilities that would generate them? This can be viewed as training a model to best fit the observed data.

In the next section, the decoding problem will be discussed. This is because it contained the most important part of the research and because we want to keep the research readable and not overwhelmed with mathematical elaborations. The likelihood problem and learning problem are used during this research, for the solutions to these problems are discussed in detail in Sections 7.1 and 7.2.

3.3.1 Solution to the decoding problem

Given the model λ = (A, B, π) and a sequence of observations $\mathcal{O}=(O_1, O_2, ..., O_T)$, our goal is to find the most likely state sequence $\mathcal{X} = (x_1, x_2, ..., x_T)$. This can be done by two approaches:

1. Maximizing the individually most likely states.

2. Maximizing the sequences of states (path) as a single entity.

In the context of this research this would be given the observation sequences inflation, unemployment rate, GDP, market index and interest rate the model tries to find the most likely (economic) state sequence (the second approach). In our weather example we would look for the most logical weather (state) sequence based on the activities (observations) 'Walking', 'Shopping', and 'Cleaning'. We start by explaining the first approach and end with why it is not optimal. Below we explain the second approach, which is used in this research.

Individually most likely states
Mathematically finding the most likely individual states given the model is given by the following equation:

$$\gamma_t(j) = P(x_t = q_j \mid \mathcal{O}, \lambda). \quad 1 \leq j \leq N, \quad 1 \leq t \leq T. \tag{11}$$

Where $\gamma_t(j)$ is the probability of being in state q_j at time t. Remember that x_t is the state of the state sequence at time t. We can use the work we did with the forward and backward variable to solve this question:

$$\gamma_t(j) = P(x_t = q_j \mid \mathcal{O}, \lambda) = \frac{P(x_t = q_j, \mathcal{O}, \lambda)}{P(\mathcal{O}, \lambda)} = \frac{\alpha_t(j)\beta_t(j)}{\sum_{i=1}^{N} \alpha_t(i)\beta_t(i)}, \quad \sum_{j=1}^{N} \gamma_t(j) = 1. \tag{12}$$

Remember that α_t is the probability of everything that comes before t and β_t is the probability of everything that comes after t. We now have to search over all $\gamma_t(j)$ to find the most likely state sequence. We find the probability of the most likely state sequence by dividing the value of the most likely state sequence by the sum of all possible state sequences. Figure 13 depicts this process. Hence, we look over all the states, 1 to N, and what is the one that maximizes the probability. The value of this state is what we call i_t^*, which can be found using the following equation:

$$i_t^* = \arg\max_j [\gamma_t(j)], \quad 1 \leq t \leq T, \quad 1 \leq j \leq N. \tag{13}$$

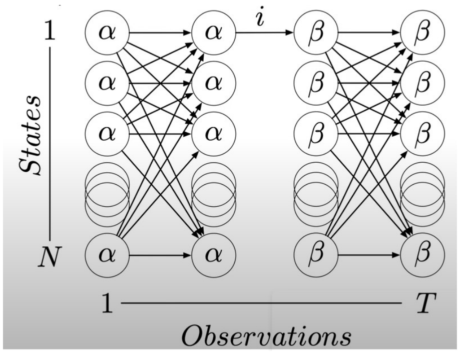

Figure 13: Visualization of combining the α and β to find the state sequence. We find the probability of the most likely state sequence by dividing the most likely state sequence path, by all possible paths (Patterson, 2020b).

This approach allows us to maximize the expected number of correct states but the result might not be in line with the properties of the hidden Markov model. Because the hidden Markov model is a model that deals specifically with sequential data, the current time t that we are in is dependent on the previous time $t-1$ that we were in. How it is now solved by using γ_t is to think of each step independently. In other words, the optimal state sequence might allow for impossible state transitions ($a_{ij} = 0$) for some individually optimal i, j. We explain what is meant here with an example, which is depicted in Figure 14. Consider a robot vacuum cleaner, this vacuum cleaner is used to vacuum the kitchen and living room. Between q_1 and q_2 is a door through which the vacuum cleaner can go from the kitchen to the living room. This is the only way this is possible. The method just discussed in this section could give a transition from q_5 to q_1, when it actually wasn't possible, the transition probability between those states is 0 (because there is no door).

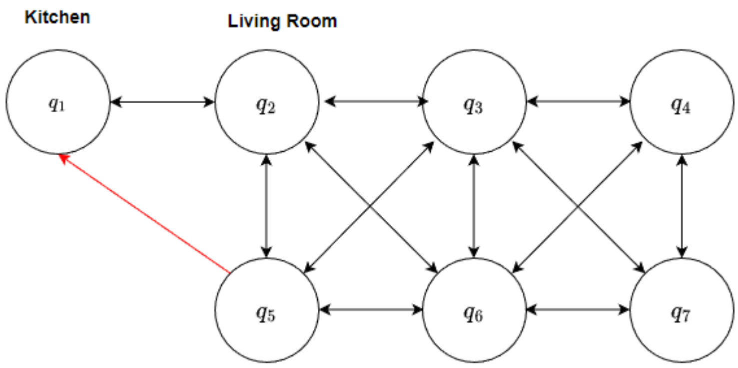

Figure 14: Schematic visualization of the example of the robot vacuum cleaner. The black arrows indicate the transitions the robot can make between the possible states q_i. The red arrow indicates an impossible transition (transition probability is equal to zero between q_5 to q_1) that can be returned with the method discussed in this section.

The Viterbi algorithm

Because looking at the individually most likely states can lead to impossible state transitions, we consider the second approach, which is looking at the sequence of states as a single entity. With this approach, we want to find the sequence of states (path), whose conditional probability as a whole is optimal given some observation sequence.

We want to find the path sequence that maximizes $P(\mathcal{X} \mid \mathcal{O}, \lambda)$. Since we are interested in maximizing the combination of the state sequence and observation sequence we can say that this is equivalent to the probability of a state sequence and an observation sequence given our model:

$$P(\mathcal{X} \mid \mathcal{O}, \lambda) = P(\{x_1, x_2...x_T\}, \{O_1, O_2...O_T\} \mid \lambda). \tag{14}$$

Equation 14 can be solved using the Viterbi algorithm (Jurafsky & Martin, 2021), where $\delta_t(j)$, represents the probability that the HMM is in state j after seeing the first t observations and passing through the most probable state sequence $x_1, ..., x_{t-1}$, given the Hidden Markov Model λ. The value of each cell $\delta_t(j)$ is calculated recursively taking the most probable path that could lead us to this point. Each patrial most probable path expresses the probability (Jurafsky & Martin, 2021):

$$\delta_t(j) = \max_{x_1,...,x_{t-1}} [P(x_1, ...x_{t-1}, x_t = q_j, O_1, O_2, ...O_t, \mid \lambda)]. \tag{15}$$

Where $\delta_t(j)$ is the sequence of states that maximizes the probability of seeing that particular states and then ending up at state q_j at time t. In our weather example this would be finding the most probable weather sequence given the activities. If only the activities 'Shopping and 'Walking' occurred in the observation sequence, the most probable path consists probably of a lot of 'Sunny' states. For $t = 1$, Equation 15 in invalid because there are no preceding states. Still this can be solved using the probability of being in that state given that $t = 1$ and the observed signal O_1:

$$\delta_1(j) = P(x_1 = j, O_1) = \pi_j b_j(O_1). \tag{16}$$

We now find the best sequences of states is extending them by one, so going from t to $t+1$. What is the state that maximized δ at the previous time step i, and then considering all the ways to get from the previous state i to the next state j. Then the step to the next state is taken by multiplying with the state transition probability a_{ij}, the probability of moving from state i to state j. Lastly, we multiply with the probability of seeing the next observation at the current state $b_j(O_j)$.

$$\delta_{t+1}(j) = [\max_i \delta_t(i) a_{ij}] * b_j(O_{t+1}). \tag{17}$$

The difference between the forward and backward algorithm and the Viterbi algorithm is that we are now keeping track of where we came from at each time step to recreate the path. To keep track of the states that maximize the $\delta' s$, we are going to use a variable called ψ, which is defined by:

$$\psi_t(j) = arg \max_i [\delta_{t-1}(i) a_{ij}]. \tag{18}$$

The ψ answers the question: Given that I am here, with which path did I most likely arrive?'. The complete Viterbi algorithm is given by the following steps:

1. Initialization:

   $\delta_1(i) = \pi_i i(O_1), \quad 1 \leq i \leq N,$
   $\psi_1(i) = 0.$

2. Recursion:

   $\delta_t(j) = \max_{1 \leq i \leq N}[\delta_{t-1}(i) a_{ij}] b_j(O_t), \quad 2 \leq t \leq T, \quad 1 \leq j \leq N,$
   $\psi_t(j) = arg \max_{1 \leq i \leq N}[\delta_{t-1}(i) a_{ij}], \quad 2 \leq t \leq T, \quad 1 \leq j \leq N.$

3. Termination:

   $P^* = \max_{1 \leq i \leq N}[\delta_T(i)],$
   $i_t^* = \psi_{t+1}(i_{t+1}^*), \quad t = T-1, T-2, ..., 1,$
   $x_T^* = q_{i_T^*}.$

4. Path backtracking:

   $i_t^* = \psi_{t+1}(I_{t+1}^*), \quad t = T-1, T-2, ..., 1,$
   $x_t^* = q_{i_t^*}.$

Where, P^* denotes the total probability of the most likely path for some observation sequence $\mathcal{O}$, i_t^* is the index of the most likely state at time t, and x_t^* is the most likely state at time t of the most likely path. The Viterbi algorithm is a lot like the forward algorithm from the likelihood problem. The difference is that the forward algorithm takes the sum of the probabilities that could lead to that point resulting in the probability of arriving at that specific point and the Viterbi algorithm takes the maximum probability of arriving at that point (most likely path). This makes sense since in the likelihood problem we want the probability of an observation sequence occurring, so we sum over the probabilities. With the decoding problem, we want to know which state sequence is most likely given an observation sequence, therefore we take the path with the highest probability, therefore we maximize with the Viterbi algorithm.

3.4 Conclusion

A hidden Markov model is a Markov model in which the system is assumed to be a Markov process, with unobservable (hidden) states. The hidden Markov model makes 2 independence assumptions, the first one is that each (economic) state only depends on the previous state. To give an example, if there is a recession in January 2022, the economic state in February 2022 only depends on January 2022 and not on the economic state in the data points before. The second independence assumption is that each observation variable depends only on the current state and not on previous states or other observations. In the context of this study, that would mean the observable variable GDP only depends on the value of the current economic state (recession or no recession for the two-state model) and not on the economic state of the data points before or the values of the other features (inflation, unemployment rate, market index, and interest rate). This is of course questionable in practice and that is why we also use higher order hidden Markov models. With higher-order hidden Markov models more historical data points are used than only the previous data point.

Furthermore, we discussed the three fundamental problems the HMM can solve, the first one being the likelihood problem. The likelihood problem is used to determine the probability that the model has generated the sequence of observations. With the likelihood problem, we can adjust the parameters of the hidden Markov model (for instance the number of states) and compare in which model the observations are most likely. This is the application we use the likelihood problem for in this research. The likelihood score can not be compared directly but can be used to calculate the Akaike Information Criterion and the Bayesian Information Criterion, with which we can compare the models. The second problem which can be solved using an hidden Markov model is the decoding problem. The decoding problem calculates the optimal state sequence given sequences of observations. It is about uncovering the hidden part of the hidden Markov model. This is probably the most relevant application of the hidden Markov models in this research. We use the decoding problem to find the optimal sequence of economic states. We validate this with the recession state since this is the only economic state for which historical data is available. We want to confirm that the hidden Markov model can identify the economic states. We validate this by comparing the historical economic state sequences with the state sequences produced by the model. Lastly, there is the learning problem. With the learning problem, we try to learn the parameters of the HMM given an observation sequence and a possible set of states. We use the learning problem in our research to calculate the initial state distribution and the economic state transition probabilities. These economic state transition probabilities can be used as alternatives for the current scenario weights used to incorporate macro-economic impact in credit risk calculations, given that the HMM can identify the different economic states. In this research, the transition probabilities for the countries United States, United Kingdom, and Japan will be calculated for the three-state hidden Markov model.

4 Methodology

In this chapter, we discuss the methodology of this research. We start by discussing which data are used, how they are processed, and how missing data are handled. Next, we explain how we validate this research. Because the recession state varies per run of the model, additional steps have to be taken to calculate the (supervised learning) performance metrics accuracy, precision, recall, and the F-score. Furthermore, we discuss how to calculate the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Finally, we discuss how to calculate the correlation of the hidden Markov model between the state sequences of the different countries.

4.1 Data preparation

We could use monthly or yearly data for the research, we chose monthly data for the following reasons:

- First, there are not many historical data available on the economic features of inflation, unemployment rate, GDP, market index, and interest rate. This also applies to the historical recession dates that are in line with the definition of the NBER. If we take annual data, this means that we have 12 times fewer data points.
- Second, it is more interesting and challenging to forecast in a shorter time frame because the values of economic features differ less than with annual data.
- Third, there are of course recessions that end in the middle of the year. However, you cannot apply this nuance to an annual model and in these cases, a choice has to be made whether the year is classified as a recession or not.

4.1.1 Observable processes used

We need input to let the hidden Markov model generate the state sequences. The following economic features have been chosen in consultation with EY:

- Inflation rate.
- GDP return (growth rate).
- Unemployment rate.
- Market index returns.
- Interest rate return.

These economic features are therefore the observation sequences that are used for uncovering the hidden states of the hidden Markov model. In Section 5.2.1 there is experimentation with these features by evaluating the difference in performance when we prepare the data in different manners. the data preparation of some features. The data of the observation sequences are collected by EY and are obtained from the data provider Reuters. Furthermore, the following values for the parameters explained in the previous chapter apply:

- T = depends on the number of data points available per country.
- N = 2, because there is a recession state and a no recession state.
- M = 5 (# features).

- $\mathcal{S}$ $= \{$recession, no recession$\}$.
- $\mathcal{V}$ $= \{$inflation, unemployment rate, GDP, market index, interest rate$\}$.

As mentioned in the previous chapter T is the length of observation and state sequence, N is the finite number of unobservable states, M is the number of distinct observation symbols, $\mathcal{S}$ are the distinct states in the Markov process and $\mathcal{V}$ is the set of possible observations. Figure 15 depicts a visual representation of the two-state hidden Markov model in the context of this research. We

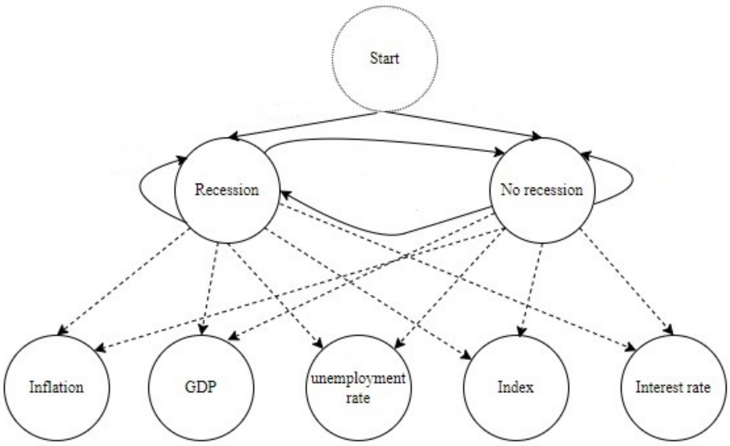

Figure 15: Schematic representation the two-state annual HMM for the US. The black arrows beginning at 'start' represent the initial state distribution, the black arrows originating from 'Recession' and 'No Recession' represent the state transition distribution, and the dotted lines the emission probabilities.

.

have a continuous hidden Markov model, these are given by the dotted lines. These observation probabilities are modeled by Gaussian mixture distributions, which are discussed in Section 3.2.4.

4.1.2 Missing data

Data collection is quite a challenge. The data for the features have been obtained per country from the Reuters data set. There are three problems with the data:

1. Not all features have the same number of historical data. For instance, GDP could be available from 1980 but the interest rate from 1990.
2. Some values are missing for several countries and GDP only quarterly data are available (so two missing values every quarter because we use monthly data).
3. In many cases, the market index has been around for a shorter period.

The above issues have been resolved in the following ways:

1. The model only starts predicting when all features are available. For example, for the US this will be from 1960, but for other countries only in 1980 or 2000.

2. The missing values are filled by the forward filling function in Python. Forward filling means filling missing values with previous data. The reason why we do not use interpolation is that the hidden Markov model takes the values of the features of time t to predict $t + 1$ and the moment we use interpolation the values of features of time $t + 1$ are already used for the prediction of the state $t + 1$. The same logic applies with backward filling, where we use the next data point to fill missing values. For GDP, our forward filling method results in three times the same value because only quarterly data is available. The first 2 rows have also been removed because they are missing for the GDP and because a row of data is lost during the data processing of the market index and the interest rate (since they are converted to relative changes).

3. The problem with the market index is solved by taking the returns or differences of the S&P500 when the market index of a particular country has less historical data available than the historical data of the rest of the features. In reality, this solution is rarely used, because most of the time the market index goes back further in time than the data of one of the other features. It is important to note that no data points are missing from the market index values of all countries.

4.2 The application of the three fundamental problems that can be solved with the hidden Markov model

In this section, we discuss how the specific applications of the hidden Markov models are applied in this research. We start with the first fundamental problem the hidden Markov model can solve; the likelihood problem.

As explained in Chapter 3 the likelihood problem allows you to choose the best match among competing models. With the likelihood problem, we can adjust the parameters of the hidden Markov model (for instance the number of states) and compare for which model the observations are most likely. This is also the application for we use the likelihood problem. We calculate the likelihood for hidden Markov models with different parameters. However, they can not be compared directly because the three-state model has far more parameters than the two-state model. We, therefore, use the likelihood to calculate the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC).

The Akaike information criterion is an estimator of out-of-sample prediction error and thereby relative quality of statistical models for a given set of data. Given a collection of models for the data, AIC estimates the quality of each model, relative to each of the other models. Thus, AIC provides a means for model selection (Zajic, 2019). The AIC is a number that can be used to determine which model is most likely relative to the other models. AIC is most frequently used where it is difficult to test the model's performance on a test set, for instance with a small data set or time series. It is important to note that the AIC adds a penalty term for the complexity of the model. The best AIC is the lowest one possible, which indicates the best balance between complexity and likelihood.

The BIC is closely related to the AIC, it is also based on the likelihood function and a criterion for model selection. The difference between the AIC and the BIC is that the penalty term of the BIC is greater than that of the AIC (Klassen, 2020). Another difference is that BIC also considers the number of observations in the model, which AIC does not. BIC is widely used for model identification in time series and linear regression, but the use of the BIC is not limited to these applications and can be used for any set of likelihood-based models (Klassen, 2020). The formulas for calculating the AIC and BIC for a hidden Markov model are given by:

$$AIC = -2 * log(L) + 2 * p \tag{19}$$

$$BIC = -2 * log(L) + p * log(T) \tag{20}$$

In these Equations, L is the likelihood of the model, p the number of independent parameters of the model, and T is the length of the time series. The number of independent parameters p can be calculated as follows:

$$p = m^2 + km - 1 \tag{21}$$

In Equation 21, m denotes the number of hidden states of the model, k a numeric value representing the number of parameters of the underlying distribution of the observation process. In our research k will be the sum of the number of features used times 2, since we use a Gaussian mixture distribution of 1 and a Gaussian mixture has a mean and standard deviation.

The second problem which can be solved using an hidden Markov model is the decoding problem. As explained in Chapter 3, the decoding problem calculates the optimal state sequence given sequences of observations. It is about uncovering the hidden part of the hidden Markov model. This probably the most relevant application of the hidden Markov models in this research. We use the decoding problem to find the optimal sequence of economic states. We validate this with the recession state, since this is the only economic state for which historical data is available.

Lastly, there is the learning problem. With the learning problem we try to learn the parameters of the HMM given an observation sequence and a possible set of states. We use the learning problem in our research to calculate the initial state distribution and the economic state transition probabilities. These economic state transition probabilities can be used as alternatives for the current scenario weights used to include macro-economic impact in credit risk calculations, given that the HMM is able to identify the different economic states. In this research the transition probabilities for the countries United States, United Kingdom, and Japan will be calculated for the three-state hidden Markov model.

4.3 Validation method

In this section, we describe how we apply validation during this research, we use several performance metrics in this research, which are explained in this section. The performance metrics for the countries United States, United Kingdom, and Japan are used to compare the output of the model with the historical recession data from the NBER, for the other countries this data is not available. The other countries are compared to the global recession periods, the following global recessions have occurred between 1960 and now :

- 1973 - 1975 (Kose et al., 2020).
- 1980 - beginning of 1983 (Kose et al., 2020).
- 1991 (Kose et al., 2020).
- 2008 - 2009 (Kose et al., 2020).
- 2020 (Mankiw, 2020).

The main goal of comparing to the global recession periods is to make the process of labeling the output of the model easier. For example, there is a greater chance that a country is in recession during a global recession, this way we can decipher more easily the recession states for the countries which do not have historical data available. Labeling the states is necessary for making the correlation matrices.

The main problem with validating our unsupervised learning algorithm with supervised learning performance metrics is that we do not know the recession state beforehand. If the recession state is 1, we can calculate the performance metrics the same way as with supervised learning algorithms. These performance metrics are calculated based on a confusion matrix. In the field of machine learning and specifically the problem of statistical classification, a confusion matrix, also known as an error matrix, is a specific table layout that allows visualization of the performance of an algorithm, typically a supervised learning one (Stehman, 1997). The name stems from the fact that it makes it easy to see whether the system is confusing two classes (i.e. commonly mislabeling one as another). In Figure 16 an example of a confusion matrix is given. There are 4 important terms (Mishra, 2018):

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)

Figure 16: Schematic representation of the confusion matrix.

- True Positives: The number of times in which the model predicted recession and the the actual economic state in that period was also a recession.
- True Negatives: The number of times in which the model predicted no recession and the the actual economic state in that period was also no recession.
- False Positives: The number of times in which the model predicted a recession but the actual economic state in that period was no recession.
- False Negatives: The number of times in which the model predicted no recession but the actual economic state in that period was a recession.

With this confusion matrix several performance metrics can be calculated (Mishra, 2018), which are discussed in this section. Accuracy is the ratio of several correct predictions to the total number of input samples, i.e.,

$$\text{Accuracy} = \frac{TP + TN}{P + N}. \tag{22}$$

Precision is the number of correct positive results divided by the number of positive results predicted by the classifier, i.e.,

$$\text{Precision} = \frac{TP}{TP + FP}. \tag{23}$$

Recall is the number of correct positive results divided by the number of all relevant samples (all samples that should have been identified as positive), i.e.,

$$\text{Recall} = \frac{TP}{TP + FN}. \tag{24}$$

The F-score is the harmonic mean of the precision and recall, i.e.,

$$\text{F-score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{25}$$

The greater the F-score, the better the performance of our model. If the recession state is 0 and not 1 we have to make adjustments to the equations. Since the state sequence is mirrored, the formulas are different. The performance metrics can still be determined, but the performance metrics are now given by the following equations:

$$Accuracy^* = \frac{FN + FP}{P + N} = 1 - Accuracy. \tag{26}$$

$$Precision^* = \frac{FN}{FN + TP}. \tag{27}$$

$$Recall^* = \frac{FN}{FN + TN}. \tag{28}$$

$$F - score^* = \frac{2 * Precision^* * Recall^*}{Precision^* + Recall^*}. \tag{29}$$

With the 3-state model there are two problems with the validation:

- First, as with the two-state model, it is not clear which state is the recession state. When the 3-state hidden Markov model is applied, it can be 0, 1, or 2, which is not the same for every run of the model.

- Second, there is now one class more than in the historical data set, where there are only 2 classes, namely a 1 and a 0. The 1 here stands for the recessions and the 0 for the no recession state.

We solve the first problem by calculating the performance metrics of all 3 scenarios; in case the recession state is 0, 1, or 2. The second problem we solve by converting the 3-state model to 3 different 2-state outputs. We validate the models with the scenario that any number can be in a recession state and then compare this with the historical recession data. In this historical recession data frame, 0 is the no recession state and 1 is the recession state. We then calculate the performance metrics for all scenarios, after which we choose the right model. Figure 17 schematically depicts this process. In short, we transform the results of the model from a multi-state output to a binary state output. This method could be extrapolated to a 4,5 or N state hidden Markov model.

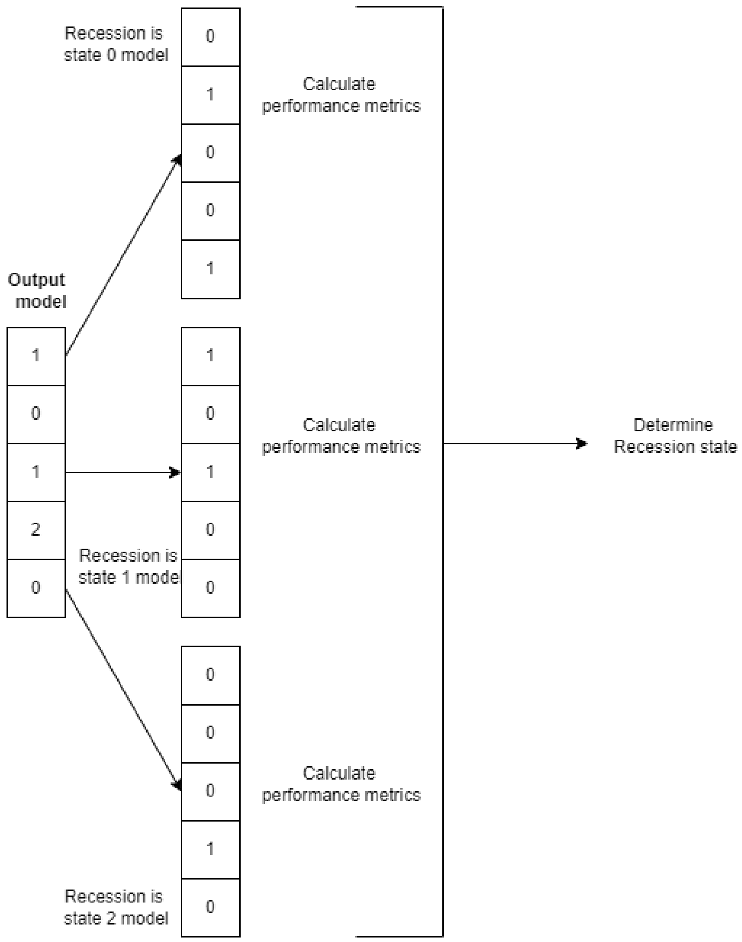

Figure 17: Schematic representation of the validation process for the 3-state hidden Markov model. The first column indicates the output of the model. After this, the performance metrics are calculated for all three possible recession state scenarios; In case 0, 1, or 2 is the recession state. Finally, it must be determined based on the performance metrics and graphs what the recession state is.

4.3.1 Benchmarks

We compare the results of the hidden Markov models with three benchmarks. The first is classifying all results as recessions, this can lead to high accuracy with unbalanced data, however for the other performance metrics this is not the case. Note that the accuracy of this method is the same as the initial state distribution for the two-state hidden Markov model, so when we reverse the method this results in 1 minus the accuracy of the other method. Classifying everything as a recession would lead to an accuracy of 0.14 for the US to name an example. The reason why we do not also include classifying everything as no-recession follows from this. Because of the data imbalance (there are more no recession periods than recession periods) classifying everything as no recession would lead to lower scores for most performance metrics. The second method is to classify all data points with a negative growth rate as a recession since negative growth rates are often associated with recessions.

The third method we use is the martingale method. A martingale is a mathematical series in which the best prediction for the next number is the current number (Yates, 2021). In our case, this results in using the current state x_t as a prediction for the next state x_{t+1}. The misclassified data points are quite intuitive to predict, namely the points that an economy moves from no recession to recession and vice versa. Based on this logic it can already be determined that the martingale method will lead to very high performance metrics because it will only misclassify the transitions between the economic states.

If you wanted to include the macro-economic impact on credit risk through a forward-looking scenario (only 6% of financial institutions (EBA, 2021) do this), the martingale method might be interesting, given that an adjustment is applied for the non-linearity effects. However, this research focuses on a multi-scenario approach, with a multi-scenario approach the martingale method cannot be applied. Our goal is to determine weights for the different scenarios that are used to include macro-economic impact in the credit risk calculations, as shown in Figure 18. However, the assumption with the martingale is that the current state with a probability of 1 is the next, so it is not a good alternative for calculating the scenario weights to include the macro-economic impact for the multi-scenario approach. Figure 19 depicts schematically the martingale method with the weights that would be used if this would be incorporated in macro-economic impact for credit risk calculations by financial institutions.

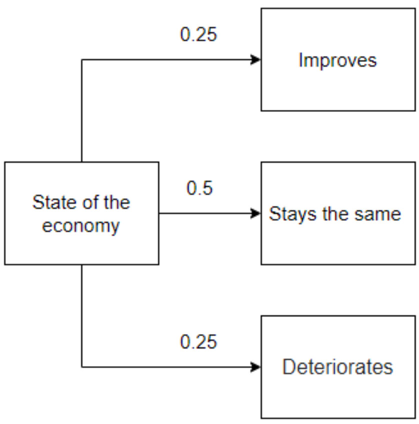

Figure 18: Schematic representation of the scenario weights currently used by most financial institutions to include macro-economic impact in their expected credit loss calculations. The numbers above the arrows represent the probabilities of moving to the particular scenario the arrow points to.

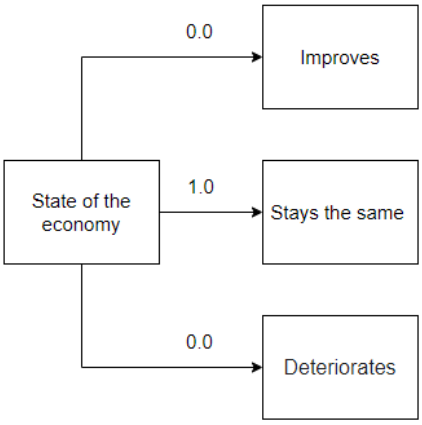

Figure 19: Schematic representation of the scenario weights used by the Martingale method. The numbers above the arrows represent the probabilities of moving to the particular scenario the arrow points to.

4.4 Correlation between the economic state of countries

In addition to identifying the economic states, we also want to evaluate the correlations between the state sequences of countries. Correlation is a way to determine if two variables are related (Lutes, 2020). In data science, the ρ value, also called Pearson's correlation coefficient is often used. This measures how closely two sequences of numbers are (linearly) correlated. The ρ-value ranges from -1 to 1. The closer to 1, the stronger the positive correlation, and the closer to -1, the stronger the negative correlation. If the value is close to 0 it means that the correlation is weak. The formula for Pearson's correlation coefficient is given(Lutes, 2020):

$$\rho = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{30}$$

Here x_i and y_i are the i-th variables and $\bar{x}$ and $\bar{y}$ are the estimates of the averages of the variables and n is the length of the sequences. To calculate Pearson's correlation coefficient, we compare the state sequences created by the model in a correlation matrix. It is important to note that the user of the model must identify himself based on the graphs, what the recession state is and what the no recession state is. Because we have the historical recession periods of the United States, the United Kingdom, and Japan, we can calculate the correlation and display it in a correlation matrix. This matrix can then be compared with the correlation output of the hidden Markov model. Figure 20 depicts the correlation matrix of the historical recession periods.

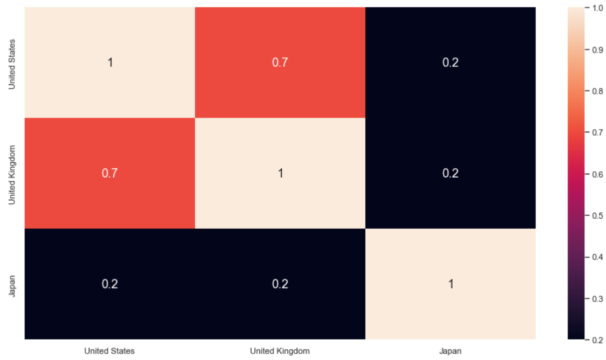

Figure 20: Correlation plot of the historical recession periods for the countries the United States, United Kingdom, and Japan.

.

Figure 20 shows that the correlation between the countries the United States and the United Kingdom is quite high (0.65). In addition, the correlation between the historical economic states of the United States and Japan (0.24) is higher than that of the United Kingdom and Japan (0.16). When we compare these values with the correlation at the same point in time of the state sequences produced by the model, this is an additional form of validation since if the model works well the correlation should also be close to the historical correlations.

4.5 Summary

In this chapter, we discuss the methodology. We explain how data are used and how we deal with missing values. Furthermore, we discuss how we apply the three fundamental problems that can be solved by the hidden Markov model in this research. We elaborate on the calculation of the AIC and the BIC, which can be used to estimate the relative quality between models. The likelihood problem is used to calculate these scores. The decoding problem is used to find the optimal hidden state sequence, we use this problem to find the most likely economic state sequence. Lastly, we use the learning problem to calculate the initial state distribution and (economic) state transition probabilities.

Next to this, we discuss how to validate the model. This is done using the performance metrics accuracy, precision, recall, and the F-score. For the two-state model, these performance metrics can be used in the normal way if the recession state is 1. When the recession state is 0, the economic state sequence is mirrored and therefore some changes in the formulas are needed to use the performance metrics. Using the three-state model, the recession state can be 0, 1, or 2. However, the recession state of the historical data is 1 and the non-recession state is 0. We turn our three-state model into a binary output if we want to use the performance metrics. We solve this by adjusting the output of the three-state model to 3 binary state outputs. We then calculate the performance metrics for all three binary state outputs. Based on the analysis of these performance metrics and the visual output of the model we determine which state is the recession state. Figure 17 schematically depicts this process. Finally, we discuss how to calculate the correlation at the same point of time between the state sequences of the countries.

5 Results

In this section, we discuss the results of this research. We start with the two-state hidden Markov base model. Here we experiment with the initialization of the initial state distribution and state transition distribution, and the number of Gaussian mixture distributions used for the emission probabilities. We generate results and analyze these results and based on these analyses, conclusions are drawn for possible areas for improvement.

Based on the conclusions of the base model, we experiment with possible methods that can improve the two-state model, after which the results of these experiments are discussed, and based on this we draw conclusions. Subse quently, the same methodology is applied to the three-state and five-state hidden Markov model; experiment, discuss and analyze results and finally draw conclusions.

5.1 The two-state hidden Markov base model

In this section, we generate the results for the base model, based on the analysis of these results we intend to find possible areas for improvement. We first give a recap of the parameters that are used, then we elaborate on the initial state distribution, transition probabilities, and the amount of Gaussian mixture distributions that are used for the base model. Next, we discuss the results of the base model and analyze these results. Finally, conclusions are drawn based on these analyses.

5.1.1 Initialization of the two-state base model

Parameters
In this section, we discuss the parameters that are used to generate the output of the 2 state base model. The observable processes that are used are:

- Inflation rate.
- GDP return (growth rate).
- Unemployment rate.
- Market index returns.
- Interest rate return.

The vectors which obtain the values of the economic features through time are therefore the observation sequences that are used for uncovering the hidden states of the hidden Markov model. As mentioned before T is the length of observation and state sequence, N is the finite number of unobservable states, M is the number of distinct observation symbols, $\mathcal{S}$ are the distinct states in the Markov process, and $\mathcal{V}$ is the set of possible observations. The following values for the parameters explained in the previous chapter apply:

- T = is given in Table 3.
- N = 2, because there is a recession state and a no recession state.
- M = 5, because there are 5 features used.
- $\mathcal{S}$ = {Recession, No Recession}.
- $\mathcal{V}$ = {Inflation, unemployment rate, GDP, market index, interest rate}.

Table 3: Length of T per country.

Country	US	CA	MX	NL	UK	FR	DE	ES
T	744	277	204	228	407	211	361	322

Country	IT	RU	CN	HK	JP	SG	KR	TW
T	211	193	205	288	331	269	255	277

Initial state distribution, state transition distribution, and the number of Gaussian mixture distributions

We start as described in Section 1.7 with the two-state base model. To give this base model the best possible performance, we have varied the following parameters:

- The initialization of the initial state distribution and state transition distribution.
- The number of Gaussian mixture distributions.
- How the features Market index and interest rate are included in the model.

We initialize the initial state distribution by giving $1 - x$ to a certain state. The value of x is then equally distributed among the other states. Suppose we have three states and we give x the value 0.10, then 90 percent will be added to one state initially and 5 percent to the other states for both. About the same applies to the state transition distribution. The transition to the same state is given the value $1 - x$ and x is distributed among the transitions to other states. This is a useful way to give the model an initial state distribution and state transition distribution because it only needs to return the value x and not the entire matrix. This way we do not have to manually adjust the whole matrices (containing the initial state distribution and state transition distribution), but only the value of x if we want to experiment with the initialization of the initial state distribution and state transition distribution. This saves a lot of time, especially as the number of hidden states increases.

The model has been tested with 0.20, 0.10, and 0.05 as values for x, for the countries Japan, United Kingdom, and the United States. For the above values of x, the same result was obtained for all countries. The value of x is therefore not of great importance (in this research), since the model converges to the same value. However, this does not mean that the value of x does not matter. In the best case, the model converges to the true initial state distribution and state transition distribution. We also tested this for the US and confirmed that the initial state distribution and state transition distribution are equal to the true initial state distribution and state transition distribution.

Data can be used in different ways for the features of the Market index and interest rate. This way you can use the levels of both features, the differences, the returns, and combinations between the variants. For example, market index returns and the difference between interest rates. For the countries Japan, the United Kingdom, and the United States, it was examined which of these different ways led to the best results. This was the combination of market index return and interest rate difference.

In conclusion, the initial state distribution and the state transition distribution will be initialized with a value for x of 0.1 For the initial state distribution this means a 90% chance you will go to a certain state and the other 10% will be divided equally between the remaining states. For the state transition distribution, this means that with a 90% chance you will remain in the same state, and with a 10% chance, you will leave the current state. The emission probabilities (the probabilities related to the observable processes) are constructed with one Gaussian mixture distribution. Lastly, the features market index return and interest rate difference will be used.

5.1.2 Results of the two-state base model

In the section, we discuss the results of the basic model. We use the performance metrics discussed in the Section 4.3.

Table 4: Results of the base model compared to classifying all data points as no recession, classifying all data points with negative growth rate as recession, and classifying the next state as the current state.

Method/model	Country	Accuracy	Precision	Recall	F-score
Base model	US	0.60	0.14	0.34	0.20
	UK	0.58	0.16	0.82	0.26
	JP	0.60	0.57	0.59	0.58
	Average	**0.59**	**0.29**	**0.58**	**0.35**
No recession	US	0.86	0.00	0.00	0.00
	UK	0.91	0.00	0.00	0.00
	JP	0.53	0.00	0.00	0.00
	Average	**0.77**	**0.00**	**0.00**	**0.00**
Growth rate	US	0.89	0.64	0.48	0.55
	UK	0.90	0.47	0.68	0.56
	JP	0.55	0.38	0.49	0.42
	Average	**0.78**	**0.50**	**0.55**	**0.51**
Martingale	US	0.98	0.91	0.91	0.91
	UK	0.99	0.92	0.92	0.92
	JP	0.95	0.95	0.95	0.95
	Average	**0.97**	**0.93**	**0.93**	**0.93**

As can be seen in Table 4, the base model only performs better for Japan compared to the GDP classification method. The historical recessions classification method only has a high accuracy which makes sense, since if all data points are classified as 'no recession' there will be no true positives, resulting in 0 for precision, recall, and the F-score.

Results for the United States
In this section, we take a closer look at the results from the United States. Figure 29 shows when the model predicted a recession and when the actual historical recessions took place in Figure 30 we see the value of the features over time compared to the average of the feature given the hidden state it is in. It can be seen in Figure 29 that the model predicts most recession periods, but that the recession periods of the model last a lot longer than the historical recession periods.

Figure 30 depicts that the model mainly bases its prediction on the unemployment rate. This can be seen from the difference in the green lines in the individual graphs in Figure 30. So for example in the recession state (state 1) of the model, the average unemployment rate is about 8% for the United States while the average unemployment rate during no recession is around 5%. The large difference between these averages says that the impact of the economic feature on the prediction of the model is significant. This could have negative effects on the model's performance, as the unemployment rate usually only rises once the impact of the recession becomes apparent. In addition, the unemployment rate usually lasts longer than the recession itself, which can lead the model to misclassify these periods. This of course leads to a lower accuracy (0.60 for the HMM base model and 0.89 for the GDP classification method) since more data points are misclassified. Furthermore, this also results in a low precision score (0.14 for the base model and 0.64 for the GDP classification method) since more data points are classified as recessions due to the unemployment rate, while these data points are not recessions, so there are many false positives.

Furthermore, it seems that the many fluctuations of the observable variables make it more difficult for the model to use them to classify the states correctly. This is also clear from Figure 30 since the difference averages of the hidden states in the unemployment rate differ a lot more than with the other features. Also, the pattern of the hidden states in Figure 29 almost exactly matches the pattern of the green line of the unemployment rate in Figure 30, indicating that the unemployment rate is seen by the model as a very important feature for determining the hidden states. However, this does not mean that the impact of the unemployment rate is positive. As just mentioned, the pattern of the hidden states created by the model is almost the same as the green line or the unemployment rate, but the actual and predicted lines in Figure 29 are certainly not equal. This implies that the model might work better without the unemployment rate.

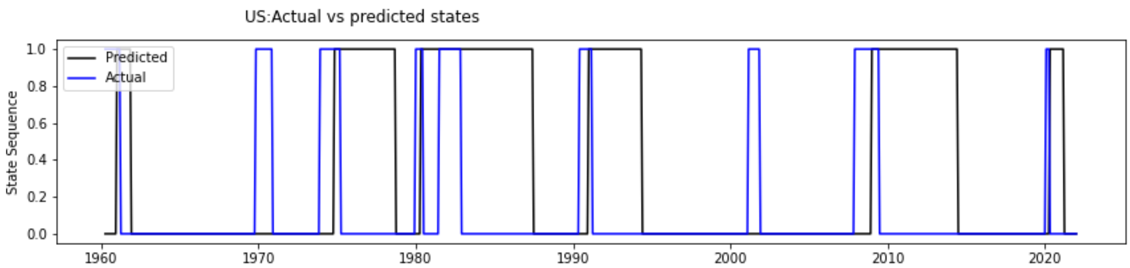

Figure 21: US: Prediction of the model versus historical recession data. The blue line in the graph shows when the historical recessions took place, if the blue line is 1, there was a recession and if it is 0 there was no recession. The black line shows when the model predicts a recession. The recession state is 1 in this case, if the black line is 0 there was no recession.

.

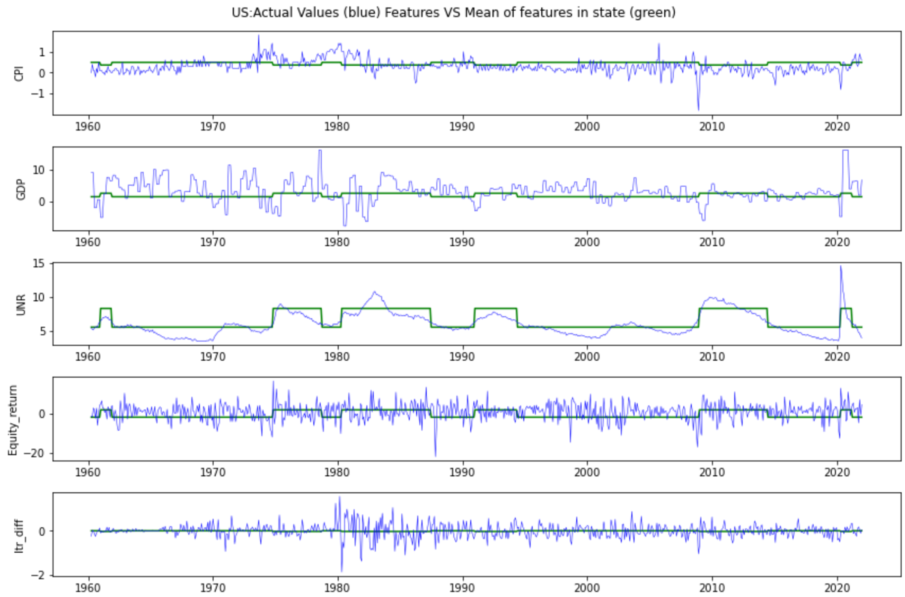

Figure 22: US: Mean of the features in state i (=0, 1) versus the actual values of the features. The blue line here is the actual value of the feature and the green line is the average value of the feature given the hidden state it is in.

.

Results for the United Kingdom

In this section, we take a closer look at the results for the UK. Figure 23 shows when the model predicted a recession and when the actual historical recessions took place in Figure 24 we see the value of the features over time compared to the average of the feature given the hidden state it is in. We can draw the same conclusions for the UK as for the US. The recession periods are identified but are a lot longer with the prediction of the model than the historical recession periods as can be seen in Figure 23. This results in many false positives and therefore in a low precision score (0.16). The recall score is a lot higher (0.82), of course, this is because the model classifies many data points as recessions and therefore also has most recession points correct, resulting in few false negatives. The accuracy (0.58) would be much higher if the recession periods predicted by the model were much shorter.

In Figure 24 it can be seen that the model mainly makes its predictions based on the features GDP and unemployment rate, since the means of the features are the furthest apart (the green line in the graphs). As discussed earlier, however, the question is whether the unemployment rate contributes positively to the model's performance. It seems that accuracy and precision (and indirectly the F-score) would be higher if the unemployment rate is not included, or if the other features are seen as more important by the model. We also see again that the features with more fluctuations are considered less important for the prediction by the model.

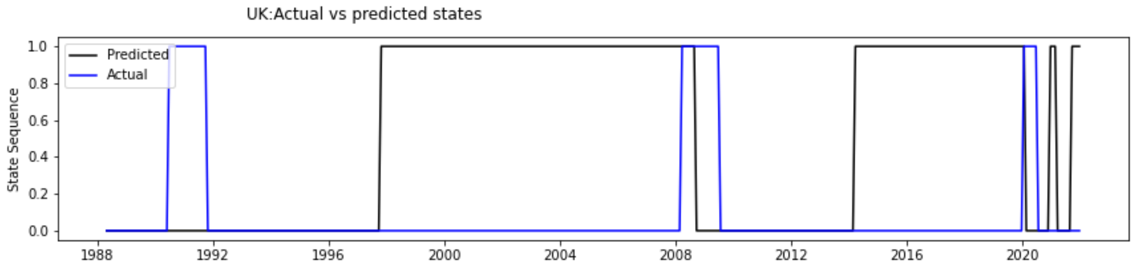

Figure 23: UK: Prediction of the model versus historical recession data. The blue line in the graph shows when the historical recessions took place, if the blue line is 1, there was a recession and if it is 0 there was no recession. The black line shows when the model predicts a recession. The recession state is 0 in this case, if the black line is 1 there was no recession.

.

Figure 24: UK: Mean of the features in state i (=0, 1) versus the actual values of the features. The blue line here is the actual value of the feature and the green line is the average value of the feature given the hidden state it is in.

.

Results for Japan

In this section, we evaluate the results for Japan. Figure 25 depicts when the model predicted a recession and when the actual historical recessions took place and the probability of the model being in a particular hidden state throughout time according to the model. In this case, the full graph has been used so that it is more visible when the model is in state 0 and state 1, this is clarified by the second and third graphs in Figure 25. In Figure 26 we see the value of the features over time relative to the average of the feature given the hidden state it is in.

Figure 25 depicts that the results are different from the results of the US and the UK. This has several reasons. First, Japan has had quite a turbulent recession pattern over the past decades as can be seen in the blue lines in Figure 26. Also, the percentage of time that Japan has been in a recession in our data is 47 percent of the time, this can also be seen from the accuracy of classifying all data points as no recession in the Table 4.

In addition, there are differences in the magnitude of the recessions that have occurred in recent decades. For example, you see large differences in the declines in GDP during the various recessions that have taken place in Japan in recent decades. Furthermore, we notice in Figure 26 that for Japan CPI and unemployment rate are seen as the most important features by the model. Again the question remains whether it benefits the performance of HMM to base the prediction almost exclusively on these two features, it is, therefore, a good idea to experiment with methods to decrease the number of fluctuations for the other features so that the model uses these features more for making predictions. It is also striking that the performance metrics of Japan are more balanced, with almost equal values for accuracy (0.60), precision (0.57), recall (0.59), and the F-score (0.58). it is also noticeable that there are only a few transitions from recession to no-recession periods in the state sequence generated by the model, some more nuance would benefit the model. This could possibly be achieved by smoothing the features with a large number of fluctuations by using moving average methods.

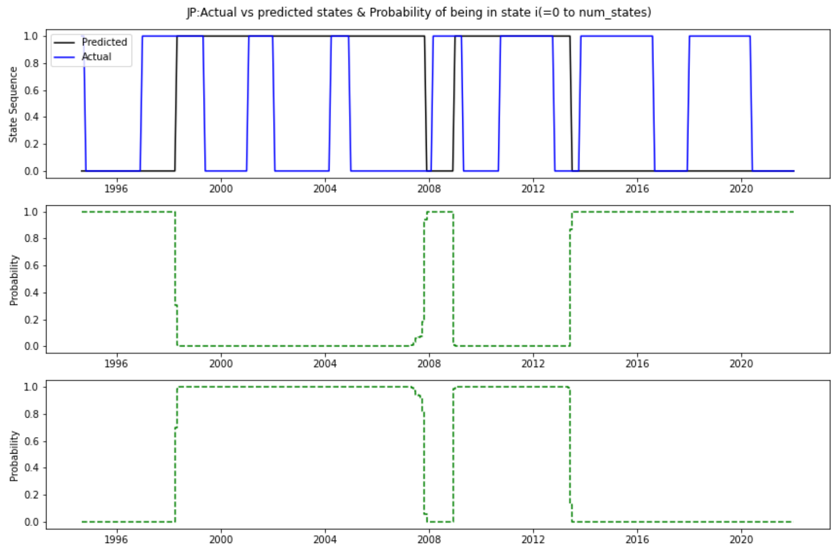

Figure 25: Prediction of the JP versus against historical recession data and probability of being in state i (=0, 1) according to the model. The blue line in the top graph shows when the historical recessions were. The black line shows when the model predicts a recession. At best, these two graphs would overlap exactly, or be exactly mirrored. This is because the model sometimes gives the recession state a 0 and the other time a 1. In this case, the recession state is 0. The other graphs in the figures indicate the probability according to the model that a country is in a particular hidden state at that moment.

.

Figure 26: JP: Mean of the features in state i (=0, 1) versus the actual values of the features. The blue line here is the actual value of the feature and the green line is the average value of the feature given the hidden state it is in.

.

Correlation between the state sequences

The hidden Markov model can generate state sequences (the economic states) as output based on the observation sequences (the features). The state sequences of the countries can then be used to create correlation matrices. In this way, it is possible to calculate the correlations between the state sequences of countries at the same point in time. These correlations can provide additional inside in the macro-economic dependencies between countries and continents. Furthermore, with the use of machine learning, the missing values for the features could be supplemented with the values of countries that have a high correlation in the economic state sequence with the country concerned.

Figure 27 shows a correlation matrix created by calculating the correlation between the different state sequences created by the model. The expectation was that there would be quite a large correlation between the countries within a continent. However, this is not the case. There is, a reasonable correlation between the United States, Canada, and Mexico, but in Europe and Asia, this appears not to be the case. Another noticeable result is Germany, which is negatively correlated with most countries. This is because the model only has two parts, the first part (1993 till 2008) which is the recession period, and the second part (2009 to 2021) is the no recession period. Finally, it is striking that several countries have a negative correlation of -1.

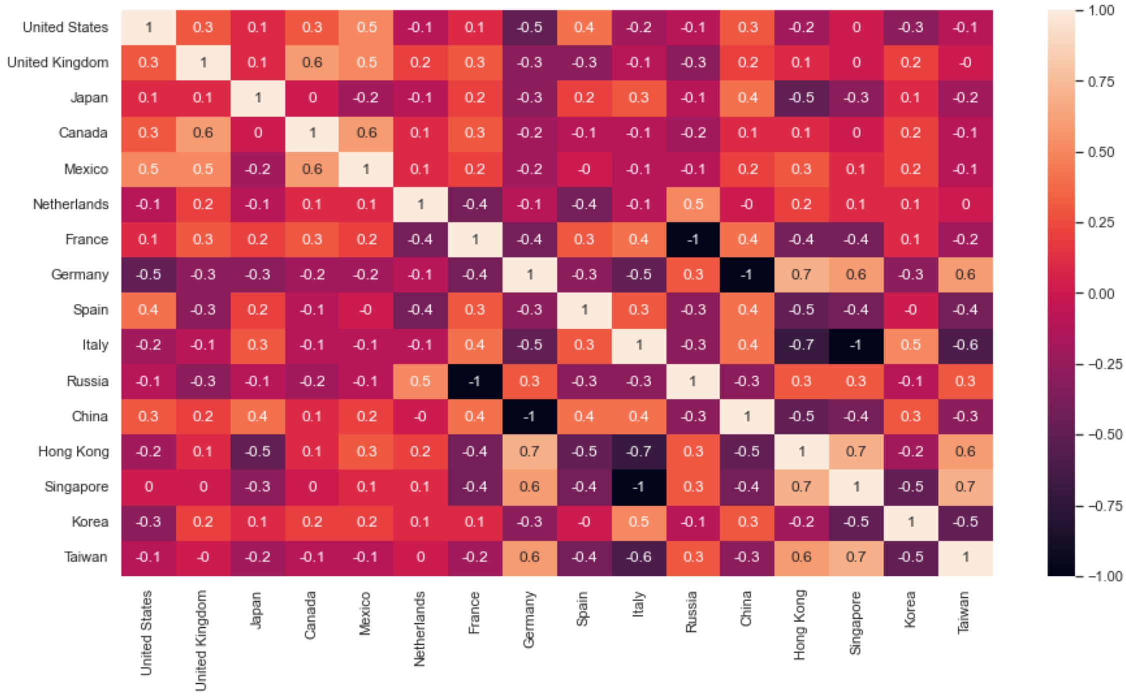

Figure 27: Correlation plot countries state sequences created by the 2 state base model.

.

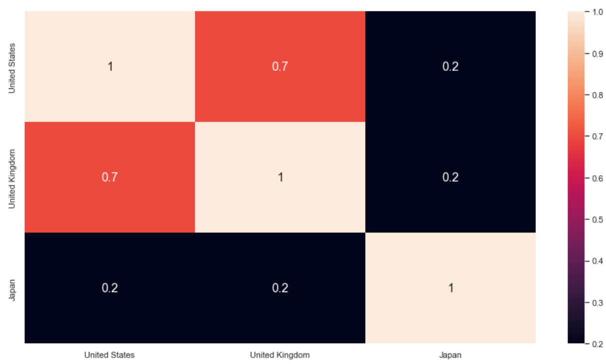

Figure 28: Correlation plot of the historical recession periods for the countries the United States, United Kingdom, and Japan.

.

The correlation of the historical recession periods between the countries the United States, United Kingdom, and Japan is in all cases (much) higher than the correlation between the state sequences of these countries generated by the model, as can be seen by comparing Figures 27 and 28.

5.1.3 Conclusions of the two-state base model

In conclusion, the model bases the state sequences logically on the values of the features (observable processes). However, there is still room for improvement since the average accuracy for the base model is 0.59, and the average accuracy of the GDP classification method is 0.78. Furthermore, the average F-score of the base model is 0.35 and the average F-score of the GDP classification method is 0.51. From analyzing the results of the base model, the following points stood out that could lead to possible improvements:

- First, the unemployment rate is seen as the most important predictor by the HMM. This is the case because the unemployment rate has the most pattern-like path. However, it is questionable whether the effect of the unemployment rate has the desired effect because it is a lagging variable. Concretely, this means that an increase in the unemployment rate is the result of a recession. Because of the unemployment rate, data points are classified as recessions while they are not according to the historical data, resulting in a lower score for the accuracy, precision and the F-score. It could therefore be that the unemployment rate negatively affects the performance of the model.
- Second, GDP does not seem to be an important predictor for the HMM. A possible reason for this is that GDP has gigantic outliers in almost all countries in 2020 during the start of the COVID-crisis. As a result, the difference in GDP in the rest of the years appears to be quite small. It may therefore be a good idea to set a maximum on these outliers of the mean plus several times the standard deviation. We evaluate the impact of the cutoff-points on the performance of the model.
- Thirdly, it is also clear to see that the model has more difficulty with pattern recognition with the features that have a lot of fluctuations. Examples of features with a lot of fluctuations are inflation, market index, and interest rate. These features would probably have more added value if we would work with a moving average of several months.
- Fourthly, the results of the correlations between the state sequences predicted by the model are unexpected in some cases, with high negative correlations between the state sequences of some countries (down to -1 in some cases). Furthermore, the correlations between the model's state sequences were compared with the correlations of the historical recession periods of the countries the United States, the United Kingdom, and Japan. This showed that the correlations of the historical recession periods were much higher than the correlations between countries for the state sequences predicted by the model.
- Fifthly, one Gaussian mixture distribution results in the best performance. The value of x does not seem to be of great importance, since the model converges to the true initial state distribution and state transition probabilities regardless of the value of x. As mentioned before we initialize the initial state distribution by giving $1 - x$ to a certain state. The value of x is then equally distributed among the other states. About the same applies to the state transition distribution. The transition to the same state is given the value $1 - x$ and x is distributed among the transitions to other states.

In the next section, we experiment with these potential areas for improvement.

5.2 Optimizing the two-state hidden Markov model

In this section, we experiment with possible improvements for the two-state hidden Markov model. We first explain the methods which could lead to possible improvements. Next, we discuss the best results of the experiments and analyze these results. Finally, conclusions are drawn based on these analyses.

5.2.1 Cutoff points, simple moving average, and exponential moving average

As discussed at the end of the previous section, there are a few points that stood out about the results of the two-state model. Firstly, the influence on the performance is limited by the severe outliers of the GDP feature in 2020. Therefore, here we evaluate the influence of these outliers and look at mitigating these outliers until a better execution of the model appears. We evaluate the influence of the outliers on the performance of the model in the following way. We are going to use cutoff points, which are calculated in the following way:

$$\text{Cutoff points} = \mu_{GDP} \mp y * \sigma_{GDP}. \tag{31}$$

In this formula, y is varied to see which leads to the best performance for the countries United States, United Kingdom, and Japan. In addition, we noticed that in many cases the unemployment rate was dominant in determining the state sequences. This is because the unemployment rate in most cases had by far the smoothest chart, while the other features went up and down faster. This was especially the case with the inflation, market index, and interest rate, with GDP this was slightly less the case. Of course, it also plays a role that GDP data points were filled because only quarterly data was available, this leads to fewer fluctuations since the same data point occurs three times. In addition, with the hidden Markov model, only the values of the current state are taken into account when predicting the next state, some patterns in the features are then lost. Using moving averages also solves this problem. The simple moving average (SMA) is the equally weighted mean of the previous M data points (Moreno, 2020). For a sequence of values, we calculate the simple moving average at period t as follows (Moreno, 2020):

$$\text{SMA}_t = \frac{O_t + O_{t-1} + O_{t-2} + ... + O_{t-M+1}}{M}. \tag{32}$$

We vary the equally weighted mean of the previous data points M for the features in the countries United States, United Kingdom, and Japan to see what the impact is and whether it improves the performance of the model. Next to experimenting with the SMA, we also experiment with the exponential moving average (EMA). With this method the weight of each element decreases progressively over time, meaning the exponential moving average gives greater weight to recent data points (Moreno, 2020). Compared to the simple moving average, the exponential moving average reacts faster to changes, since it is more sensitive to recent movements. For a sequence of values, we calculate the exponential moving average at the period t as follows (Moreno, 2020):

$$\text{EMA}_t \begin{cases} O_0, & \text{if} \quad t = 0, \\ sO_t + (1-s)EMA_{t-1} & \text{if} \quad t > 0. \end{cases} \tag{33}$$

Here:

- O_t is the observation at time period t.
- EMA_t is the exponential moving average at time t.
- s is the smoothing factor. The smoothing factor has a value between 0 and 1 and represents the weighting applied to the most recent period.

We experiment with different values for the smoothing factor s to see what the impact is on the performance of the model for the countries the United States, the United Kingdom, and Japan.

5.2.2 Results of the improved two-state model

In this section, we discuss the results of the improved 2-state hidden Markov model. We have experimented with the methods described in Section 5.2.1 as well as the number of features, Gaussian mixture distributions, and the value of x (Section 5.1.1). The results of all experiments and the parameters used to generate the results are given in Section 7.4. We used the performance metrics discussed in Section 4.3 to compare the performance of the experiments. These experiments showed that the unemployment rate for all three countries had a hurt performance. It also turned out that a Gaussian mixture distribution led to the best result. In addition, it appears that a general approach (the same parameters for all three countries) does not give the best results. The parameters and the inclusion of the amount of past data must therefore be country-specific to achieve the best performance for each country. Note that all results are given in Table 5 are without the unemployment rate. The experiments could confirm that the hypothesis of the negative influence of the feature unemployment rate was correct. Leaving out the unemployment rate feature significantly improved the performance of the model.

We compare the results of the improved model by classifying all results as recessions, classifying all data points with negative growth rates as recession, and the results of the base model. Table 5 depicts the results of the methods.

Table 5: Results of the improved 2 state HMM compared to classifying all data points as no recession, classifying all data points with negative growth rate as recession, classifying the next state as the current state, and the 2 state base model. The run defines the experiment run, this makes it easier to find the corresponding parameters of the results in Section 7.4.

Method/model	Run	Country	Accuracy	Precision	Recall	F-score
Base model	N/A	US	0.60	0.14	0.34	0.20
	N/A	UK	0.58	0.16	0.82	0.26
	N/A	JP	0.60	0.57	0.59	0.58
	N/A	**Average**	**0.59**	**0.29**	**0.58**	**0.35**
Two-state	9	US	0.80	0.41	0.92	0.57
	21	UK	0.79	0.30	0.95	0.46
	43	JP	0.80	0.90	0.64	0.75
	N/A	**Average**	**0.80**	**0.54**	**0.84**	**0.59**
No recession	N/A	US	0.86	0.00	0.00	0.00
	N/A	UK	0.91	0.00	0.00	0.00
	N/A	JP	0.53	0.00	0.00	0.00
	N/A	**Average**	**0.77**	**0.00**	**0.00**	**0.00**
Growth rate	N/A	US	0.89	0.64	0.48	0.55
	N/A	UK	0.90	0.47	0.68	0.56
	N/A	JP	0.55	0.38	0.49	0.42
	N/A	**Average**	**0.78**	**0.50**	**0.55**	**0.51**
Martingale	N/A	US	0.98	0.91	0.91	0.91
	N/A	UK	0.99	0.92	0.92	0.92
	N/A	JP	0.95	0.95	0.95	0.95
	N/A	**Average**	**0.97**	**0.93**	**0.93**	**0.93**

Table 5 shows that the model with the improvements discussed in Section 5.2 has a much better performance than the two-state base model. The average accuracy and the F-score of the two-state improved model are higher (accuracy is 0.80 and the F-score is 0.59) than that of the two-state base model (accuracy is 0.59 and the F-score 0.35). The two-state improved model has a much better performance than the GDP-based method, for Japan. The accuracy of Japan is 0.80 for the two-state improved model while it is 0.55 for the GDP classification method. Next to this, the F-score for the two-state improved model is 0.75 and that of the GDP classification method is 0.42. However, for the US and UK, the GDP classification method still outperforms the 2-state HMM, with the accuracy being about 0.10 higher for the GDP classification method. The F-score for the GDP classification method is 0.02 lower for the US and 0.10 higher for the UK.

Results for the United States

In this section, we discuss the results of the improved 2-state hidden Markov model for the United States. In Figure 29 a timeline can be seen with the prediction of the model compared to the historical recession periods, an extended version of this graph with also the probability of being at a hidden state given the time can be seen in Figure 55. In Figure 30 we see the value of the features over time compared to the average of the feature given the hidden state it is in. Table 5 shows that the results for the US of the improved two-state model are better than the results of the two-state base model for all performance metrics. The accuracy is 0.20 higher and the F-score is 0.37 higher.

In Figure 29 it can be seen that the model correctly predicts most recessions. However, there are still two points that stand out in a negative way:

1. Even without the unemployment rate feature, the recessions of the model are of longer duration than the historical recession periods.
2. There are also 2 recessions in the late 1980s while no historical recessions have taken place. When we look at Figure 30 we see that in the first recession the market index has a sharp decline and in the second recession GDP declines, which explains why the model sees these periods as recessions.

The two aforementioned points result in a larger number of false positives and therefore a still moderate precision score (0.41) and indirectly lower F-score (0.57).

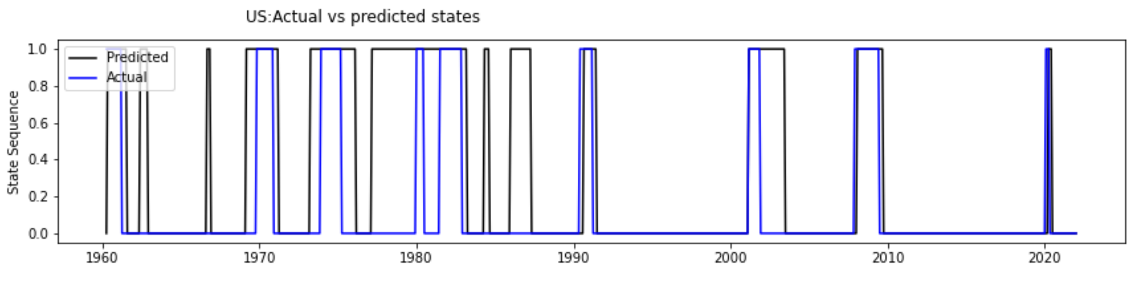

Figure 29: US: Prediction of the model versus historical recession data. The blue line in the graph shows when the historical recessions took place, if the blue line is 1, there was a recession and if it is 0 there was no recession. The black line shows when the model predicts a recession. The recession state is 1 in this case, if the black line is 0 there was no recession.

.

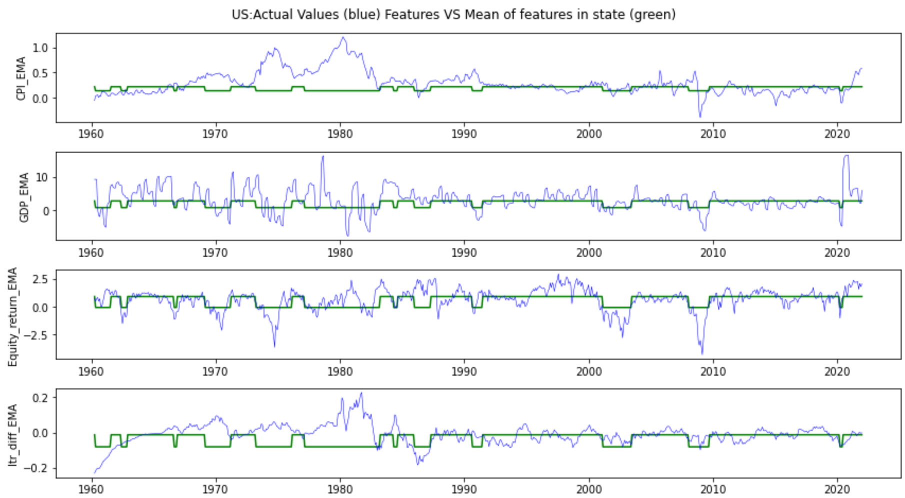

Figure 30: US: Mean of the features in state i (=0, 1) versus the actual values of the features for the model. The blue line here is the actual value of the feature and the green line is the average value of the feature given the hidden state it is in.

.

Results for the United Kingdom
In this section, we discuss the results of the three-state hidden Markov model for the United Kingdom. In Figure 31 a timeline can be seen with the prediction of the model compared to the historical recession periods, an extended version of this graph with also the probability of being at a hidden state given the time can be seen in Figure 56. In Figure 32 we see the value of the features over time compared to the average of the feature given the hidden state it is in. Table 5 shows that the results for the UK of the improved two-state model are better than the results of the two-state base model for all performance metrics. The accuracy of the two-state improved model is 0.79 while that of the base model is 0.58. Furthermore, the F-score of the two-state improved model is 0.46 while that of the base model is 0.26.

In Figure 31 it can be seen that the model identifies the historical recessions, however, two points stand out again:

1. Even without the unemployment rate feature, the recessions predicted by the model are of longer duration than the actual historical recession periods.
2. According to the model, a recession took place in 2012 when there is no historical recession as can be seen from Figure 31. The model seems to look at the stability of the features in the UK, when we look at Figure 32 we see that the features show few fluctuations in the period 1993 to 2008 and the period 2013 to 2020 while in the rest of the periods there are many fluctuations. The model, therefore, classifies the periods having a lot of fluctuations as recessions.

The two aforementioned points result in a large number of false positives and therefore a still moderate precision score (0.30) and indirectly lower F-score (0.46).

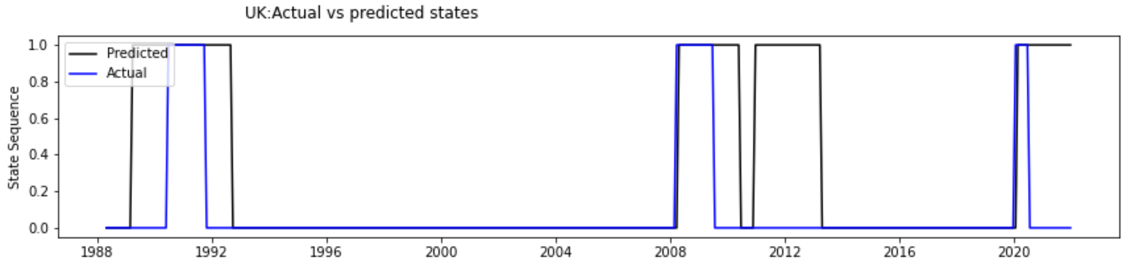

Figure 31: UK: Prediction of the model versus historical recession data. The blue line in the graph shows when the historical recessions took place, if the blue line is 1, there was a recession and if it is 0 there was no recession. The black line shows when the model predicts a recession. The recession state is 1 in this case, if the black line is 0 there was no recession.

.

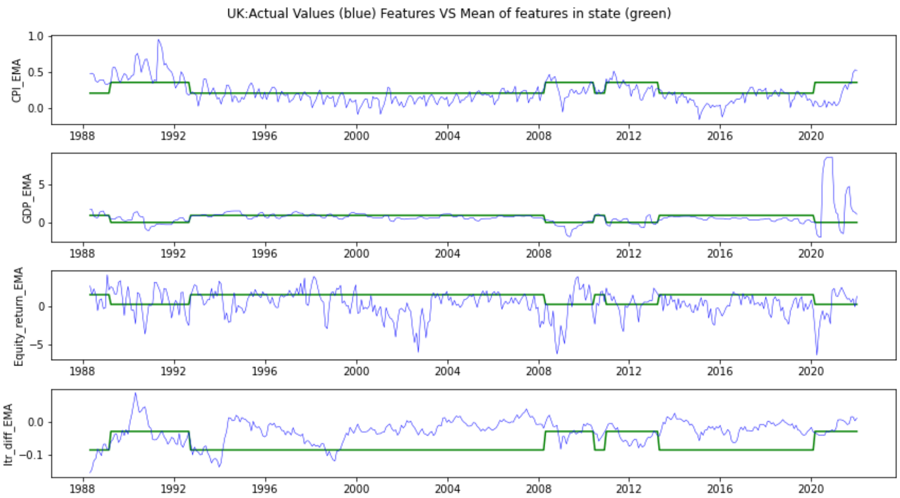

Figure 32: UK: Mean of the features in state i (= 0, 1) versus the actual values of the features for the model. The blue line here is the actual value of the feature and the green line is the average value of the feature given the hidden state it is in.

.

Results for Japan

In this section, we take a closer look at the results for Japan. Figure 33 shows when the model predicted a recession and when the actual historical recessions took place and the probability of the model being in a particular hidden state throughout time according to the model. In this case, the full Figure has been used so that it is more visible when the model is in state 0 and state 1, this is clarified by the second and third graphs in Figure 33. In Figure 34 we see the value of the features over time compared to the average of the feature given the hidden state it is in. Table 5 shows that the results for Japan of the improved two-state model are better than the results of the two-state base model for all performance metrics. The accuracy of the two-state improved model is 0.80, while the accuracy of the base model is 0.60. Furthermore, the F-score of the two-state improved model is 0.75, while the F-score of the base model is 0.58.

We see in Figure 33 that many recessions occurred in Japan since the 1990s. Most of the recessions are correctly predicted by the model, except the recessions in 2001 and 2004. When we look at the features in Figure 34, we see that the decline in GDP in the recessions in 2001 and 2004 is a lot less than the decline in GDP for the other recessions. You could describe these recessions as milder recessions and the other recessions as severe recessions. Since the data mainly contains moderate economic periods for Japan, it becomes more difficult for the hidden Markov model to identify all recessions, especially because they differ in magnitude.

It is striking that the recessions predicted by the model are not of a longer duration, as is the case in the US and the UK. This results in fewer false positives and a higher precision score (0.90). However, the recall is lower (0.64) than in the US (0.92) and UK (0.95) because the milder recessions are not classified as recessions. As a result, there are more false negatives and the recall score is lower than in the US and UK.

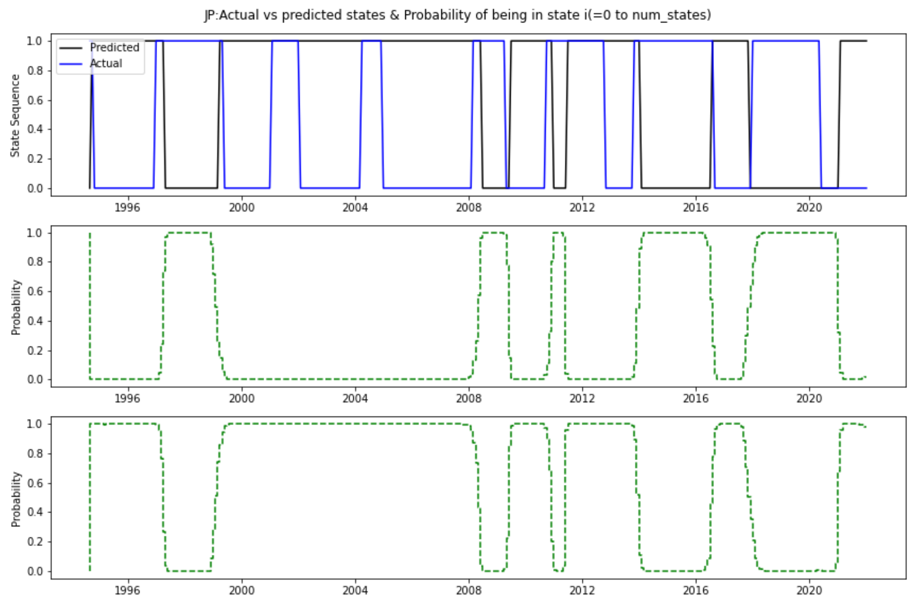

Figure 33: Prediction of the JP versus historical recession data and probability of being in state i (=0, 1) according to the model. The blue line in the top graph shows when the historical recessions were. The black line shows when the model predicts a recession. The other graphs in Figure indicate the probability according to the model that a country is in the particular hidden state at that moment.

.

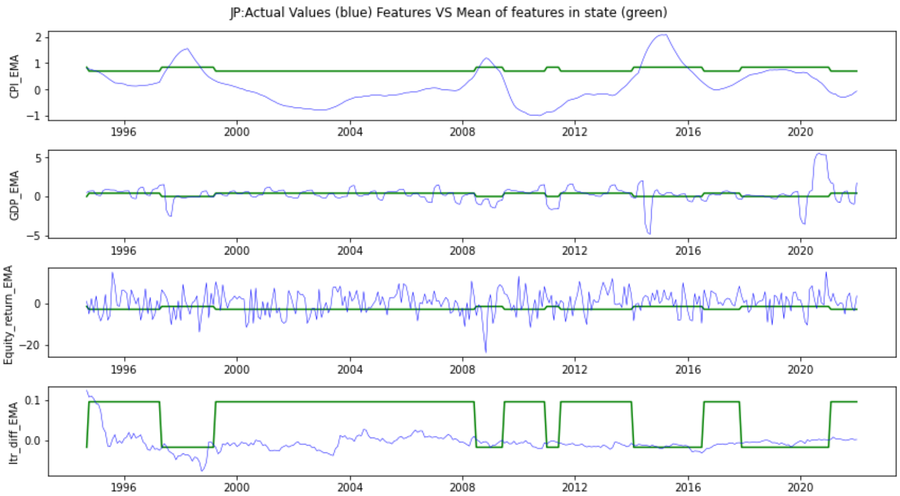

Figure 34: JP: Mean of the features in state i $(= 0, 1)$ versus the actual values of the features for the model. The blue line here is the actual value of the feature and the green line is the average value of the feature given the hidden state it is in.

.

Correlation between the state sequences of the improved two-state model
Also in the improved model, the correlations between the state sequences are plotted in a correlation matrix, which can be seen in Figure 36. These correlations can provide additional inside in the macro-economic dependencies between countries and continents. Furthermore, with the use of machine learning, the missing values for the features could be supplemented with the values of countries that have a high correlation in the economic state sequence with the country concerned. Even with the improved model, the correlations between countries within continents do not appear to be very high, with a few exceptions. The high negative correlations (-1) have disappeared.

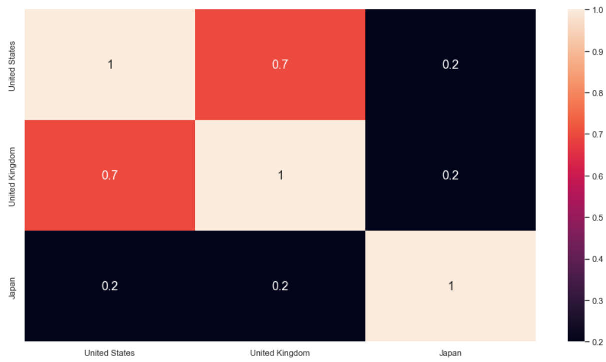

Figure 35: Correlation plot of the historical recession periods for the countries the United States, United Kingdom, and Japan.

.

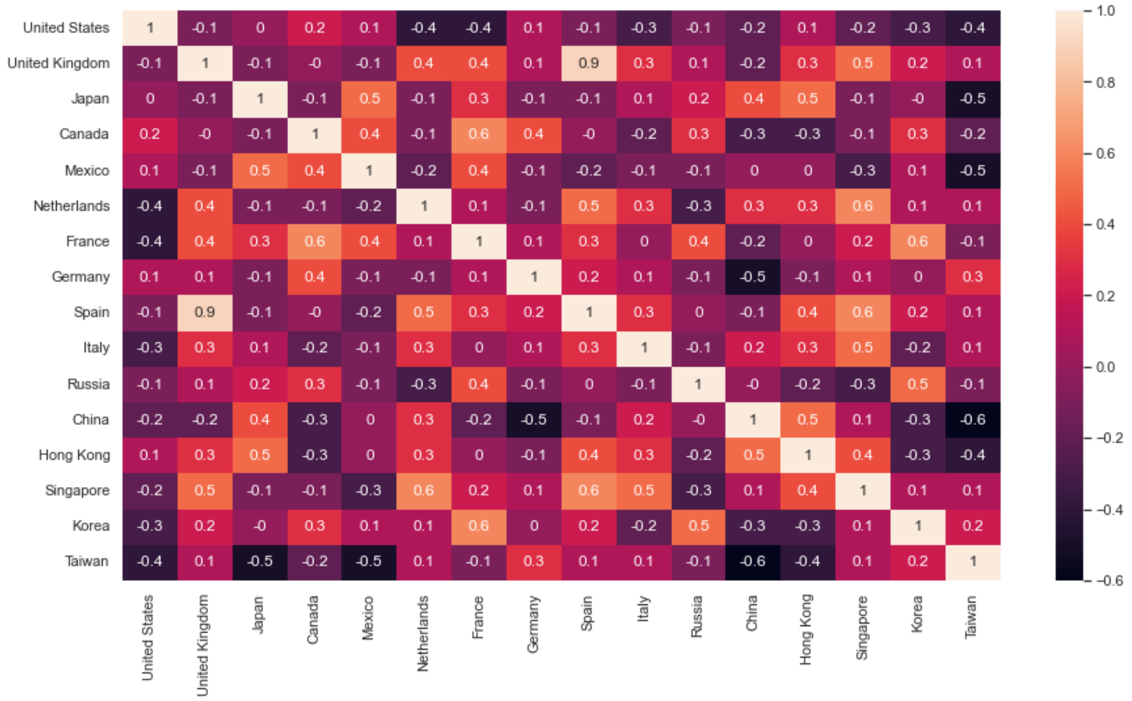

Figure 36: Correlation plot countries state sequences for the improved model.

.

The correlations between the state sequences of the countries of the base model were:

- 0.3 for the United States - the United Kingdom.
- 0.1 for the United States - Japan.
- 0.1 for the United Kingdom - Japan.

It is striking that the correlations of the base model are closer to the historical correlations than the correlations from the improved model, as can be seen by comparing Figure 36, Figure 35, and the aforementioned correlations of the base model. This is counter-intuitive because the values for the performance metrics are better than the base model. So you would expect that the correlations of the two-state improved model would also be closer to the historical correlations.

5.2.3 Conclusions of the improved two-state model

From these experiments with the improved 2-state hidden Markov model we were able to draw several conclusions:

- Table 5 shows that the results are greatly improved compared to the base model. The average accuracy of the two-state improved model is 0.80 and that of the base model is 0.59. Furthermore, the average F-score of the two-state improved model is 0.59 and that of the base model is 0.35. Furthermore, the two-state improved model also outperforms the GDP classification method for Japan. The accuracy (0.80) of the two-state improved model, as well as the F-score (0.75), are higher than the accuracy (0.55) and the F-score (0.42) for the GDP classification method for Japan. However, the results of the UK and US are not yet as good as those of the GDP classification method. The accuracies of the US and UK for the two-state improved model are lower than the accuracies of the GDP classification method (0.89 and 0.90 respectively). Additionally, the F-score of the UK is also lower for the two-state improved model (0.46 for the two-state improved model and 0.56 for the GDP classification method). The F-scores for the US are approximately the same for the two-state improved model and the GDP classification method, with a score of 0.57 for the two-state improved model and a score of 0.55 for the GDP classification method. We can conclude that the results of the two-state improved model are a lot better than the results of the base model. Still, the GDP classification method outperforms the HMM for two out of three countries. The two-state model remains a binary model where there is little room for nuance, it is therefore expected that the three-state model (adding 1 hidden state) will improve the performance.

- It is noticeable that a general approach comes at the expense of the average performance of the countries. It is, therefore, better to take a country-specific approach in which the incorporation of past data differs per country.

- During the experiments, it was noticed that the unemployment rate still seems to have a (too) large impact on the performance of the model for the countries, especially in the United Kingdom and Japan. The unemployment rate generally lasts much longer than the recession, which results in more false positives and therefore a lower score for accuracy, precision, and the F-score. It makes sense that the unemployment rate lasts longer than a recession since companies do not immediately go back to the numbers of employees from before a recession. Hiring new employees is a gradual process, where after a recession there will be cautious optimism first and then there will be application procedures involved. The fact remains, however, that the model predicts a longer recession period due to the feature unemployment rate and that the performance is better without this feature for all countries.

- We experimented with the simple moving average and the exponential moving average models to order to create smoother time series for the features. Both increase the performance and the results between these two methods are similar, with both methods resulting in approximately the same values for all performance metrics. However, with the exponential moving average, there is no data loss while this is the case with the simple moving average because the first data points can not be used with the simple moving average model. For this reason, the exponential moving average method is used for further experimentation.

- Furthermore, as mentioned earlier, Japan has had a rather turbulent economic history in recent years. Japan had many recessions since the 1990s, these recessions differ in magnitude. Using the two-state model, the model only predicts relatively large recessions. This is not surprising because the model with two states can add little nuance. It is therefore interesting to see what the results of the model will be for several states. Next to this, because there is little growth in the time frame available it becomes more difficult for the hidden Markov model to distinguish all the recessions.

- The high negative correlations (-1) have disappeared, indicating a better performance for the HMM since it does not seem realistic for countries to have a maximum negative correlation. It is striking that the correlations of the improved model are further away from the results of the correlations between the historical recession periods of the United States, United Kingdom, and Japan, for the base model.

5.3 The three-state hidden Markov model

In this section, we experiment with a hidden Markov model with 3 unobservable states. In the weather example mentioned in Section 3.1 we had 2 unobservable (hidden) states; 'Sunny' and 'Rainy'. We could also add a hidden state to this, for example, 'Cloudy', this would result in a three-state model. In the previous section, we experimented with the two-state model, where the hidden states consisted of recession and no-recession. Although the performance was adequate for the two-state improved model, it remains a binary model that has to split the economy into 2 parts, which results in little nuance. It is therefore possible that the three-state model with an additional hidden state will lead to a better performance, which is what we investigate in this section.

We go through the same process as with the two-state model. We experiment with the features used and the methods discussed in Section 5.2.1. The quantitative results of all experiments can be seen in Section 7.5.1, and the parameters used to obtain these results can be seen in Section 7.5.2.

We start by presenting the best results of the 3-state hidden Markov model and comparing these with the aforementioned methods (GDP method and classifying everything as no recession) and the results of the two-state hidden Markov model and the base model. Next, we take a closer look at the results of the United States, United Kingdom, and Japan using visual results. Finally, we draw conclusions from these results.

5.3.1 Results of the three-state model

In this section, we experiment with the higher order of the hidden Markov model (the incorporation of past data), such as different values for the smoothing factor s of the exponential moving average (EMA) for the different economic features. In addition, experiments were carried out with the number of features, as well as the number of Gaussian mixture distributions and the value of x. The results of all experiments done with the 3 state hidden Markov model can be found in Section 7.5.1 and the parameters used for this can be found in Section 7.5.2.

The experiments again showed that the unemployment rate had a negative impact on the model's performance. In addition, one Gaussian mixture distribution again worked best, and a general approach does not provide the best results for all three countries. Also, the parameters and the inclusion of past data that led to the best results with the two-state HMM, do not do this for the three-state HMM. So when we adjust the number of hidden states we have to experiment again to see which parameters lead to the best performance.

Table 6: Results of the 3 state HMM compared to the other methods and models. The run defines the experiment run, this makes it easier to find the corresponding parameters of the results in Section 7.5.2.

Method/model	**Run**	**Country**	**Accuracy**	**Precision**	**Recall**	**F-score**
Base model	N/A	US	0.60	0.14	0.34	0.20
	N/A	UK	0.58	0.16	0.82	0.26
	N/A	JP	0.60	0.57	0.59	0.58
	N/A	**Average**	**0.59**	**0.29**	**0.58**	**0.35**
Two-state	9	US	0.80	0.41	0.92	0.57
	21	UK	0.79	0.30	0.95	0.46

	43	JP	0.80	0.90	0.64	0.75
	N/A	**Average**	**0.80**	**0.54**	**0.84**	**0.59**
Three-state	6	US	0.91	0.61	0.95	0.75
	2	UK	0.87	0.43	1.00	0.60
	54	JP	0.77	0.87	0.58	0.70
	N/A	**Average**	**0.85**	**0.64**	**0.85**	**0.68**
No recession	N/A	US	0.86	0.00	0.00	0.00
	N/A	UK	0.91	0.00	0.00	0.00
	N/A	JP	0.53	0.00	0.00	0.00
	N/A	**Average**	**0.77**	**0.00**	**0.00**	**0.00**
Growth rate	N/A	US	0.89	0.64	0.48	0.55
	N/A	UK	0.90	0.47	0.68	0.56
	N/A	JP	0.55	0.38	0.49	0.42
	N/A	**Average**	**0.78**	**0.50**	**0.55**	**0.51**
Martingale	N/A	US	0.98	0.91	0.91	0.91
	N/A	UK	0.99	0.92	0.92	0.92
	N/A	JP	0.95	0.95	0.95	0.95
	N/A	**Average**	**0.97**	**0.93**	**0.93**	**0.93**

In Table 6 it can be seen that the 3 state hidden Markov model also outperforms the GDP classification method. The average accuracy of the three-state HMM is higher (0.85 for the three-state HMM and 0.78 for the GDP classification method) as well as the average F-score (0.68 for the three-state HMM and 0.51 for the GDP classification method). It is also apparent that the three-state HMM outperforms the two-state HMM. The average accuracy of the three-state HMM is 0.85 and the average accuracy of the two-state HMM is 0.80. Next to this, the F-score of the three-state HMM is 0.68 and that of the two-state HMM is 0.59. Furthermore, it is striking to see that Japan has a higher accuracy for the two-state HMM (0.80) than for the three-state HMM (0.77).

The state transition probabilities

Our goal is to calculate the weights of the scenarios of the macro-economic impact on credit risk calculations on a quantitative basis. We have confirmed that the three-state hidden Markov model is capable of identifying the economic states with an average accuracy of 0.85 and an average F-score of 0.68. The next step is therefore the calculation of the weights of the economic scenarios, which is done in the form of transition probabilities of the three-state hidden Markov model. Note, that the current approach adopted by almost all financial institutions has a weight of 0.50 for the most likely scenario, and 0.25 for the other two scenarios. Figure 37 depicts the transition probabilities for transition between the economic states which are calculated for the three-state hidden Markov model. It is noticeable that the weights are not in line with the current method used, since the probabilities are never in line with the current weight distribution used by financial institutions (0.5, 0.25, 0.25). This result indicates that the current method might not be optimal and further research about testing the weights of the HMM might improve the current credit risk calculations.

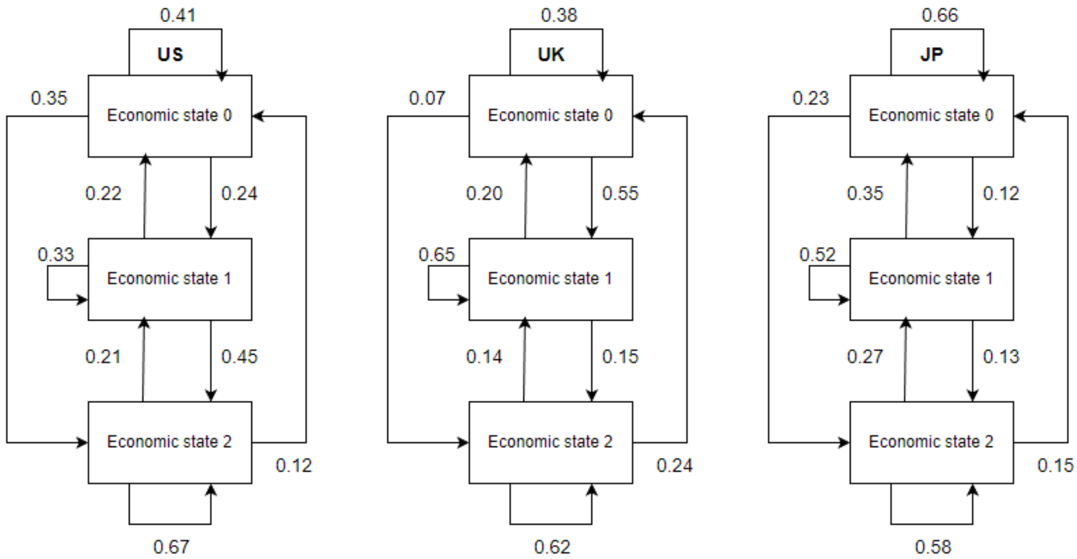

Figure 37: Transition probabilities calculated by the HMM adjusted for yearly transitions between economic states. It is noticeable that the probabilities calculated by the HMM are not in line with the weight distribution currently used (0.5, 0.25, 0.25) by most financial institutions.

.

Results for the United States
In this section, we discuss the results of the three-state hidden Markov model for the United States. In Figure 38 a timeline with the prediction of the model compared to the historical recession periods, an extended version of this graph with also the probability that the model assigns to being in that state given the time can be seen in Figure 57. Figure 39 depicts the value of the features over time compared to the average of the feature given the hidden state it is in. As can be seen in Table 6, the three-state model has better values for all performance metrics than the two-state model.

As can be seen in Figure 38, the model predicts recessions well. An exception here is the recession that took place in 1980. When we look at the second graph in Figure 39 we see the explanation for this; GDP has already started to fall earlier, so it is not surprising that the model also classifies the period before 1980 as a recession. This longer recession results in false positives, and a higher number of false positives results in a lower precision score (0.61). Apart from this long recession, there are no noticeable things that are apparent from the performance metrics as well as the visual validation methods.

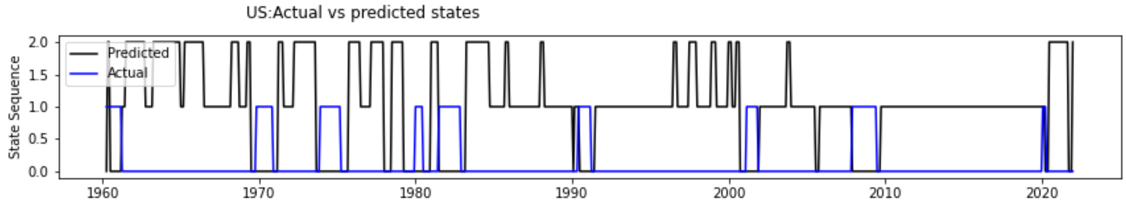

Figure 38: US: Prediction of the model versus historical recession data. The blue line in the graph shows when the historical recessions took place, if the blue line is 1, there was a recession and if it is 0 there was no recession. The black line shows when the model predicts a recession. The recession state is 0 in this case, if the black line is 1 or 2 there was no recession.

.

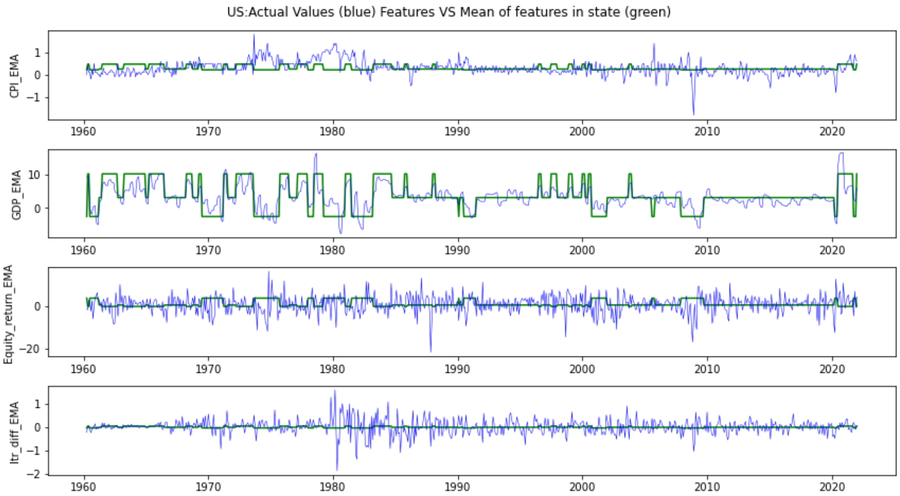

Figure 39: US: Mean of the features in state i (= 0 to 2) versus the actual values of the features for the model. The blue line here is the actual value of the feature and the green line is the average value of the feature given the hidden state it is in.

.

Results for the United Kingdom

In this section, we discuss the results of the 3 state hidden Markov model for the United Kingdom. In Figure 40 a timeline with the prediction of the model compared to the historical recession periods, an extended version of this graph with also the probability that the model assigns to being in that state given the time can be seen in Figure 58. In Figure 41 we see the value of the features over time compared to the average of the feature given the hidden state it is in.

As can be seen in Figure 40, the recession periods predicted by the model are of longer duration than the historical recession periods. This results in more false positives and therefore a lower precision score (0.43). Furthermore, all recessions are identified and all data points which are recessions according to the historical data are also classified as recessions. This results in no false negatives and therefore the maximum score (1.0) for recall.

In addition, the model's predictions show a recession in 2012, while no historical recession has occurred there. The explanation for this is a decrease in GDP, market index, and interest rate in 2012, as can be seen in Figure 41, the model therefore also sees this period as a recession period. This additional recession has a negative influence on the score of the performance metrics: accuracy, precision, and the F-score.

Figure 40: UK: Prediction of the model versus historical recession data. The blue line in the graph shows when the historical recessions took place, if the blue line is 1, there was a recession and if it is 0 there was no recession. The black line shows when the model predicts a recession. The recession state is 2 in this case, if the black line is 0 or 1 there was no recession.

.

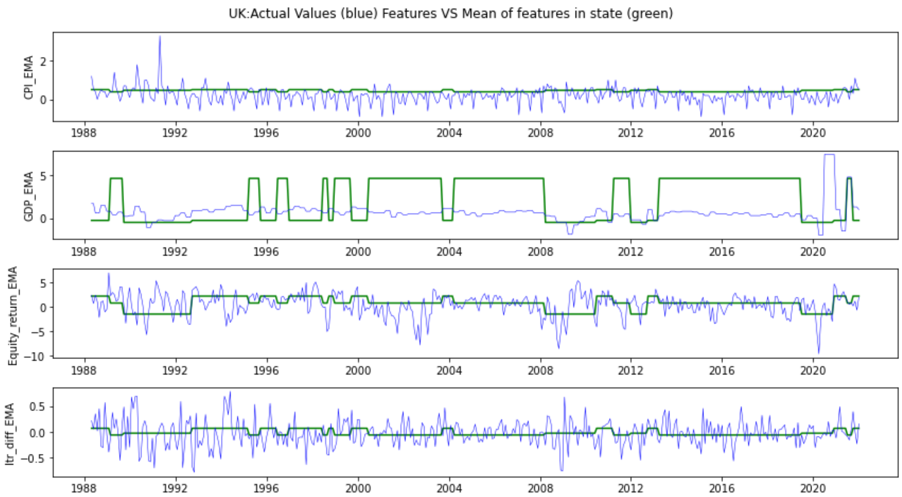

Figure 41: UK: Mean of the features in state i (=0 to 2) versus the actual values of the features for the model. The blue line here is the actual value of the feature and the green line is the average value of the feature given the hidden state it is in.

.

Results for Japan

In this section, we discuss the results of the three-state hidden Markov model for Japan. In Figure 42 a timeline with the prediction of the model compared to the historical recession periods. An extended version of this graph with also the probability that the model assigns to being in that state given the time can be seen in Figure 59. In Figure 43 we see the value of the features over time compared to the average of the feature given the hidden state it is in.

As can be seen in Figure 42, the model does not predict all recessions. Also, the exact duration of the historical recessions does not match the recession periods predicted by the model. The reason for this is that Japan has been going through a difficult economic period since the 1990s. This is also clearly reflected in the number of recessions Japan has had. There is a difference in how severe these recessions are. For example, in Figure 43 you can also see that in some recessions the inflation rises a lot more than in other recessions. Furthermore, the decline in GDP and the decline in market index differ per recession. For instance, the increase in inflation around 2015 is much greater than after 2004. Next to this, there is also a much steeper decline in GDP. Because of the absence of periods of economic growth in Japan's historical data, the model only predicts severe recessions and not minor recessions. If more historical data from Japan, including periods of high economic growth, Japan's performance would probably be better.

It is also striking that Japan again has a higher recall (0.85) than precision (0.64), in contrast to the US (recall of 0.95 and precision of 0.61) and UK (recall of 1.0 and precision of 0.43). In addition, it is notable that Japan has a higher accuracy with the improved two-state model (0.80) compared to the three-state model (0.77). However, the question is whether the supervised learning performance metrics we use for this study are also applicable to Japan. As mentioned before in Section 4.3, we transform the three-state output into a binary state output for the three-state hidden Markov model so that the performance metrics can still be used. However, in this case, you might also see State 1 as a mild recession state, therefore the performance metrics give a distorted picture.

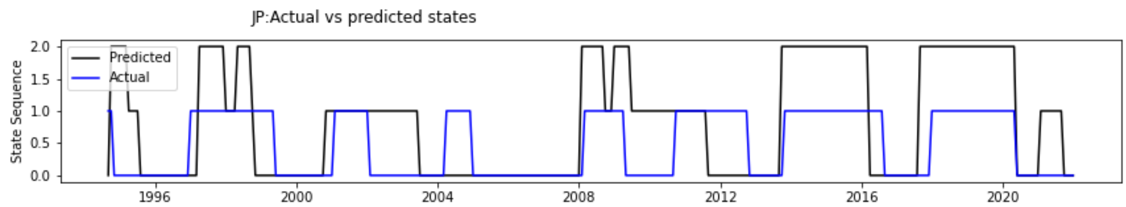

Figure 42: JP: Prediction of the model versus historical recession data. The blue line in the graph shows when the historical recessions took place, if the blue line is 1, there was a recession and if it is 0 there was no recession. The black line shows when the model predicts a recession. The recession state is 2 in this case, if the black line is 0 or 1 there was no recession.

.

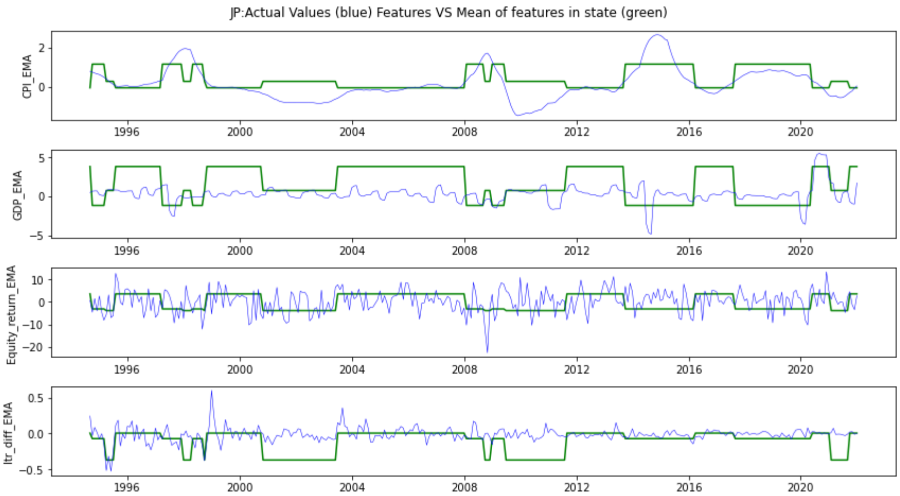

Figure 43: JP: Mean of the features in state i (= 0 to 2) versus the actual values of the features for the model. The blue line here is the actual value of the feature and the green line is the average value of the feature given the hidden state it is in.

.

Correlation between the state sequences of the three-state model
In this section we discuss the correlations between the state sequences of the three-state HMM. The correlations can provide additional inside in the macro-economic dependencies between countries and continents. Furthermore, with the use of machine learning, the missing values for the features could be supplemented with the values of countries that have a high correlation in the economic state sequence with the country concerned. Furthermore, the correlations of the HMM can be compared to the historical correlations, which results in an additional method of validation.

Not all countries can be divided into 3 states given the data length and shape of the data. With the countries that are divided into three states, the correlations are shown in Figure 44. When we compare the Figures 44 and 45 for the countries United States, United Kingdom, and Japan, we see that the correlations now match quite well. However, the correlations of the model are still somewhat lower than the correlations of the historical recession periods.

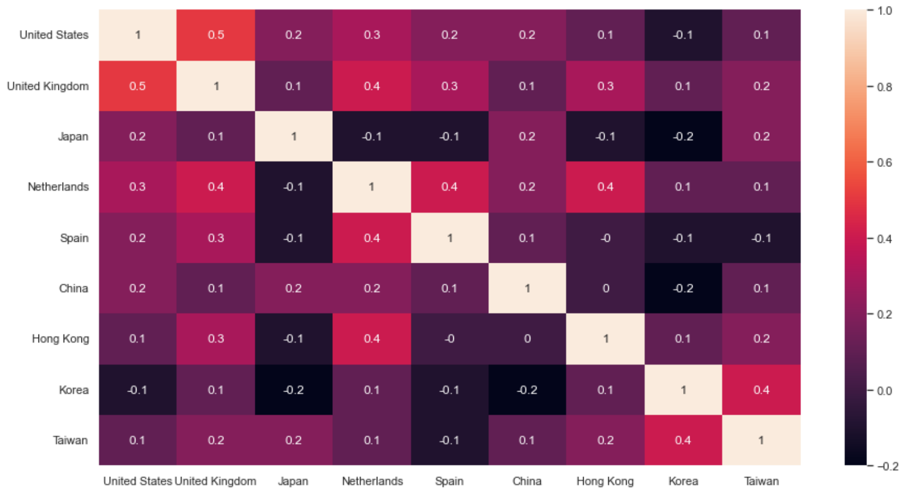

Figure 44: Correlation plot of the countries state sequences of the three-state hidden Markov model.

.

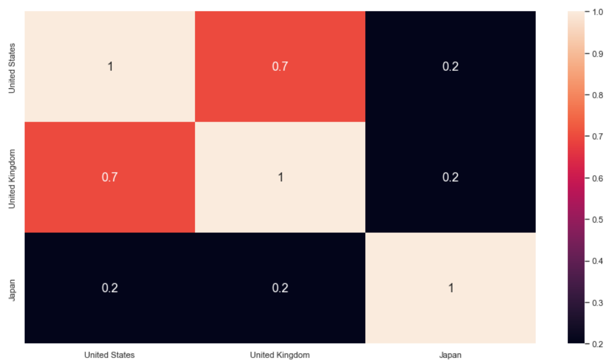

Figure 45: Correlation plot of the historical recession periods for the countries the United States, United Kingdom, and Japan.

.

5.3.2 Conclusions of the three-state model

From the experiments with the three-state HMM the following conclusions are drawn:

- The 3 state hidden Markov model outperforms all previously discussed methods and models. The accuracy, as well as the F-score, are higher. The average accuracy of the three-state HMM is 0.85 compared to 0.80 for the two-state improved HMM and 0.78 for the GDP classification method. Next to this, the average F-score for the three state HMM is 0.68 compared to 0.59 for the improved-two state HMM and 0.51 for the GDP classification method. There is however one exception; Japan. For Japan, the two-state hidden Markov model results in higher accuracy (0.80 compared to 0.77 for the three-state HMM) and F-score (0.75 compared to 0.68 for the three-state HMM). This difference in performance could be explained by the fact that

Japan has had a large number of recessions since the 90s. The economic data consists mostly of moderate economic growth and recessions. The three-state hidden Markov model thus distinguishes more between the smaller and bigger recessions, resulting in lower performance metrics with the three-state model for Japan. It is important to note that the research is about predicting economic states and not recessions. If the HMM distinguishes between two recessions states because that is most likely this is not a problem for the scenarios and their transition probabilities between them. The validating on recession states is only to see if the model identifies the economic states well.

- As mentioned before, the smoothing factor is the weighting applied to the most recent period. The smoothing factor s of the exponential moving average that is used for making the time series of the features more smooth. The smoothing factors that led to the best results for the 2-state HMM, are not the same as those that led to the best results for the 3-state HMM. When the number of states changes, the smoothing factors must therefore be reconsidered.
- Not all countries could be divided into three states given the length and shape of the data. The more data, the easier it is for the hidden Markov model to divide the country into multiple states. The same goes for the 'shape' of the data, when a country has a fairly stable economy and therefore also fairly stable economic features, it becomes a lot more difficult for the hidden Markov model to divide the country into several economic states.
- The correlations between the countries United States, United Kingdom, and Japan are quite close to the historical correlations for the three-state hidden Markov model, which is another indication that the 3 state model performs well.

5.4 The five-state hidden Markov model

Because the three-state model led to better performance, it was interesting to see if the five-state HMM would also lead to better performance. In this section, we, therefore, discuss the results and mainly discuss why the five-state model does not work. The 5 state hidden Markov model does not work properly for the following reasons:

- The main reason is that the complexity and subjectivity of the model increase considerably. By complexity, we mean that it becomes a lot more difficult to identify the recession states and calculate the performance metrics. Subjectivity means that the recession states are less obvious than with a smaller number of states, it is up to the user to label the states, and an HMM with a higher number of states results in more recession states, which makes it more complex to label. The number of recession states is not always the same and the recession states are not always unambiguous, hence the subjectivity.
- The five-state hidden Markov model does not find a solution for the majority of the countries when the methods for the inclusion of past data and dealing with extreme outliers (GDP cutoff points, EMA, SMA) are used. Only for the United States solutions are found. As can be seen in the experiments with the two-state and three-state hidden Markov model, higher order hidden Markov models (inclusion of more past data with the use of EMA or SMA) result in a better performance of the model.
- Even without the use of the EMA or SMA, the 5-state hidden Markov model does not find a solution in many countries. The reason for this is that a large number of countries only have a limited number of data points and if we want to divide the periods into 5 states, it becomes a challenge for the HMM.

The correlation matrix shown in Figure 46 is because of the aforementioned reasons made without the EMA, SMA, and GDP cutoff points. As can be seen when we compare the correlation matrices in Figures 46 and 47 we also see that the correlations of the 5 state hidden Markov model for the countries the US, the UK and Japan are far from the correlations of the historical recession periods, this is another indication that the five-state HMM is not performing well. The reason why the HMM does not find a solution is that the HMM does generate 5 states, but some of them are never visited. This results in a division by zero, and therefore no solution.

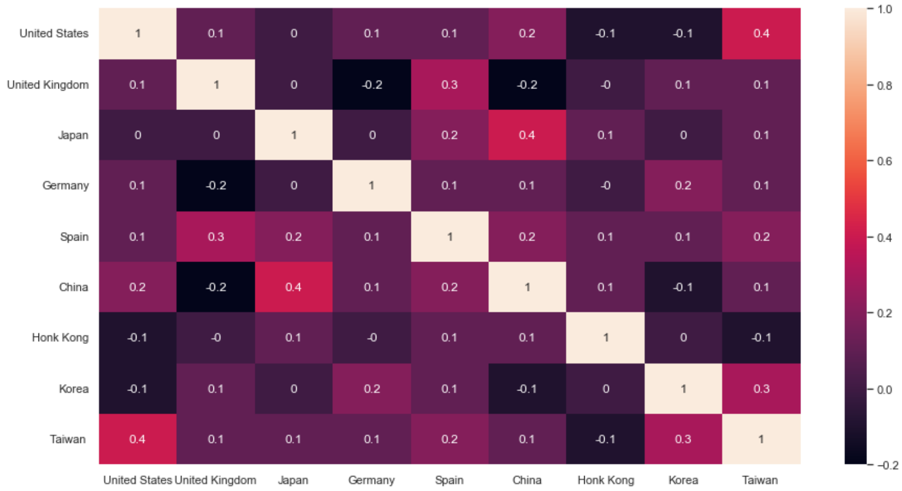

Figure 46: Correlation matrix of the state sequences produced by the 5 state hidden Markov model.

.

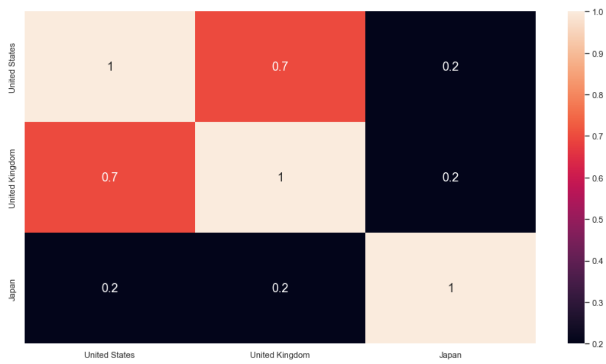

Figure 47: Correlation plot of the historical recession periods for the countries the United States, United Kingdom, and Japan.

.

5.5 AIC and BIC of the hidden Markov models

In this section, we provide the scores for the AIC and BIC for the different hidden Markov models used in this research. In Section 4.2 we discussed the AIC and the BIC in detail, and we briefly recap both methods. The AIC and BIC are based on the likelihood function with the purpose to estimate the relative quality of models. The lower the score for both methods, the better. Both the AIC and BIC add a penalty term for the complexity of the model. Note that, the AIC and BIC are independent of the recession identification process and the scores for the performance metrics that come with it. They measure separate aspects of the performance of the model. Table 7 gives the results for the AIC and the BIC for the base model, two-state model, and three-state model.

Table 7: Results of the AIC and BIC for the base model, the two-state model, and the three-state model.

Model	**Country**	**AIC**	**BIC**
Base model	US	10754	10771
	UK	4402	4414
	JP	3586	3596
	Average	6247	6260
Two-state	US	2016	2033
	UK	226	176
	JP	1350	1360
	Average	1197	1190
Three-state	US	7518	7546
	UK	2494	2514
	JP	2112	2129
	Average	4041	4063

As can be seen in Table 7 the two-state hidden Markov model has the lowest values for the AIC and BIC, indicating the best performance. The base model has the highest scores for the AIC and BIC. It makes sense that the two-state model performs better than the base model since the penalty term for the complexity is not increased for the two-state model compared to the base model. The penalty term is not increased since the number of parameters is not increased. The difference between the base model and the two-state HMM is that more past data is included when making the predictions.

There are several reasons that the AIC and BIC of the three-state model are higher than the two-state model:

1. The complexity goes up (there are a lot more parameters), and therefore also the penalty term for the complexity.
2. It is easier to divide the economic periods into two states than three states. Therefore the paths are easier to predict, which results in more likely paths and a higher log-likelihood value.

Another point that has to be taken into account is that the performance metrics of the three-state HMM are already chosen out of three binary state outputs, as discussed in Section 4.3.

5.6 Summary

We experimented with the initialization of the initial state distribution, transition probabilities, and the number of Gaussian mixture distributions that are used for the emission probabilities. Furthermore, we experiment with the number of hidden states and higher-order of the hidden Markov model. These experiments aimed to see which hidden Markov model leads to the best performance metrics and why. This is with a view to the higher goal of this research; to see if the hidden Markov models can predict the economic states. If this is the case, the transition probabilities could be calculated. These probabilities can be used as new weights for the scenario-based approach that financial institutions use to include macro-economic impact in expected credit risk losses. The analysis of the experiments with the two-state base model showed that the unemployment rate and outliers of the GDP feature might have a negative influence on the model. Furthermore, due to many fluctuations of features, pattern recognition becomes more difficult, resulting in sub-optimal performance of the model.

From the experiments conducted with the length of the incorporation of past data and use of features (or absence), we conclude that the model performed better without the unemployment rate. In addition, cutoff points for GDP's extreme outliers improved the model's performance. Finally, we conclude that the simple moving average and exponential moving average applied to the features make pattern recognition easier for the model, resulting in a better performance of the model.

Although the performance was adequate for the two-state improved model, it remains a binary model that has to split the economy into 2 parts, which results in limited nuance. We therefore also experimented with the three-state HMM to see whether increasing the number of hidden states with one would increase the performance of the model. We applied the same method of experimentation for the hidden Markov model with 3 hidden states of experimentation. Because the three-state HMM had significantly higher performance metrics, we conclude that the three-state HMM outperformed the two-state HMM. Because adding one hidden state led to improved performance, we evaluated whether the five-state model also led to better performance, but this was not the case. The complexity and subjectivity of the model increase considerably since it is a lot more subjective to label the states. Furthermore, the five-state hidden Markov model is not able to find a solution for most of the countries when the aforementioned EMA, SMA, and the cutoff points for the GDP feature are used, while with the two-state and three-state models these methods significantly increased the performance of the model.

We also calculated the AIC and the BIC, the purpose of the AIC and BIC is to estimate the relative quality of the models. We calculated the AIC and BIC for the base model, the two-state model, and the three-state model. The two-state model had the lowest (indicating the best performance for AIC and BIC) value. There are two reasons for the two-state model outperforming the three-state model. The first reason is that the complexity term goes up, which results in a higher penalty term and a higher score for the AIC and BIC. The second reason is that it is easier to divide the economic periods into two states than three states. Therefore the paths are easier to predict, which results in more likely paths and a higher log-likelihood value.

In short, we can conclude that the two-state, and especially the three-state higher-order hidden Markov model can predict economic states. The three-state model outperforms all viable benchmarks and other HMMs with an average accuracy of 0.85 and an average F-score of 0.68. Compared to an average accuracy of 0.78 and an average F-score 0.51 of for the GDP classification method and an average accuracy of 0.77 for the no recession method and an average F-score of 0.00.

Lastly, the economic state transition probabilities calculated by the HMM are not in line with the weight distribution currently used by most financial institutions (0.5, 0.25, 0.25). The state transition probabilities calculated by the HMM are not close to the weight distribution currently used for the United States, United Kingdom, and Japan.

6 Conclusion, discussion, and further research

The goal of this research is to determine a quantitative way of predicting economic states and the transition probabilities between them, with means of the features inflation, unemployment rate, GDP, market index, and interest rate.

During this research, the lack of data and the definition of the NBER played a major role in choosing the most suitable method. Since the NBER does not have an exact definition and only a few countries have historical recession periods available, supervised learning was not an option. Of the considered unsupervised learning methods, the hidden Markov model seemed to be the most suitable.

We experimented with 2, 3, and 5 hidden states and the higher order of the HMM (incorporation of past data) of the features, as well as the initialization of the initial state distribution, the initialization of the state transition distribution, and the number of Gaussian mixtures used for the observation distribution. We validated the performance of the models using the performance metrics accuracy, recall, precision, and the F-score. Additionally, we used the AIC and BIC to look at the relative performance of the different HMMs.

In addition, the correlations of the countries between the state sequences produced by the model are also examined. For the countries United States, United Kingdom, and Japan, these have been compared with the correlations of the historical recession periods of the NBER. The historical recession periods of these countries were available and were used in this study.

In this chapter, the main research question is answered, the discussion and contribution to theory and practice are given and finally, advice is given for possible further research.

6.1 Conclusion

We first briefly recap the goal of this thesis. In this thesis we analyze the prediction of economic states on behalf of EY Technology Consulting. Our goal is to identify economic states and calculate transition probabilities, for transitioning between these economic states. For the IFRS9 financial institutions have to take macro-economic impact on expected credit loss into account with a forward-looking view. The IFRS9 standard requires financial institutions (or other companies with financial assets like loans) to estimate potential credit losses with a forward-looking view. Although the standard does not prescribe any specific ways of doing so, most financial institutions are taking a scenario-based approach to include forward-looking macro-economic impact. The weighting of such scenarios is often quite basic where the most likely scenario (baseline) accounts for 50% and the remaining two (upside and downside) share the other 50% equally. The percentages assigned to these scenarios are not determined via a quantitative method. The aim of this thesis is therefore to find a method that determines these weights on a quantitative basis, in order to do this it is important that the method is able to identify the different economic scenarios. This results in the formulation of the main research question:

"In what way and to what extent can machine learning be applied using the economic features inflation, GDP, unemployment rate, market index, and interest rate for the prediction of the economic state of countries over a monthly time frame defining the performance of the model by the performance metrics accuracy, precision, recall, and the F-score?"

To answer the main research question we considered different supervised and unsupervised machine learning methods, hidden Markov models seemed to be the most suitable because in addition to being able to classify periods in economic states, it is also able to calculate transition probabilities (the weights for the different scenarios). A hidden Markov model (HMM) is a statistical Markov model in which the system is assumed to be a Markov process, with unobservable (hidden) states. The HMM requires that there is an observable process whose outcomes are influenced by the outcomes of the unobservable process in a known way. The goal is to learn the unobservable process using the observable process. In this research, the economic states (recession and no recession for the two-state model) are the unobservable processes. These unobservable processes, according to the definition of the NBER, cannot be determined with an exact value or class label. However, with our observable processes (inflation, unemployment rate, GDP, market index, and interest rate) we can learn the unobservable process. Three fundamental problems can be solved using the HMM:

1. Likelihood: the likelihood problem allows you to choose the best match along competing models. We use the likelihood problem to calculate the AIC and BIC of the base model, the two-state HMM, and the three-state HMM.

2. Decoding: the decoding problem allows you to discover the hidden part of the hidden Markov model. We use the decoding problem to find the optimal economic state sequences for the countries. For the countries United States, United Kingdom, and Japan we validate the economic state sequences by comparing them with the historical recession data available.

3. Learning: the learning problem allows you to find the optimal model parameters that maximize the probability of the observation sequences. The learning problem can be viewed as training the model to best fit the observed data. We use the learning problem to calculate the initial state distribution and economic state transition probabilities. The economic state transition probabilities are compared with the current weights used by financial institutions to take macro-economic impact into account in the expected credit loss calculations.

We started with a hidden Markov model with two hidden states (recession and no recession) since the states of the binary model are easier to decipher than a multi-state hidden Markov model. We validate the performance of the models in this research by the performance metrics: accuracy, precision, recall, and the F-score. We benchmark the performance of the HMM against three methods:

1. Classifying all results as recessions. This can lead to high accuracy (average of 0.77) with unbalanced data, however, for the other performance metrics, this is not the case (average of 0.00 for all other performance metrics). Note that the accuracy of this method is the same as the initial state distribution for the two-state hidden Markov model, so when we reverse the method (classifying everything as a recession) this results in 1 minus the accuracy of classifying everything as no recession.

2. Classifying all data points with negative growth as a recession (we call this GDP classification method in this research). A negative growth rate is a common indicator of a recession. Companies could use forecasted growth rates as an indicator of economic states. A forecasted negative growth rate could be used as a predictor for a recession state. The average accuracy for this method is 0.78 and the average F-score is 0.51.

3. The martingale method, In our case this results in the using the current state x_t as a prediction for the next state x_{t+1}. Keep in mind that our goal is to determine weights for the different scenarios that are used to include macro-economic impact in the credit risk calculations. However, the assumption with the martingale is that the current state with a probability of 1 is

the next state, which is not in line with the requirements of the IFRS9, which states that it is not allowed to have only one scenario unless there are adjustments made to compensate for the non-linearity in the expected credit risk losses. The martingale method does score very good on the performance metrics, with an average accuracy of 0.97 and an average F-score of 0.93.

We started with the two-state base model. The goal of this base model was to run with the features and analyze the results for possible improvement. The average accuracy of this base model is 0.59 and the average F-score is 0.35. These scores of the performance metrics of the base model are lower than that of the GDP classification method (accuracy of 0.78 and F-score of 0.51), which indicates that there is room for improvement. By analyzing and experimenting with the two-state model, possible areas of improvement were found:

1. The unemployment rate seems by far the most important feature for the predictions of the model. This is the case because the unemployment rate has the most pattern-like path. However, it is questionable whether the effect of the unemployment rate has the desired effect because it is a lagging variable, removing the feature unemployment rate might increase the performance of the model.

2. GDP does not seem to influence the model much. The reason for this is that GDP has extreme outliers for almost all countries in 2020 during the start and end of the pandemic. As a result, the difference in GDP in the rest of the years appears to be quite small. It may therefore be a good idea to use cutoff points for the extreme outliers.

3. It is also clear to see that the model has more difficulty with pattern recognition with the features that have a lot of fluctuations like market index and interest rate than with the features that do not have a lot of fluctuations like the unemployment rate. These features would probably have more added value if we would work with a moving average of several months.

The use of the cutoff points, moving averages (the inclusion of more past data to make the predictions), and the removal of the unemployment rate resulted in higher scores for all performance metrics for all countries. The average accuracy of the two-state improved model is 0.80 and the average F-score 0.59 compared to an average accuracy of 0.59 and an average F-score of 0.35 for the base model. Although the performance was adequate for the two-state improved model, it remains a binary model that has to split the economy into 2 parts, which results in little nuance. We therefore also experimented with the three-state HMM (adding one hidden state). The three- state hidden Markov model outperforms all previously discussed methods and models (except the martingale, but this method can not be used directly as discussed). The accuracy (0.85), as well as the F-score (0.68), are higher. There is however one exception; Japan. For Japan, the two-state hidden Markov model results in higher accuracy (0.80 compared to 0.77 for the three-state HMM) and F-score (0.75 compared to 0.70 for the three-state HMM). This could be explained by the fact that Japan has had an enormous amount of recessions since the 90s. The economic data consists mostly of moderate economic growth and recessions. The three-state hidden Markov model thus distinguishes more between the smaller and bigger recessions, resulting in lower performance metrics with the three-state model for Japan.

Since increasing the number of hidden states from two to three increased the performance of the model we also evaluated whether the five-state HMM would again lead to better performance, which was not the case. The complexity and subjectivity of the model increased significantly while performance lagged. Also, the model was usually unable to converge for many countries when cutoff points or exponential moving average are used. This while these methods greatly increased the performance of the two-state and three-state model. Even without these methods, the five-state HMM did not find a solution for many countries. We can conclude that it is not the case that the more states the hidden Markov model has, the better the performance.

Next to identifying the recessions state using various hidden Markov models, we also calculated the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). We used the hidden Markov model to calculate the log-likelihood which is used as a parameter for the calculation of the AIC and BIC. The purpose of the AIC and BIC is to estimate the relative quality of the models, therefore the absolute values are not of great importance. The lower the score for the AIC and BIC, the better the performance. Both the AIC and the BIC add a penalty score for the complexity of the models. The more parameters the more complex the model, which results in a higher penalty score which is added to the AIC and BIC. We calculated the AIC and BIC for the base model, the two-state model, and the three-state model. The two-state model has the lowest (indicating the best performance for AIC and BIC) value and the base model has the highest score. There are several reasons for the two-state model outperforming the three-state model:

1. For the three-state model, the complexity term goes up, which results in a higher penalty term and a higher score for the AIC and BIC.
2. It is easier to divide the economic periods into two states than three states. Therefore the paths are easier to predict, which results in more likely paths and a higher log-likelihood value.

As mentioned in the research question we defined the performance using the aforementioned performance metrics. In short, we can conclude that the two-state, and especially the three-state higher-order hidden Markov model can predict economic states. The three-state model outperforms all previously discussed methods with an average accuracy of 0.85 and an average F-score of 0.68. Compared to an average accuracy of 0.78 and an average F-score 0.51 of for the GDP classification method and an average accuracy of 0.77 for the no recession method and an average F-score of 0.00. Only the martingale method has a better performance with an average accuracy of 0.97 and an average F-score of 0.93. However, as previously discussed a single forward-looking scenario (i.e the most likely scenario) does not meet the requirements of IFRS9 unless there is an adoption of an adjustment to reflect non-linearity in the credit loss distribution for alternative scenarios. We also calculated the correlations between the state sequences (output of the HMM) of the countries for all models, and compared those to the correlations of the historical data, resulting in an additional method of validation. For the three-state HMM these correlations were quite close. The correlation between the US-UK is 0.5 for the three-state HMM and the historical correlation is 0.7. For the US-JP this is 0.2 for the three-state HMM and 0.2 for the historical correlations. Lastly, the correlation for the UK-JP is 0.1 for the three-state HMM compared to 0.2 for the historical correlation. The closeness of the correlation values for the economic state sequences is another indication that the three-state HMM is working properly.

Considering the economic features, we can conclude that the unemployment rate does not work well as an input for the HMM. The reason for this is that the unemployment rate only rises after the recession has been going on for a while, and lasts a lot longer than the recession. For the other economic features, the inclusion of more past data with the use of moving averages (SMA or EMA) to create a more pattern-like path, benefits the performance of the model. Furthermore, the removal of extreme outliers for the GDP feature leads to higher scores for all performance metrics.

Lastly, the economic state transition probabilities calculated by the HMM are not in line with the weight distribution currently used by most financial institutions (0.5, 0.25, 0.25). The state transition probabilities calculated by the HMM are not close to the weight distribution currently used for the United States, United Kingdom, and Japan. This deviation indicates that the current method is not optimal and that it is worth to research the impact of using the weights calculated by the HMM for the expected credit loss calculations.

6.2 Discussion

Our main limitation was the lack of data. The data for the economic features used during this research are obtained via EY from the data provider Reuters. For many countries, there is limited availability of historical economic data for the features inflation, unemployment rate, GDP, market index, and interest rate. And if there is data available, the data usually did not go far back, which in many cases results in few data points. This is one of the reasons why monthly data have been chosen instead of annual data. In addition, the NBER has historical recession data available for a few countries. If more data are available, a kind of supervised learning method could be used for more countries (not only for the US, UK, and Japan), by comparing the historical recession with the state sequences provided by the model. This lack of data does not mean that this method can not be used right now. We have confirmed that the HMM can identify the recession states, the use of the models is therefore not restricted anymore to countries that have recession data available consistent with the definition of the NBER. Furthermore, with the use of machine learning, the missing values for the features could be supplemented with the values of countries that have a high correlation in the economic state sequences with the country concerned. Concluding, we do not have to wait until more data becomes available before the model can be applied to other countries, it just requires additional data pre-processing.

Another limitation of the research is that the results of the hidden Markov model do not provide a literal classification of the economic state (like recession or no recession for the two-state HMM), but divide the periods into numbers (0 or 1 for the two-state HMM). It is up to the user of the HMM to decipher the state associated with this number. This can be done using the graphs that the model gives as output and by evaluating the performance metrics of multiple scenarios, for this task some familiarity with numbers and graphs is required. This does not mean that a lot of knowledge is required of the user, for someone with some familiarity with graphs, the interpretation is quite straightforward. A downside of the HMM is thus that it is more complex and time-consuming than data that are labeled beforehand (like supervised machine learning). On the other hand, this is also the strength of the model, since the definition according to the NBER is not exact, and for the other economic states, no definition is available. It is therefore not possible to use a model which exactly classifies the economic states according to this definition. Using the HMM it is therefore still possible to predict the economic states and provide meaningful predictions for credit risk models.

6.3 Contributions to theory and practice

Our research has both a theoretical contribution and a practical one. As far as the theory goes, there has been very little research into applying machine learning to macro-economic conditions. However, we have shown that there are applications. We have shown that economic conditions can be predicted one month ahead, using unsupervised machine learning. The following economic features have been used to achieve this: inflation, GDP, market index, and interest rate. Cutoff points for the GDP features and exponential moving average for all features were used to increase the performance of the model. Furthermore, we have shown that supervised machine learning performance metrics with some modifications can be used for hidden Markov models. The output of the HMM are not clear labels, but instead, the output of the HMM are numbers (signifying an economic scenario) depending on the number of hidden states in your HMM. Still, if historical data is available you can calculate the performance metrics of both scenarios and decipher based on visual output and the performance metrics what the right label is. Furthermore, we showed that higher-order hidden Markov models (created with the use of the simple moving average and the exponential moving average) and the removal of extreme outliers improve pattern recognition for the hidden Markov resulting in a better performance. This could also be applied in other applications for the HMM.

Our goal is to find a method that determines the economic scenarios quantitative basis with corresponding transition probabilities. We confirmed that higher-order hidden Markov models are able to determine economic states. The three-state HMM has a particular good performance since it outperforms all other models and viable benchmarks. Furthermore, we have transition probabilities calculated by the HMM are not in line with the current weight distribution (0.5, 0.25, 0.25) used by most financial institutions, as can be seen in Figure 37. The last step would be testing the weights (transition probabilities) of the HMM on the credit risk models and comparing them with the historical credit risk calculations which use the traditional weights, and evaluating the results. Unfortunately, this comparison could not be done in this research because EY estimated that the data collection and comparison would take approximately another six months. It is expected that the method with the weights following from the HMM calculations will lead to better credit risk calculations, because the current method has no quantitative basis. The weights of the HMM are based on quantitative methods, and have proven to be able to correctly identify the economic states.

6.4 Recommendations for further research

Our aim was to use the hidden Markov model to see if there are quantitative ways to identify the economic scenarios and the weights associated with arriving in a scenario. It is of course important that the model is then able to distinguish the economic states. We have shown with this research that this is possible, provided that a three-state higher-order hidden Markov model is used with cutoff points for extreme outliers. Furthermore, we showed that the weights based on quantitative bases (the ones calculated with the HMM), are not in line with the weight distribution currently used. However, as mentioned in the previous section, given the time it was not possible to test these weights on the expected credit risk calculations as well. The first and most important point for further research would therefore be to test the weights (transition probabilities) of the hidden Markov model to see what the influence of this would be on credit risk calculations. This could be done by comparing the credit risk calculations with the weights (transition probabilities) of the HMM with the historical credit risk calculations which are using the traditional weights where the most likely scenario (baseline) accounts for 50% and the remaining two (upside and downside) share the other 50% equally. The next step would be evaluating which of the credit risk amounts would have been more sufficient.

The second point for further research has to do with data, many things could be further explored here. The impact on the performance of the HMM with additional economic features could be evaluated, even though the data collection of these will be challenging in many cases. Furthermore, other ways of dealing with missing data could be looked into. For instance, it would be interesting to see if the performance of the HMM would improve is the expectation maximization technique was used to manage the missing data. The expectation maximization technique first calculates the means, variances, and covariances of features for which the data is complete and than uses maximum likelihood procedures to estimate the missing values of the features which do not have all data available.

Keeping track of the data of the features for many countries started not so long ago, another data related point for further research would be to prolong this time frame with the use of data of features from countries which do have a longer time frame available. This could be done by using data from those countries which have a high correlation in the state sequences series produced by the HMM (or historical recession data if available) and prolong the data with machine learning methods.

Lastly, it was noticeable that the martingale method had a very good performance. As mentioned before, a martingale is a mathematical series in which the best prediction for the next number is the current number. In our case this results in the using the current state x_t as a prediction for the next state x_{t+1}. The martingale results in a single forward-looking scenario, which is not allowed unless adjustments to reflect non-linearity in the credit loss distribution are used. It would be an interesting point for further research to use the martingale method with adjustments, to validate if the method would outperform the current method.

References

Anderson, S. (2022). Guide to economic recession. Retrieved from https://www.investopedia.com/terms/r/recession.asp

ANP. (2020). Oud-topman Hamers van ING wordt vervolgd voor witwassen. Retrieved from https://www.1limburg.nl/oud-topman-Hamers-van-ing-wordt-vervolgd-voor-witwassen#:~:text=ING%20kreeg%20in%20september%202018,vervolging%20van%20individuele%20ING%2Dmedewerkers.

Buysse, K., & Essers, D. (2020). De impact van Covid-19 op de wereldeconomie: kunnen de opkomende economieën de motor aan de praat houden? Retrieved from https://www.nbb.be/nl/artikels/de-impact-van-covid-19-op-de-wereldeconomie-kunnen-de-opkomende-economieen-de-motor-aan-de

Castañón, J. (2019). 10 machine learning methods that every data scientist should know. Retrieved from https://towardsdatascience.com/10-machine-learning-methods-that-every-data-scientist-should-know-3cc96e0eeee9

Coulombe, et al. (2020). How is machine learning useful for macroeconomic forecasting?

CPB. (2021). Saffier 3.0: technische achtergrond. Retrieved from https://www.cpb.nl/saffier-3-0-technische-achtergrond

EBA. (2021). IFRS9 monitoring report. Retrieved from www.eba.europa.eu

EY. (2021). EY financial services Brexit tracker: Brexit activity muted over 2021, with total relocations revised down and few major operational announcements. Retrieved from https://www.ey.com/en_uk/news/2021/12/ey-financial-services-brexit-tracker-brexit-activity-muted-over-2021-with-total-relocations-revised-down-and-few-major-operational-announcements

Gareth, J., et al. (2021). *An Introduction to Statistical Learning* (2th ed.).

Garg, A. (2018). Complete guide to association rules. Retrieved from https://towardsdatascience.com/association-rules-2-aa9a77241654

Hall, M. (2021). Explaining the world through macroeconomic analysis. Retrieved from https://www.investopedia.com/insights/macroeconomic-analysis/

Joy, A. (2022). Pros and cons of supervised machine learning. Retrieved from https://pythonistaplanet.com/pros-and-cons-of-supervised-machine-learning/

Jurafsky, D., & Martin, J. (2021). Speech and language processing.

Klassen, M. (2020). What is BIC? Retrieved from https://towardsdatascience.com/an-intuitive-explanation-of-the-bayesian-information-criterion-71a7a3d3a5c5

Kose, M., et al. (2020). Global recessions. Retrieved from https://openknowledge.worldbank.org/handle/10986/33415

Kumar, A. (2020). Great mind maps for learning machine learning. Retrieved from https://vitalflux.com/great-mind-maps-for-learning-machine-learning/

Lutes, J. (2020). Correlation is simple with Seaborn and Pandas. Retrieved from https://towardsdatascience.com/correlation-is-simple-with-seaborn-and-pandas-28c28e92701e

Mankiw, N. G. (2020). The Covid-19 recession of 2020. Retrieved from https://scholar.harvard.edu/files/mankiw/files/covid-19_recession_of_2020_aug2020.pdf

Microsoft. (2022). *AI business school for financial services.* Retrieved from https://docs.microsoft.com/en-us/learn/paths/ai-business-school-financial-services/

Mishra, A. (2018). Metrics to evaluate your machine learning algorithm. Retrieved from https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234

Moreno, A. I. (2020). Moving averages with Python. Retrieved from https://towardsdatascience.com/moving-averages-in-python-16170e20f6c

Müller, A., & Guido, S. (2016). *Introduction to machine learning with Python: A guide for data scientists.*

NBER. (2008). Business cycle dating committee announcement. Retrieved from https://www.nber.org/news/business-cycle-dating-committee-announcement-january-7-2008

NU.nl. (2022a). 'perfect storm' leidt tot kilometerslange Brexit-files voor haven van Dover. Retrieved from https://www.nu.nl/brexit/6180988/perfect-storm-leidt-tot-kilometerslange-brexit-files-voor-haven-van-dover.html

NU.nl. (2022b). Spanning rond Oekraïne en grote vraag drijven olieprijs verder op. Retrieved from https://www.nu.nl/economie/6181195/spanning-rond-oekraine-en-grote-vraag-drijven-olieprijs-verder-op.html

Nykamp, D. (2019). State space definition. Retrieved from https://mathinsight.org/definition/state_space

ODSC. (2020). The a − z of supervised learning, use cases, and disadvantages. Retrieved from https://opendatascience.com/the-a-z-of-supervised-learning-use-cases-and-disadvantages/

Palachy, S. (2019). Stationarity in time series analysis.. Retrieved from https://towardsdatascience.com/stationarity-in-time-series-analysis-90c94f27322

Patterson, D. J. (2020a). Hidden Markov models 09: the forward-backward algorithm. Retrieved from https://www.youtube.com/watch?v=gYma8Gw38Os&list=PLix7MmR3doRo3NGNzrq48FItR3TDyuLCo&index=9&ab_channel=djp3

Patterson, D. J. (2020b). Hidden Markov models 11: the Viterbi algorithm. Retrieved from https://www.youtube.com/watch?v=s9dU3sFeE40&list=PLix7MmR3doRo3NGNzrq48FItR3TDyuLCo&index=11&ab_channel=djp3

Piatetsky-Shapiro, G. (1991). Discovery, analysis, and presentation of strong rules.

Stamp, M. (2004, 01). A revealing introduction to hidden Markov models. *Science*, 1-20.

Stehman, S. (1997). Selecting and interpreting measures of thematic classification accuracy. Retrieved from https://www.sciencedirect.com/science/article/abs/pii/S0034425797000837?via%3Dihub

Sutton, C., & McCallum, A. (2012). An introduction to conditional random fields. *Foundations and Trends® in Machine Learning*, *4*(4), 267-373. Retrieved from http://dx.doi.org/10.1561/2200000013 doi: 10.1561/2200000013

Wikipedia. (2022). Hidden Markov model. Retrieved from https://en.wikipedia.org/wiki/Hidden_Markov_model#Weather_guessing_game

Witten, I. H., et al. (2016). *Data Mining, Fourth Edition: Practical Machine Learning Tools and Techniques* (4th ed.). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Xiao, J., Liu, B., & Wang, X. (2005). Principles of non-stationary hidden Markov model and its applications to sequence labeling task". In *Second international joint conference on natural language processing: Full papers.* Retrieved from https://aclanthology.org/I05-1072 doi: 10.1007/11562214_72

Yates, T. (2021). 4 ways to predict market performance. Retrieved from https://www.investopedia.com/articles/07/mean_reversion_martingale.asp

Zajic, A. (2019). Introduction to AIC. Retrieved from https://towardsdatascience.com/introduction-to-aic-akaike-information-criterion-9c9ba1c96ced

Škoda, P., & Adam, F. (2020). *Knowledge Discovery in Big Data from Astronomy and Earth Observation.* Retrieved from https://www.sciencedirect.com/book/9780128191545/knowledge-discovery-in-big-data-from-astronomy-and-earth-observation

7 Appendix

7.1 Solution to the likelihood problem

Let λ = (A, B, π) be a given model, $\mathcal{O}$=$(O_0, O_1, ..., O_T)$ be a series of observations and $\mathcal{X} = (x_0, x_1, ..., x_T)$ be a state sequence. We want to find $P(\mathcal{O} \mid \lambda)$. First, we have to compute the joint probability of a particular state sequence $\mathcal{X}$ generating a particular sequence $\mathcal{O}$ of observations. We have (Jurafsky & Martin, 2021):

$$P(\mathcal{O}, \mathcal{X} \mid \lambda) = P(\mathcal{O} \mid \mathcal{X}, \lambda)P(\mathcal{X} \mid \lambda) = \prod_{t=1}^{T} P(O_t \mid x_t) * \prod_{t=1}^{T} P(x_t \mid x_{t-1}). \tag{34}$$

Note that the one-to-one mapping in the formula is a result of the Markov assumption given in the Assumption 1. Now that we know how to get the joint probability distribution of the observations with a particular hidden state sequence we can compute the total probability of the observations by summing over all possible hidden state sequences (Jurafsky & Martin, 2021). By summing over all possible state sequences we obtain:

$$P(\mathcal{O} \mid \lambda) = \sum_{\mathcal{X}} P(\mathcal{O}, \mathcal{X} \mid \lambda) = \sum_{\mathcal{X}} P(\mathcal{O} \mid \mathcal{X}, \lambda)P(\mathcal{X} \mid \lambda). \tag{35}$$

For an HMM with N hidden states and an observation sequence of T observations, there are N^T possible hidden sequences. This computation is generally infeasible since it requires too many multiplications (Stamp, 2004).

We go back to the weather example to illustrate the multiplication problem. We begin with the calculation of the probability of the observed sequence $\mathcal{O} = \{Cleaning, Shopping, Walking\}$ given the parameters of our HMM $\lambda = (A, B, \pi)$, which are given in Figure 49. As such, we are looking for the probability $P(\mathcal{O} = O_1, O_2, O_3)$. To compute this we need to consider all the sequences of hidden states that might produce this observed sequence. Take for instance the state sequence $\mathcal{X} = \{Rainy, Sunny, Sunny\}$. The computation for the probability of this state sequence is the following:

$$\pi_1 b_1(O_1) a_{12} b_2(O_2) a_{22} b_2(O_3) = 0.6 * 0.5 * 0.3 * 0.3 * 0.6 * 0.6 = 0.00972.$$

Figure 48 visualizes the aforementioned process.

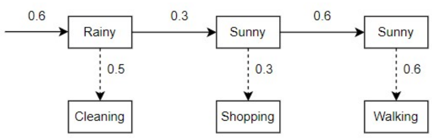

Figure 48: Example of the calculation of the likelihood for one particular sequence of observations and states. The first black arrow indicates with what probability the sequence starts in the "Rainy" state (π_1), the other black arrows give the state transition probabilities (a_{12} and a_{22}) and the dotted arrows give the emission probabilities ($b_1(O_1)$, $b_2(O_2)$, $b_2(O_3)$).

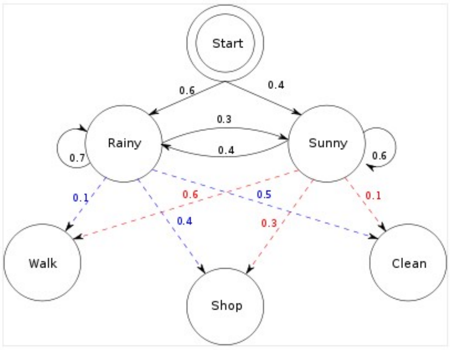

Figure 49: For the convenience of the reader we repeat Figure 9.

An example is given above about the calculation of the likelihood problem. However, this is just an example of a scenario with a rather small sequence. If we calculate all likelihoods of all possible state sequences we are already on $2^3 = 8$ (2 possible hidden states and sequences of 3 possible state sequences). If we calculate everything in this way, the number of multiplications becomes huge, it would be $2TN^T$ where T is the length of the observed sequence and N is the number of hidden states. So, the number of multiplications grows exponentially with the number of hidden states, this is illustrated in Figure 50.

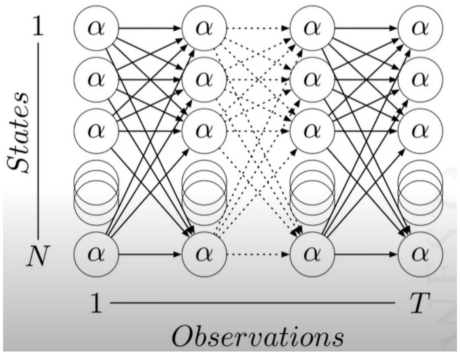

Figure 50: Example of the growth of multiplications with the number of states N and length of the sequence T. There can be seen that a lot of the sequences follow largely the same path, and thus can be calculated more efficiently with the forward algorithm (Patterson, 2020a).

The forward algorithm

As mentioned in the previous section, the number of multiplications grows exponentially with the number of hidden states. Hence a more efficient way of doing these multiplications is needed. In Figure 50 can be seen that a large part of the sequences is the same, the forward algorithm makes use of this. The forward algorithm or α-pass stores the results and uses them in future multiplications (Jurafsky & Martin, 2021).

$\alpha_t(j)$ represents the probability of being in state j after seeing the first t observations, given the HMM λ. The value of each cell $a_t(j)$ is computed by summing over the probabilities of every path that could lead us to this cell:

$$\alpha_t(j) = P(O_1, O_2...O_t, x_t = q_j \mid \lambda), \quad 1 \leq t \leq T, \quad 1 \leq j \leq N. \tag{36}$$

Here, $x_t = j$ means the t^{th} state in a sequence of states in state q_j. We compute this probability $a_t(j)$ by summing over all extensions of all paths that lead to the current cell. For a given state x_j at time t, the value $a_t(j)$ is computed as:

$$\alpha_t(j) = \sum_{i=1}^{N} \alpha_{t-1}(i) a_{ij} b_j(O_t). \tag{37}$$

Here:

- $\alpha_{t-1}(i)$ = the previous forward path probability from the previous time step.
- a_{ij} = the transition probability from previous state x_i to current state x_j.
- $b_j(O_t)$ = the state observation likelihood of the observation symbol O_t given the current state j.

Therefore, all t, j the $\alpha_t(j)$ can be computed recursively:

1. Initial step:
$$\alpha_1(j) = \pi_j b_j(O_j), \quad 1 \leq j \leq N. \tag{38}$$
2. For $1 \leq t \leq T-1$ and $1 \leq j \leq N$:
$$\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(i) a_{ij}\right] b_j(O_{t+1}). \tag{39}$$
3. Termination. From the definition of $\alpha_j(t)$, we find:
$$P(\mathcal{O} \mid \lambda) = \sum_{j=1}^{N} \alpha_T(j). \tag{40}$$

The backward algorithm
Where a defines what the probability is of being in a state given everything that has happened before, β is the probability of being in a state given what is coming ahead. β is the probability we are going to see the sequence of observations that we know is coming given that we are starting in a state right now.

We can define the backward variable in similar fashion. Let $\mathcal{O} = (O_{t+1}, O_{t+2}, ..., O_T)$ be the observation sequence from time $t+1$ until time T. Instead of moving forward step by step we will move backward step by step. we can define the backward variable:

$$\beta_t(i) = P(O_{t+1}, O_{t+2}...O_T, x_t = q_i \mid \lambda). \tag{41}$$

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(O_{t+1}) \beta_{t+1}(j). \tag{42}$$

Figure 51 illustrates the calculation of the backward variables.

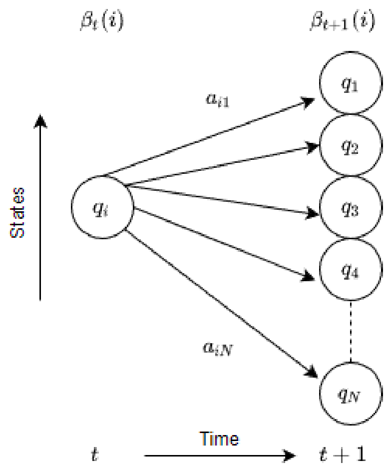

Figure 51: To calculate the value of the backward variable at time t, you only need the values of the forward variables for all states at time $t+1$. In this figure q are the values of the distinct states, a_{ij} is the probability of transitioning from state i to j, and β is the backward variable.

So the probability of $\beta_t(i)$ is the sum over all of the states that we might go to of the parameter β_{t+1} and we move backwards into state j and therefore use the state observation likelihood $b_j(O_t)$ and multiply that with the probability a_{ij}, the probability that we go from i to j. Note that this is the opposite of the forward algorithm; given the probability of where we could possibly be, what is the probability of having seen the observation there and what is the probability of moving there. So the $\beta_t(i)$ can be calculated recursively:

1. Initial step:
$$\beta_T(i) = 1, \quad 1 \leq i \leq N. \tag{43}$$

2. Inductive step: for $t = T-1, T-2, ..., 1$ and $1 \leq j \leq N$,
$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(O_{t+1}) \beta_{t+1}(j). \tag{44}$$

It is possible to combine the forward and backward algorithm in which case we would get:

$$P(\mathcal{O} \mid \lambda) = \sum_{j=1}^{N} \alpha_t(j) \beta_t(j). \tag{45}$$

7.2 Solution to the learning problem

Lastly, there is the learning problem. Given an observation sequence $\mathcal{O}$=$(O_0, O_1, ..., O_T)$ and a possible set of states in the HMM $\mathcal{S} = \{q_0, q_1, ..., q_N\}$, learn the HMM parameters A, B and π. In other words, given a set of observation sequences, how do we learn the model parameters that would generate them? Formally this is given by:

$$\lambda^* = arg \max_{\lambda} P(\mathcal{O} \mid \lambda). \tag{46}$$

There is no method to solve for the globally optimal parameters of λ. We, therefore, search for a locally optimal result, a result that is a good answer but is not guaranteed to be the best answer.

The input for this problem would be an unlabeled sequence of observations $\mathcal{O}$ and the potential hidden states $\mathcal{S}$. The standard algorithm for the HMM learning problem is the forward-backward, or Baum-Welch algorithm, a special case of the Expectation-Maximization or EM algorithm. The EM algorithm is iterative, computing an initial estimate for the probabilities, then using those estimates to compute a better estimate, and so on, iteratively improving the probabilities that it learns (Jurafsky & Martin, 2021). To do this we define:

$$\xi_t(i,j) = P(x_t = q_i, x_{t+1} = q_j \mid \mathcal{O}, \lambda). \tag{47}$$

The variable ξ captures the probability that at time t, we are are at state q_i and in time $t+1$ we are in state q_j given our set of observation sequences $\mathcal{O}$ and our model λ. For the calculation of ξ many of the previous sections are repeated. The following equation is able to determine the value of ξ:

$$\xi_t(i,j) = \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{P(\mathcal{O} \mid \lambda)}. \tag{48}$$

We explain the formula below. First a recap of the parameters used in this Equation 48:

- $\alpha_t(i)$ is the forward variable. It is the probability of being in state i after seeing the first t observations, given the model λ.
- a_{ij} is the state transition probability from previous state x_i to current state x_j.
- $b_j(O_{t+1})$ is the state observation likelihood of the observation symbol O_t given the current state j.
- $\beta_{t+1}(j)$ is the backward variable. it is the probability of being in state j at time t given everything that comes after t, given the model λ.
- $P(\mathcal{O} \mid \lambda)$ is the probability of an observation sequence occurring given the model λ.

Figure 52 visualizes the calculation of ξ. Note all the states we could be in q_1 to q_N lined up and regardless of how we got there, ξ is capturing the probability that we end up in q_i, given our observations. For this part of the calculation the forward variable $\alpha_t(i)$ can be used. The next step is to transition to q_j, for this transition the $a_{ij} b_j(O_{t+1})$ comes in, where α_{ij} is the probability of transitioning from i to j and we multiply this by the probability of seeing our observation in time $t+1$ given that we are now in state j. And backward variable $\beta_{t+1}(j)$ is capturing the right half of the transition of the path. The $P(\mathcal{O} \mid \lambda)$ is the normalizing factor to account for the fact that we want to calculate a probability. We can get $P(\mathcal{O} \mid \lambda)$ by summing over all the probabilities of all i's and j's at the given time, so therefore we sum over i and j, resulting in the following formula for ξ.

$$\xi_t(i,j) = \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}. \tag{49}$$

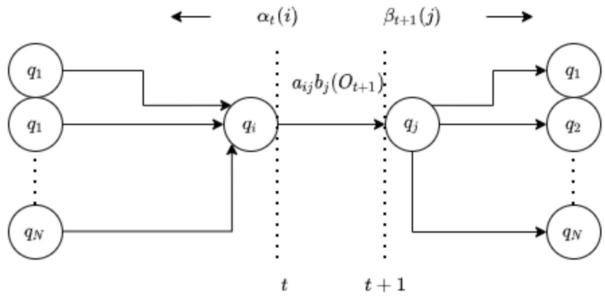

Figure 52: Visualization of the calculation of ξ.

In Equation 11 we gave the formula for $\gamma_t(i)$, which is the probability of being in state q_1 at time t. We recap the formula for $\gamma_t(i)$ in the equation below.

$$\gamma_t(j) = P(q_t = x_j \mid \mathcal{O}, \lambda), \quad 1 \leq j \leq N.$$

Note that there is a clear relationship between γ and ξ. γ is just one point in time, so if we sum over all possible destinations of being in i and going to j, we sum over all possible j's therefore must be $\gamma_t(i)$. Formally:

$$\gamma_t(i) = \sum_{j=1}^{N} \xi_t(i,j). \tag{50}$$

Since ξ is the probability of transitioning from q_1 to q_j at time t, we could sum over all t to acquire a number that can be treated as the expected number of times q_i transitions to q_j. If we sum over all t for $\gamma_t(i)$ we get the expected number of transtions from q_i, resulting in the following equations:

$$\mathbb{E}[\text{ number of transitions from } q_i \text{ to } q_j] = \sum_{t=1}^{T-1} \xi_t(i,j). \tag{51}$$

$$\mathbb{E}[\text{ number of transitions from } q_i] = \sum_{t=1}^{T-1} \gamma_t(i). \tag{52}$$

So, how can we use this to improve the model $\lambda(A, B, \pi)$? We start with the initial state distribution π. π is the probability of being in state q_i at time $t = 1$, therefore the re-estimation formula for the initial state distribution is given by:

$$\pi_i^* = \gamma_1(i), \quad 1 \leq i \leq N. \tag{53}$$

The state transition distribution a_{ij} is the expected number of transitions from q_i to q_j divided by the expected total number of transitions out of q_i:

$$a_{ij}^* = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum^{T-1} \gamma_t(i)}, \quad 1 \leq j \leq N, \quad 1 \leq j \leq M. \tag{54}$$

The observation probability distribution $b_j(k)$ for a discrete HMM is the expected number of visits to state q_j, where k is the observed signal, divided by the expected total number of visits to state q_j:

$$b_j^*(k) = \frac{\sum_{t=1}^{T} I_t^k \gamma_t(i)}{\sum_{t=1}^{T} \gamma_t(i)}. \tag{55}$$

Here,

$$I_t^k = \begin{cases} 1, & \text{if } O_t = v_k \\ 0, & \text{otherwise} \end{cases} \tag{56}$$

the numerator is subject to $O_t = v_k$. The learning process can now be defined as follows:

1. Initialization of the model $\lambda = (A, B, \pi)$.
2. Re-estimation of the state transition distribution A, the observation probability distribution B, and the initial state distribution π and define the adjusted model $\lambda^*(A^*, B^*, \pi^*)$.
3. If $P(\mathcal{O} \mid \lambda^*) > P(\mathcal{O} \mid \lambda)$ then $\lambda = \lambda^*$ and proceed to Step 1 until a limiting point is reached.

7.3 The two-state hidden Markov base model

7.3.1 Visual results

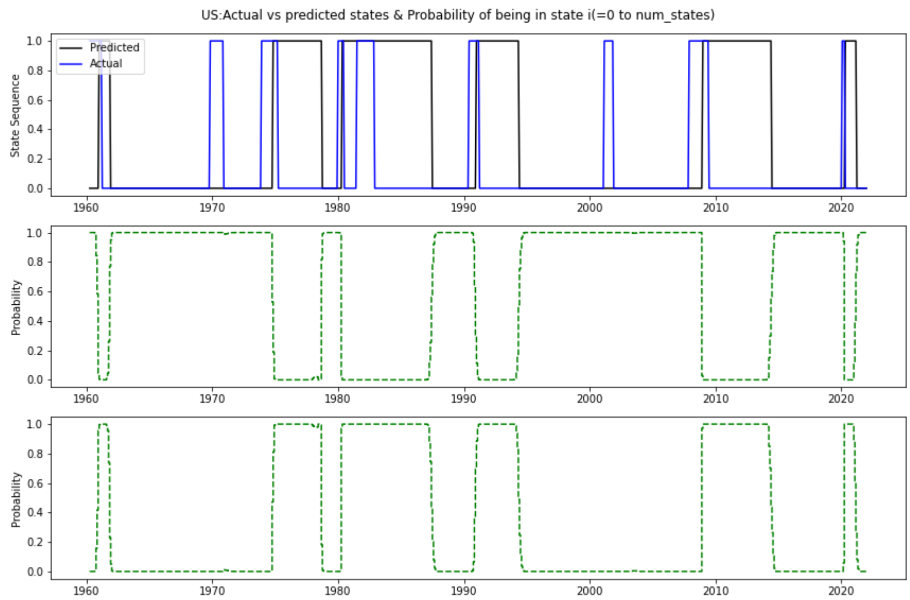

Figure 53: Prediction of the US versus against historical recession data and probability of being in state i (= 0 to 1) according to the model. The blue line in the top graph shows when the historical recessions were. The black line shows when the model predicts a recession. At best, these two graphs would overlap exactly, or be exactly mirrored. This is because the model sometimes gives the recession state a 0 and the other time a 1. The other graphs in the Figure indicate the probability according to the model that a country is in a particular hidden state at that moment.

.

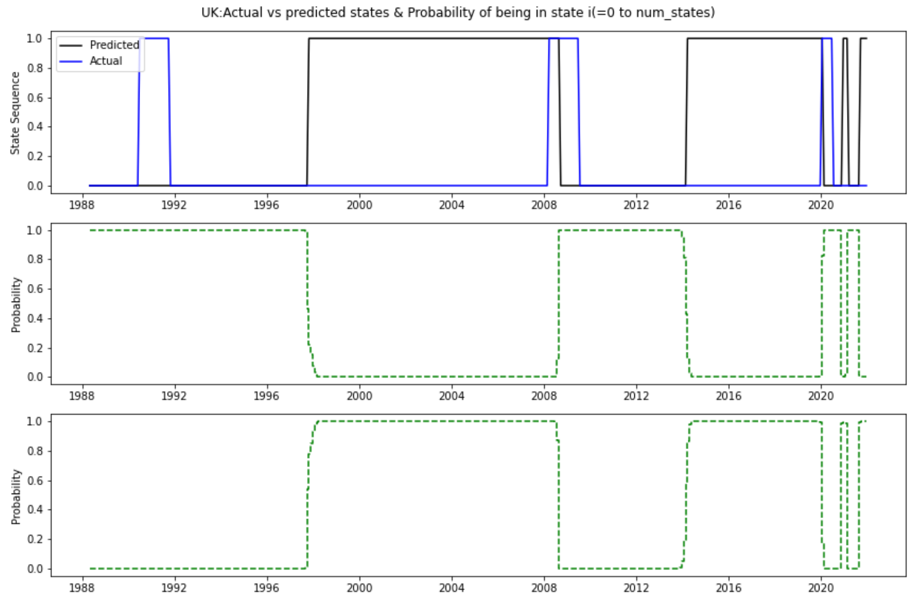

Figure 54: Prediction of the UK versus against historical recession data and probability of being in state i (= 0 to 1) according to the model. The blue line in the top graph shows when the historical recessions were. The black line shows when the model predicts a recession. At best, these two graphs would overlap exactly, or be exactly mirrored. This is because the model sometimes gives the recession state a 0 and the other time a 1. The other graphs in the Figure indicate the probability according to the model that a country is in a particular hidden state at that moment.

.

7.4 The improved two-state hidden Markov model

7.4.1 Quantitative results of all experiments

Table 8: Results performance metrics for the experiments of the United States, United Kingdom, and Japan for the two-state hidden Markov model.

Run	Country	Accuracy	Precision	Recall	F-score
1	US	0.80	0.41	0.82	0.54
1	UK	0.70	0.17	0.58	0.27
1	JP	0.58	0.55	0.55	0.55
2	US	0.53	0.11	0.32	0.16
2	UK	0.77	0.29	0.95	0.44
2	JP	0.61	0.62	0.44	0.52
3	US	0.59	0.16	0.44	0.23
3	UK	0.69	0.23	0.95	0.37
3	JP	0.76	0,84	0.60	0.70
4	US	0.79	0.39	0.92	0.55
4	UK	0.41	0.13	0.92	0.23
4	JP	0.56	0.54	0.49	0.51
5	US	0.77	0.36	0.83	0.50
5	UK	0.70	0.17	0.58	0.27
5	JP	0.56	0.53	0.50	0.51
6	US	0.48	0.16	0.61	0.25
6	UK	0.60	0.19	0.97	0.31
6	JP	0.59	0.57	0.54	0.55
7	US	0.79	0.39	0.92	0.55
7	UK	0.70	0.17	0.58	0.27
7	JP	0.56	0.53	0.51	0.52
8	US	0.80	0.41	0.82	0.54
8	UK	0.68	0.21	0.87	0.34
8	JP	0.58	0.55	0.55	0.55
9	US	0.80	0.41	0.92	0.57
9	UK	0.78	0.24	0.61	0.34
9	JP	0.58	0.55	0.51	0.53
10	US	0.78	0.39	0.92	0.55
10	UK	0.87	0.35	0.45	0.40
10	JP	0.57	0.55	0.48	0.51
11	US	0.79	0.39	0.92	0.55
11	UK	0.65	0.21	1.00	0.35
11	JP	0.63	0.61	0.62	0.61
12	US	0.48	0.18	0.76	0.29
12	UK	0.60	0.13	0.58	0.21
12	JP	0.57	0.54	0.47	0.50
13	US	0.66	0.29	0.93	0.44
13	UK	0.80	0.25	0.58	0.35
13	JP	0.57	0.54	0.51	0.53

14	US	0.77	0.37	0.94	0.54
14	UK	0.77	0.29	1.00	0.45
14	JP	0.54	0.51	0.36	0.42
15	US	0.66	0.29	0.93	0.44
15	UK	0.75	0.27	0.97	0.42
15	JP	0.57	0.54	0.51	0.53
16	US	0.66	0.29	0.93	0.44
16	UK	0.80	0.25	0.58	0.35
16	JP	0.57	0.54	0.51	0.53
17	US	0.63	0.20	0.55	0.29
17	UK	0.77	0.29	0.95	0.44
17	JP	0.64	0.63	0.58	0.61
18	US	0.75	0.34	0.80	0.48
18	UK	0.87	0.22	0.16	0.18
18	JP	0.69	0,65	0.70	0.68
19	US	0.63	0.20	0.56	0.30
19	UK	0.54	0.14	0.76	0.24
19	JP	0.57	0.56	0.40	0.47
20	US	0.44	0.20	0.96	0.33
20	UK	0.78	0.28	0.84	0.42
20	JP	0.60	0.60	0.42	0.49
21	US	0.64	0.27	0.92	0.42
21	UK	0.79	0.30	0.95	0.46
21	JP	0.60	0.57	0.61	0.59
22	US	0.63	0,27	0.92	0.42
22	UK	0.75	0.27	1.00	0.42
22	JP	0.54	0.51	0.49	0.50
23	US	0.66	0.27	0.83	0.41
23	UK	0.79	0.30	0.95	0.46
23	JP	0.54	0.52	0.28	0.36
24	US	0.66	0.28	0.93	0.43
24	UK	0.75	0.27	0.97	0.42
24	JP	0.56	0.54	0.43	0.48
25	US	0.70	0.31	0.92	0.47
25	UK	0.64	0.20	0.92	0.33
25	JP	0.67	0.63	0.71	0.67
26	US	0.64	0.27	0.92	0.42
26	UK	0.78	0.29	0.95	0.44
26	JP	0.50	0.47	0.58	0.52
27	US	0.64	0.27	0.93	0.42
27	UK	0.75	0.27	0.97	0.42
27	JP	0.62	0.59	0.64	0.61
28	US	0.66	0.10	0.17	0.13
28	UK	0.49	0.10	0.58	0.17
28	JP	0.61	0.59	0.57	0.58
29	US	0.59	0.12	0.30	0.17
29	UK	0.79	0.30	0.95	0.46
29	JP	0.52	0.49	0.49	0.49

30	US	0.64	0.27	0.92	0.42
30	UK	0.79	0.30	0.95	0.46
30	JP	0.60	0.57	0.61	0.59
31	US	0.58	0.15	0.43	0.23
31	UK	0.68	0.22	0.95	0.36
31	JP	0.65	0.64	0.56	0.60
32	US	0.60	0.16	0.44	0.24
32	UK	0.76	0.28	0.97	0.44
32	JP	0.68	0.84	0.40	0.54
33	US	0.47	0.16	0.63	0.25
33	UK	0.68	0.22	0.95	0.36
33	JP	0.64	0.61	0.66	0.63
34	US	0.51	0.16	0.58	0.25
34	UK	0.58	0.17	0.95	0.30
34	JP	0.68	0.69	0.59	0.64
35	US	0.58	0.16	0.44	0.23
35	UK	0.68	0.22	0.95	0.36
35	JP	0.75	0.83	0.59	0.69
36	US	0.59	0.15	0.43	0.23
36	UK	0.69	0.23	0.95	0.37
36	JP	0.76	0.84	0.60	0.70
37	US	0.59	0.15	0.43	0.23
37	UK	0.69	0.23	0.95	0.37
37	JP	0.76	0.84	0.59	0.69
38	US	0.58	0.15	0.43	0.23
38	UK	0.69	0.23	0.95	0.37
38	JP	0.76	0.84	0.60	0.70
39	US	0.68	0.28	0.82	0.42
39	UK	0.69	0.23	0.95	0.37
39	JP	0.75	0.84	0.57	0.68
40	US	0.59	0.16	0.44	0.23
40	UK	0.69	0.23	0.95	0.37
40	JP	0.76	0.84	0.59	0.69
41	US	0.59	0.16	0.44	0.23
41	UK	0.69	0.23	0.95	0.37
41	JP	0.76	0.84	0.60	0.70
42	US	0.59	0.16	0.44	0.23
42	UK	0.69	0.23	0.95	0.37
42	JP	0.76	0.84	0.60	0.70
43	US	0.61	0.19	0.54	0.28
43	UK	0.75	0.26	0.95	0.41
43	JP	0.80	0.90	0.64	0.75

7.4.2 Parameters used for all experiments

Table 9: Parameters used for all experiments of the 2 state hidden Markov model.

Run	**Cutoff point**	**CPI s=**	**UNR s=**	**GDP s=**	**index s=**	**itr_diff s=**	**# gm**	**x**
1	3SD	0.1	N/A	0.8	0.1	0.05	1	0.1
2	4SD	0.1	N/A	0.8	0.3	0.05	1	0.1
3	4SD	0.1	N/A	0.8	1	1	1	0.1
4	2SD	0.1	N/A	0.8	0.1	0.05	1	0.1
5	4SD	0.1	N/A	0.8	0.1	0.05	1	0.1
6	3.5SD	0.1	N/A	0.8	0.1	0.05	1	0.1
7	3SD	0.1	N/A	0.7	0.1	0.05	1	0.1
8	3SD	0.1	N/A	0.9	0.1	0.05	1	0.1
9	3SD	0.2	N/A	0.8	0.1	0.05	1	0.1
10	3SD	0.3	N/A	0.8	0.1	0.05	1	0.1
11	3SD	0.2	N/A	0.8	0.2	0.05	1	0.1
12	3SD	0.2	N/A	0.8	0.05	0.05	1	0.1
13	3SD	0.2	N/A	0.8	0.1	0.1	1	0.1
14	3SD	0.2	N/A	0.8	0.1	N/A	1	0.1
15	3SD	0.2	N/A	0.8	0.1	0.05	2	0.1
16	3SD	0.2	N/A	0.8	0.1	0.05	1	0.05
17	3.5SD	0.1	N/A	0.8	0.3	0.05	1	0.1
18	NVT	0.1	N/A	0.8	0.3	0.05	1	0.1
19	4SD	0.1	1	0.8	0.3	0.05	1	0.1
20	4SD	N/A	N/A	0.8	0.3	0.05	1	0.1
21	4SD	0.2	N/A	0.8	0.3	0.05	1	0.1
22	4SD	0.3	N/A	0.8	0.3	0.05	1	0.1
23	4SD	0.2	N/A	0.7	0.3	0.05	1	0.1
24	4SD	0.2	N/A	0.6	0.3	0.05	1	0.1
25	4SD	0.2	N/A	0.8	0.2	0.05	1	0.1
26	4SD	0.2	N/A	0.8	0.4	0.05	1	0.1
27	4SD	0.2	N/A	0.8	0.3	0.1	1	0.1
28	4SD	0.2	N/A	0.8	0.3	0.01	1	0.1
29	4SD	0.2	N/A	0.8	0.3	0.05	2	0.1
30	4SD	0.2	N/A	0.8	0.3	0.05	1	0.05
31	3SD	0.1	N/A	0.8	1	1	1	0.1
32	4SD	0.2	N/A	0.8	1	1	1	0.1
33	4SD	0.05	N/A	0.8	1	1	1	0.1
34	4SD	0.1	1	0.8	1	1	1	0.1
35	4SD	0.1	N/A	0.7	1	1	1	0.1
36	4SD	0.1	N/A	0.9	1	1	1	0.1
37	4SD	0.1	N/A	1	1	1	1	0.1
38	4SD	0.1	N/A	0.8	0.9	1	1	0.1
39	4SD	0.1	N/A	0.8	0.8	1	1	
40	4SD	0.1	N/A	0.8	1	0.9	1	0.1
41	4SD	0.1	N/A	0.8	1	1	2	0.1
42	4SD	0.1	N/A	0.8	1	1	1	0.05 43
4SD	0.1	N/A	0.8	1	0.05	1	0.10	

7.4.3 Visual results of best experiments

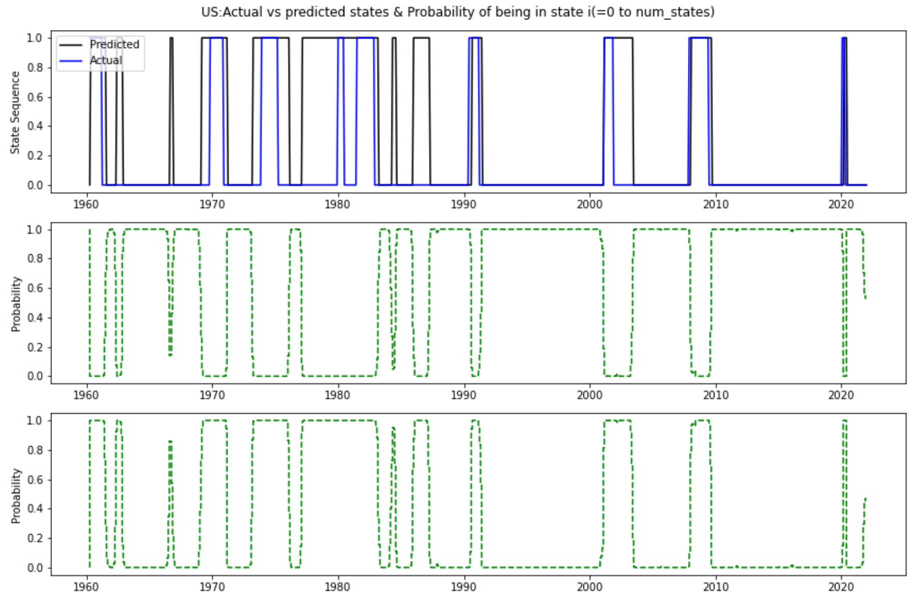

Figure 55: Prediction of the US versus against historical recession data and probability of being in state i (= 0 to 1) according to the model. The blue line in the top graph shows when the historical recessions were. The black line shows when the model predicts a recession. At best, these two graphs would overlap exactly, or be exactly mirrored. This is because the model sometimes gives the recession state a 0 and the other time a 1. The other graphs in Figure indicate the probability according to the model that a country is in a particular hidden state at that moment.

.

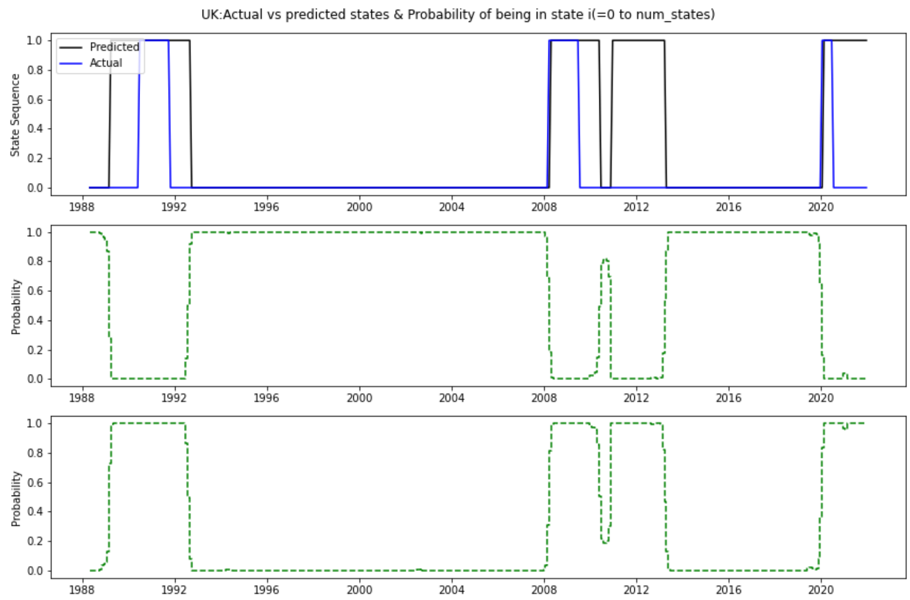

Figure 56: Prediction of the UK versus against historical recession data and probability of being in state i (= 0 to 1) according to the model. The blue line in the top graph shows when the historical recessions were. The black line shows when the model predicts a recession. At best, these two graphs would overlap exactly, or be exactly mirrored. This is because the model sometimes gives the recession state a 0 and the other time a 1. The other graphs in Figure indicate the probability according to the model that a country is in a particular hidden state at that moment.

.

7.5 The three-state hidden Markov model

7.5.1 Quantitative results of all experiments

Table 10: Results of all experiments for the 3 state hidden Markov model.

Run	Country	Accuracy	Precision	Recall	F-score
1	US	0.89	0.58	0.92	0.71
1	UK	0.85	0.36	0.82	0.50
1	JP	0.61	0.73	0.28	0.40
2	US	0.90	0.59	0.90	0.72
2	UK	0.87	0.43	1.0	0.60
2	JP	0.57	0.56	0.32	0,41
3	US	0.88	0.55	0.90	0.68
3	UK	Does not converge			
3	JP	0.58	0.56	0.55	0.55
4	US	Does not converge			
4	UK	Does not converge			
4	JP	Does not converge			

5	US	0.88	0.55	0.93	0.69
5	UK	0.82	0.32	0.76	0.45
5	JP	0.59	0.63	0.29	0.39
6	US	0.91	0.61	0.95	0.75
6	UK	0.79	0.30	0.92	0.45
6	JP	0.60	0.68	0.27	0.39
7	US	0.91	0.61	0.95	0.75
7	UK	0.79	0.30	0.92	0.45
7	JP	0.60	0.68	0.27	0.39
8	US	0.89	0.57	0.91	0.71
8	UK	0.74	0.20	0.61	0.30
8	JP	0.66	0.69	0.5	0.58
9	US	Does not converge			
9	UK	Does not converge			
9	JP	Does not converge			
10	US	0.81	0.41	0.81	0.54
10	UK	0.75	0.24	0.76	0.36
10	JP	0.56	0.60	0.19	0.29
11	US	0.83	0.44	0.88	0.59
11	UK	0.83	0.34	0.79	0.47
11	JP	0.69	0.86	0.40	0.54
12	US	0.69	0.18	0.34	0.24
12	UK	Does not converge			
12	JP	0.60	0.59	0.51	0.54
13	US	0.66	0.09	0.15	0.11
13	UK	0.90	0.47	0.39	0.43
13	JP	0.62	0.60	0.56	0.58
14	US	0.63	0.25	0.82	0.39
14	UK	0.91	0.52	0.66	0.58
14	JP	0.55	0.53	0.31	0.39
15	US	0.88	0.55	0.80	0.65
15	UK	Does not converge			
15	JP	0.64	0.64	0.53	0.58
16	US	0.62	0.23	0.72	0.35
16	UK	0.89	0.38	0.24	0.29
16	JP	0.67	0.71	0.49	0.58
17	US	0.68	0.22	0.50	0.30
17	UK	0.74	0.22	0.68	0.33
17	JP	0.60	0.59	0.52	0.55
18	US	0.82	0.42	0.77	0.55
18	UK	0.75	0.25	0.82	0.38
18	JP	0.60	0.59	0.47	0.53
19	US	0.81	0.41	0.75	0.53
19	UK	0.77	0.21	0.53	0.30
19	JP	0.68	0.83	0.40	0.54
19	US	0.85	0.49	0.78	0.60
19	UK	0.71	0.20	0.71	0.32
19	JP	0.63	0.76	0.31	0.44

20	US	0.83	0.45	0.84	0.58
20	UK	0.86	0.05	0.03	0.04
20	JP	Does not converge			
21	US	0.87	0.53	0.70	0.61
21	UK	0.88	0.25	0.16	0.19
21	JP	0.56	0.53	0.44	0.48
22	US	0.91	0.61	0.96	0.75
22	UK	0.83	0.34	0.89	0.50
22	JP	0.63	0.73	0.33	0.46
23	US	0.89	0.57	0.94	0.71
23	UK	0.70	0.19	0.66	0.29
23	JP	0.49	0.45	0.36	0.40
24	US	0.90	0.60	0.95	0.74
24	UK	0.84	0.36	0.92	0.51
24	JP	0.61	0.70	0.29	0.41
25	US	0.88	0.56	0.84	0.67
25	UK	0.86	0.29	0.32	0.30
25	JP	0.67	0.80	0.38	0.52
26	US	0.89	0.58	0.93	0.72
26	UK	0.82	0.31	0.74	0.43
26	JP	0.60	0.71	0.23	0.34
27	US	0.91	0.61	0.96	0.75
27	UK	0.82	0.33	0.87	0.48
27	JP	0.60	0.76	0.20	0.32
28	US	0.91	0.61	0.96	0.75
28	UK	0.82	0.33	0.87	0.48
28	JP	0.62	0.80	0.26	0.39
29	US	0.83	0.45	0.83	0.58
29	UK	Does not converge			
29	JP	0.66	0.66	0.58	0.62
30	US	Does not converge			
30	UK	0.84	0.31	0.58	0.40
30	JP	0.53	0.49	0.36	0.41
31	US	0.63	0.25	0.81	0.38
31	UK	0.75	0.13	0.29	0.18
31	JP	0.63	0.66	0.42	0.51
32	US	Does not converge			
32	UK	0.84	0.31	0.58	0.40
32	JP	0.53	0.49	0.36	0.41
33	US	0.63	0.25	0.82	0,39
33	UK	0.91	0.52	0.66	0,58
33	JP	0.55	0.52	0.41	0,46
34	US	0.63	0.25	0.82	0.39
34	UK	0.91	0.52	0.66	0.58
34	JP	0.53	0.50	0.38	0.43
35	US	0.90	0.61	0.89	0.72
35	UK	0.89	0.45	0.66	0.53

35	JP	0.60	0.61	0.43	0.50
36	US	0.90	0.59	0.90	0.71
36	UK	0.87	0.42	0.97	0.59
36	JP	0.52	0.49	0.33	0.39
37	US	Does not converge			
37	UK	0.83	0.33	0.76	0.46
37	JP	0.56	0.55	0.37	0.44
38	US	0.89	0.58	0.90	0.70
38	UK	0.84	0.37	0.95	0.53
38	JP	0.53	0.50	0.23	0.32
39	US	0.85	0.46	0.51	0.49
39	UK	0.88	0.42	0.66	0.52
39	JP	0.59	0.60	0.36	0.45
40	US	0.88	0.55	0.91	0.69
40	UK	0.90	0.47	0.76	0.58
40	JP	0.60	0.59	0.43	0.50
41	US	0.88	0.55	0.90	0.69
41	UK	0.85	0.38	0.97	0.55
41	JP	0.61	0.61	0.45	0.52
42	US	0.88	0.55	0.90	0,69
42	UK	0,87	0.43	1.00	0.60
42	JP	0.53	0.50	0.36	0.42
43	US	0.82	0.42	0.77	0.55
43	UK	0.75	0.25	0.82	0.38
43	JP	0.60	0.59	0.47	0.53
44	US	0.84	0.46	0.73	0.57
44	UK	Does not converge			
44	JP	0.71	0.81	0.49	0.61
45	US	Does not converge			
45	UK	0.76	0.28	0.97	0.44
45	JP	0.66	0.76	0.39	0.52
46	US	0.87	0.52	0.90	0.66
46	UK	0.83	0.35	0.92	0.51
46	JP	0.58	0.56	0.53	0.54
47	US	0.84	0.46	0.72	0.57
47	UK	0.89	0.37	0.29	0.32
47	JP	0.57	0.54	0.53	0.54
48	US	0.81	0.40	0.72	0.52
48	UK	0.80	0.26	0.63	0.37
48	JP	0.59	0.59	0.41	0.48
49	US	0.85	0.48	0.76	0.59
49	UK	Does not converge			
49	JP	0.73	0.83	0.54	0.65
50	US	0.86	0.49	0.82	0.62
50	UK	Does not converge			
50	JP	0.73	0.83	0.53	0.65
51	US	0.88	0.55	0.96	0.70
51	UK	Does not converge			

51	JP	0.72	0.82	0.51	0.63
52	US	0.87	0.51	0.82	0.63
52	UK	Does not converge			
52	JP	0.74	0.84	0.55	0.67
53	US	0.84	0.47	0.76	0.58
53	UK	Does not converge			
53	JP	0.76	0.88	0.56	0.68
54	US	0.80	0.40	0.72	0.51
54	UK	0.81	0.27	0.61	0.38
54	JP	0.77	0.87	0.58	0.70
55	US	0.84	0.45	0.76	0.57
55	UK	0.85	0.31	0.45	0.37
55	JP	0.68	0.85	0.38	0.52
56	US	0.80	0.40	0.73	0.52
56	UK	0.81	0.27	0.61	0.38
56	JP	0.74	0.86	0.53	0.65
57	US	0.80	0.40	0.73	0.51
57	UK	0.81	0.27	0.61	0.38
57	JP	0.77	0.87	0.58	0.70

7.5.2 Parameters used for all experiments

Table 11: Parameters used for all experiments of the 3 state hidden Markov model.

Run	Cutoff point	CPI s=	UNR s=	GDP s=	index s=	itr_diff s=	# gm	x
1	4SD	1	N/A	1	1	1	1	0.1
2	3SD	1	N/A	1	0.5	1	1	0.1
3	3SD	1	N/A	1	0.25	1	1	0.1
4	3SD	1	N/A	1	0.80	1	1	0.1
5	3SD	1	N/A	1	1	1	1	0.1
6	3SD	1	N/A	0.8	1	1	1	0.1
7	3SD	1	N/A	0.75	1	1	1	0.1
8	2.5SD	1	N/A	0.75	1	1	1	0.1
9	3SD	N/A	N/A	0.75	N/A	N/A	1	0.1
10	3SD	N/A	N/A	0.75	1	N/A	1	0.1
11	3SD	1	N/A	0.75	1	N/A	1	0.1
12	4SD	1	1	0.8	1	1	1	0.1
13	4SD	1	1	0.8	0.75	1	1	0.1
14	4SD	0.75	1	0.8	0.75	0.75	1	0.1
15	4SD	0.5	1	0.8	0.5	0.5	1	0.1
16	4SD	0.5	1	0.8	1	0.5	1	0.1
16	N/A	0.5	1	0.8	1	0.5	1	0.1
17	N/A	0.5	N/A	0.8	1	0.5	1	0.1
18	4SD	0.5	N/A	0.8	1	0.5	1	0.1
19	4SD	0.5	N/A	0.8	1	0.5	1	0.1
20	3SD	0.5	N/A	0.8	0.05	0.05	1	0.1
21	3SD	0.05	N/A	0.5	0.05	0.05	1	0.1

22	3SD	0.8	N/A	0.8	1	1	1	0.1
23	3SD	0.8	N/A	0.8	0.9	1	1	0.1
24	3SD	0.8	N/A	0.8	1	0.9	1	0.1
25	3SD	0.8	N/A	0.7	1	1	1	0.1
26	4SD	0.8	N/A	0.8	1	1	1	0.1
27	3SD	0.8	N/A	0.8	1	1	2	0.1
28	3SD	0.8	N/A	0.8	1	1	1	0.05
29	4SD	0.75	1	0.8	0.6	0.75	1	0.1
30	3SD	0.75	1	0.8	0.6	0.75	1	0.1
31	3SD	0.75	0.9	0.8	0.75	0.75	1	0.1
32	3SD	0.75	1	0.8	0.75	0.75	2	0.1
33	3SD	0.75	1	0.8	0.75	0.75	2	0.1
34	3SD	0.75	1	0.8	0.75	0.75	2	0.05
35	2SD	1	N/A	1	0.5	1	1	0.1
36	3SD	1	N/A	0.9	0.5	1	1	0.1
37	3SD	1	N/A	1	0.4	1	1	0.1
38	3SD	1	N/A	1	0.65	1	1	0.1
39	3SD	1	N/A	1	0.5	0.1	1	0.1
40	3SD	1	N/A	1	0.5	1	2	0.1
41	3SD	1	N/A	1	0.5	1	3	0.1
42	3SD	1	N/A	1	0.5	1	1	0.05
43	N/A	0.5	N/A	0.8	1	0.5	1	0.1
44	4SD	0.5	N/A	0.8	0.9	0.5	1	0.1
45	4SD	0.5	N/A	0.8	0.8	0.5	1	0.1
46	4SD	0.5	N/A	0.8	0.7	0.5	1	0.1
47	4SD	0.5	N/A	0.7	0.9	0.5	1	0.1
48	4SD	0.5	N/A	0.9	0.9	0.5	1	0.1
49	4SD	0.5	N/A	0.8	0.9	0.6	1	0.1
50	4SD	0.5	N/A	0.8	0.9	0.7	1	0.1
51	4SD	0.6	N/A	0.8	0.9	0.6	1	0.1
52	4SD	0.4	N/A	0.8	0.9	0.6	1	0.1
53	4SD	0.3	N/A	0.8	0.9	0.6	1	0.1
54	4SD	0.2	N/A	0.8	0.9	0.6	1	0.1
55	4SD	0.1	N/A	0.8	0.9	0.6	1	0.1
56	4SD	0.2	N/A	0.8	0.9	0.6	2	0.1
57	4SD	0.2	N/A	0.8	0.9	0.6	1	0.05

7.5.3 Visual results of best experiments

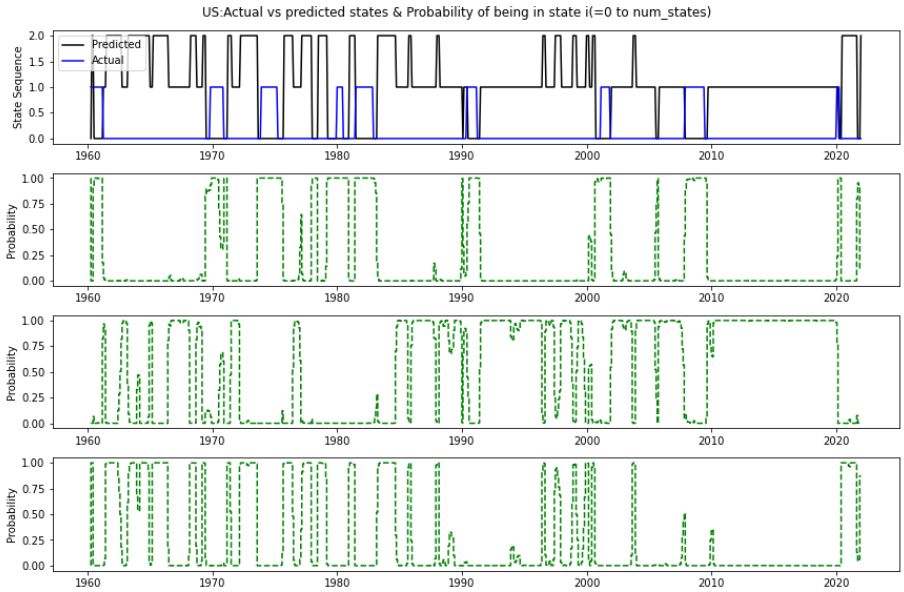

Figure 57: Prediction of the US versus historical recession data and probability of being in state i (= 0 to 2) according to the model. The blue line in the top graph shows when the historical recessions were. The black line shows when the model predicts a recession. The other graphs in Figure indicate the probability according to the model that a country is in a particular hidden state at that moment.

.

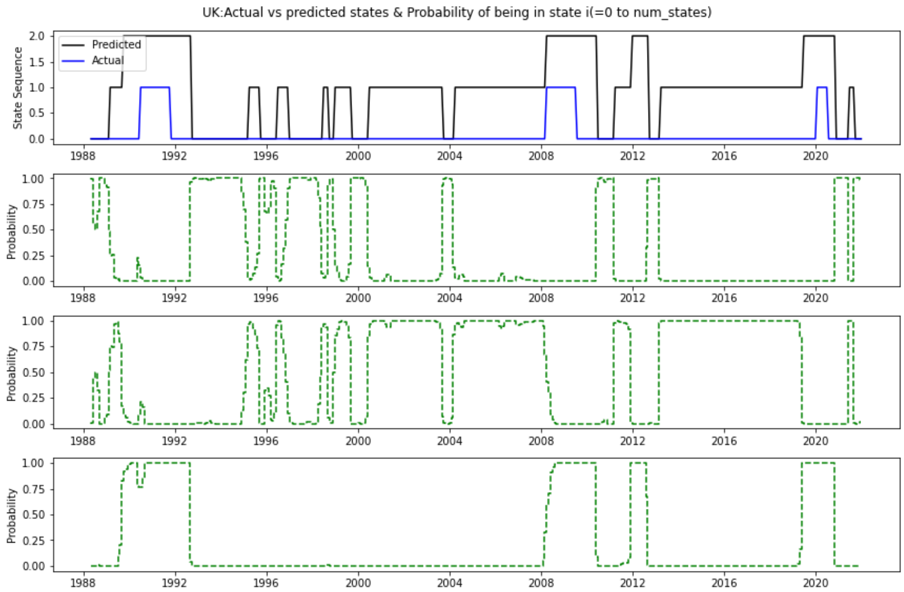

Figure 58: Prediction of the UK versus historical recession data and probability of being in state i (= 0 to 2) according to the model. The blue line in the top graph shows when the historical recessions were. The black line shows when the model predicts a recession. The other graphs in Figure indicate the probability according to the model that a country is in the particular hidden state at that moment.

.

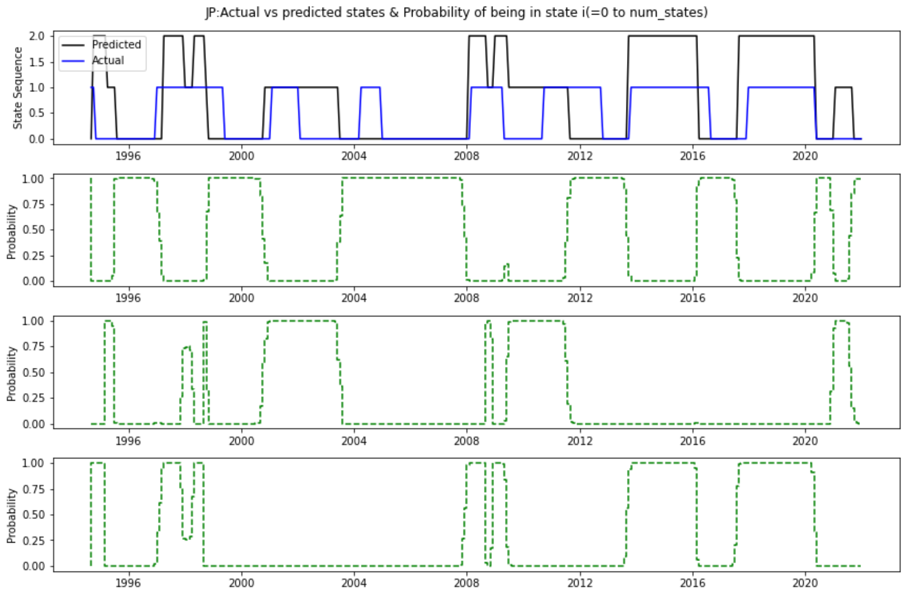

Figure 59: Prediction of the JP versus historical recession data and probability of being in state i (= 0 to 2) according to the model. The blue line in the top graph shows when the historical recessions were. The black line shows when the model predicts a recession. The other graphs in Figure indicate the probability according to the model that a country is in the particular hidden state at that moment.

.