# ***Page intentionally left blank***

# **Introduction**

There is a long tradition in business cycle analysis of separating periods in which there is broad economic growth, called expansions, from periods of broad economic contraction, called recessions. Understanding these phases and the transitions between them has been the focus of much macroeconomic research over the past century.

In the United States, the National Bureau of Economic Research (NBER) establishes a chronology of “turning point” dates at which the shifts between expansion and recession phases occur. The chronology identifies the dates of peaks and troughs that frame economic recessions and expansions. These dates are nearly universally used in work requiring a definition of U.S. business cycle phases.[[1]](#footnote-1)

This study compares two widely used recession forecasting models. The first approach is a parametric dynamic factor time series model that captures expansion and recession phases as unobserved regime shifts in the mean of the common factor. The unobserved state variable controlling the regime shifts is modeled as following a Markov process as in Hamilton (1989). This dynamic-factor Markov-switching model (DFMS), as developed in Chauvet (1998), produces a probability that the economy is in an expansion or a recession at any point in time. These probabilities can then be used to establish turning point dates using a rule for converting probabilities into a binary variable defining which regime the economy is in at any particular time.

The second approach is a non-parametric long short-term memory (LSTM) neural network, which is an artificial recurrent neural network (RNN) architecture used in the field of deep learning. It was first proposed by Sepp Hochreiter and Jürgen Schmidhuber (1997) to deal with the vanishing gradient problem[[2]](#footnote-2) that can occur when training a traditional RNN. RNN is a construct of connections between nodes that form a directed graph along a temporal sequence. This allows it to exhibit temporal dynamic behavior. For this, LSTM neural networks are well-suited to deal with time series prediction problems.

For the prediction of business cycle turning points, one set of coincident economic indicators and one set of coincident economic indicators are used, each of which is widely used as indicators that provide signals of the current state and the future state, respectively, in aggregate economic activity. Coincident economic indicators are designed to capture the common component of the business cycle as they rise and fall together, in a coincident fashion (Moore (1978)). Thus, they can provide information about whether the economy is currently experiencing a recession or a slowdown, recovery or a boom. Whereas leading indicators, for instance, should systematically provide a precise indication of the future state of the economy and send early economic signals in this regard.

For this study, four monthly coincident economic indicators, industrial production (IP), total personal income less transfer payments (PILTP), total manufacturing and trade sales (MTS), and employees on non-farm payroll (ENAP) are used as the first set of inputs to the introduced models. Those are the same economic variables utilized by the NBER and the Department of Commerce, which were initially selected based on Burn and Mitchell’s work (1946).

The set of leading economic indicators refer to yield curve parameters. The slope of the Treasury yield curve has often been cited as a leading economic indicator, with inversion of the curve being thought of as an indicator of recession (see e.g. Dueker (1997), Gurkaynak, Sack, Wright (2012), Johansson, Meldrum (2018)). Historically, the ten-year less three three-month term spread has exhibited a negative statistical relationship with real GDP growth over subsequent quarters, and a positive statistical relationship with recession probabilities (Estrella and Hardouvelis (1991) and Estrella and Mishkin (1996).

The paper is organized as follows: The subsequent section provides a review of related literature. The third section discusses the data. The fourth section introduces the two outlined forecasting models and the estimation procedure employed in this study. In the fourth section, the empirical results are presented and interpreted for both models. Furthermore, both models are tested for out-of-sample performance. The fifth section concludes.

# **Literature review**

There is large literature that investigates prediction of business cycle dates. The possibility of a set of indicators providing early signals of change in aggregate economic activity is important to any business or government affected by expansions and contractions. In their pioneering studies on this research topic, Burns and Mitchell (1946) provide a careful statistical description of the cyclical aspects of various time series, and classify macroeconomic variables as lagging, leading or coincident with economic activity (Chauvet, 1998). In their research, Burns and Mitchell (1946) identified two fundamental elements with respect to the business cycle, namely co-movements among economic variables through the business cycle as well as different behavior of the economy during different phases of the business cycle.

This framework has opened up two new research areas at that time regarding the use of statistical models. The first gave rise to the formation of dynamic factor models and composition of indices, as in Stock and Watson (1989), and the latter inspired the use of nonlinear regime switching models, as in Hamilton (1989). In this regard, Stock and Watson (1989) developed a multivariate dynamic factor model in which business cycles are measured by co-movements in various economic variables. Recessions and expansions are then generated by negative or positive symmetric shocks to a linear and dynamically stable time series. On the other hand, Hamilton (1989) developed a univariate Markov switching model which, in contrast to the dynamic factor model, considers nonlinearities in business cycles and assumes that business cycle expansions and recessions could be viewed as different regimes. The transition between recession and expansion is driven by an unobserved state variable that follows a Markov process.

Later on, Kim and Yoo (1995) extended Hamilton’s univariate Markov-switching model to a multivariate factor model, where a common factor series is driven by a two state Markov-switching process. Diebold and Rudebusch (1996) were the first to suggest a unified model that captures these two business cycle features from a set of economic indicators. They argued that co-movements among the individual economic indicators can be modelled by using the linear coincident indicator approach described in Stock and Watson (1993), while the existence of two separate business cycle regimes can be modelled by using the Markov-switching specification described by Hamilton (1989). Integrating these approaches, Kim and Yoo (1995), Chauvet (1998) and Kim and Nelson (1998) combined the dynamic factor and Markov-switching frameworks to implement a model that simultaneously captures both co-movements and regime shifts.

Neural networks have been successfully applied to economic problems in the last years and also has been widely used in forecasting business cycles with satisfactory results (see e.g. Shen, Du, Ji (2019) and Puglia, Tucker (2020)). The circumstance that traditional regression based single factor models as described above were not able to adequately model the asymmetries in business cycles gave initially rise to an implementation of neural networks due to their ability to recognize patterns (Kiani (2006)). As outlined in the literature, neural networks tend to outperform traditional forecasting methods in the short-term and for monthly and quarterly time series. Hill, O’Connor and Remus (1996) compared six traditional forecasting models to one neural network model and showed that the neural network model outperformed the traditional statistical methods. Another example of the tendency of neural networks to outperform traditional statistical approaches was given by Shen, Du and Ji (2019). In their paper, neural network models were used to classify recessions as indicated by the NBER’s business cycle dating methodology in the U.S. using 12 financial indices with a striking result. They achieved 98% accuracy in the training set, 97% accuracy in the validation set and 100% accuracy in the testing set. Given that the NBER announces business cycle turning points with a significant lag (Baris, 2018), those forecasting approaches as described above can help to detect contractions before NBER’s announcement.

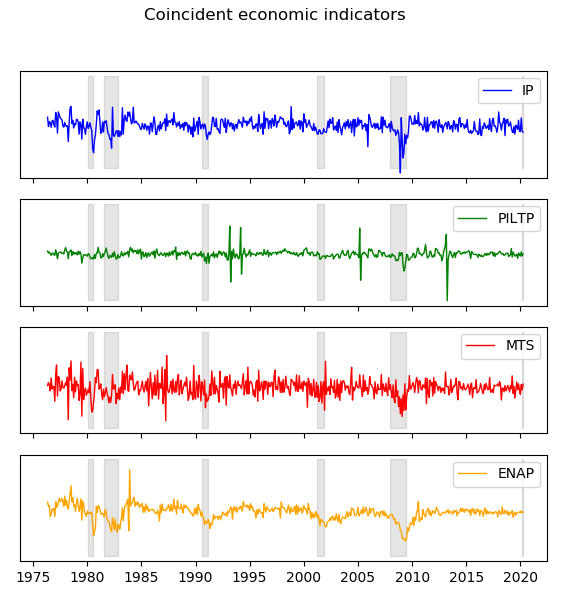
# **Data**

To test the forecasting performance of the dynamic-factor Markov-switching model and the LSTM neural network approach, each model was trained on a dataset of four coincident economic indicators and on a dataset of yield curve data to forecast recessions. In the following, those two selected data inputs are presented and possible differences in their application to the outlined models are described.

## **Coincident economic indicators**

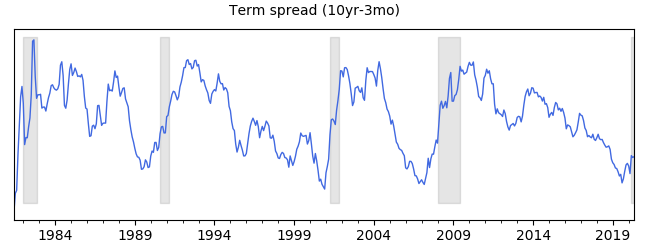
The first set of economic indicators, consisting of four monthly coincident variables ((1) non-farm payroll employment (ENAP), (2) industrial production (IP), (3) real manufacturing and trade sales (MTS), and (4) real personal income excluding transfer payments (PILTP)), was obtained from the public database of the Federal Reserve Bank of St. Louis (FRED). The monthly data is dated from 1967 until 2020 in order to provide an adequate time horizon to encompass as many recessions as possible for training and testing. All variables are seasonally adjusted. In order to achieve stationarity, the data was transformed by computing the first difference of natural logarithms, which is consistent to the approach taken by Chauvet (1998). The transformed data was standardized in the following step.

In the figure down below the four differenced and standardized monthly coincident economic indicators are plotted with the NBER recession dates shaded in grey. From a first visual analysis, the coincident fashion of those variables can be clearly identified. It is interesting to note, that there are differences in the sensitivity of changes among the economic variables. E.g. the industrial production (IP) variable and the real manufacturing and trade sales (MTS) variable seem to be most sensitive to business cycles, whereas the real personal income excluding transfer payments (PILTP) variable seems to be less sensitive, at least regarding its magnitude of changes within business cycles.



**Figure 1:** Differenced coincident economic indicators

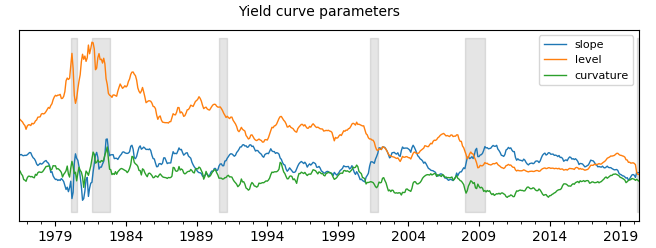
## **Yield curve data**

The second set of economic indicators used in this study refers to yield curve data which also was obtained from the public database of the FRED. For the LSTM neural network application, the slope of the yield, defined as the difference between the 10-year and the 3-month

**Figure 2:** Term spread between 1982 until 2020

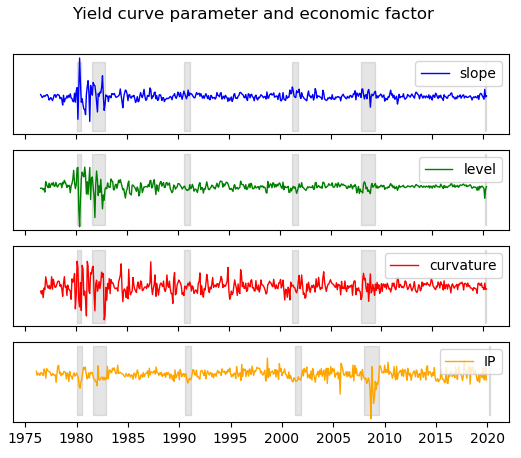
treasury rates, was used. In the figure above, the slope of the yield curve (or term spread) is plotted with the recession dates shaded in grey. For our analysis, the term spread was further processed by calculating the first difference of natural logarithms in order to obtain adequate stationarity. Lastly, the differenced data was scaled.

For the application with the dynamic-factor Markov-switching model, however, another approach was taken. As proposed in Chauvet and Senyuz (2012), not only the term spread but also the information extracted from other components of the yield curve and from real economic activity were taken into account. The reason is that this approach offers the possibility to take the interrelationship between the yield curve and economic activity through the dynamic factors and the Markov processes into account. (Chauvet and Senyuz (2012)). In particular, besides the slope, the other components of the yield curve refer to the level and the curvature.



**Figure 3:** Yield curve parameter from 1976 until 2020.

The described components of the yield curve are plotted in the figure above with the shaded regions referring to the recession dates indicated by the NBER. The level of the yield curve is highly persistent and considered to capture the long-run development of the yield curve (Chauvet and Senyuz (2012)). The curvature and the slope are considered to capture medium and short run developments, respectively. According to Diebold and Li (2006) the level of the yield is associated with inflation expectations, whereas the slope of the yield curve is related to economic activity. Lastly, the curvature of the yield curve indicates some correlation with the recession dates, although not as strong as the slope of the yield curve.

The monthly industrial production variable is considered as a proxy for the economic activity, as this is the most commonly used measure of economic activity. Those factors are then combined as a joint factor and estimated simultaneously from the observable variables and from their interrelationship. As a preprocessing step, the observable variables are differenced and scaled afterwards. In the figure below the differenced and scaled input factors to the dynamic-factor Markov-switching model are illustrated with the shaded regions indicating the recession periods as dated by the NBER. From a first visual analysis, the magnitude of changes in the yield curve components seem to have decreased after the mid 1980s. This observation is consistent with the observation of other studies as they have shown that the predictive content of the yield curve is not stable over time and that the forecasting ability of the yield curve dropped since the mid-1980s (see e.g. Giacomini and Rossi (2006)).

**Figure 4:** Illustration of difference yield curve parameter and industrial production from 1976 until 2020.

# **Methodology**

## **Markov Switching Models**

The dynamic-factor Markov-switching model incorporates both co-movements and business-cycle shifts into a statistical model (see e.g. Kim and Yoo (1995), Chauvet (1998) and Kim and Nelson (1998)). The model proposes that a vector of macroeconomic variables, which are hypothesized to move coincidentally with the overall economic conditions can be decomposed as the sum of two components. The first component refers to an unobserved time series variable, which corresponds to the common factor among the observable variables. The second component is a time series vector, which represents the idiosyncratic movements in the time series (Chauvet (1998)). This suggests the formulation:

**Formula 1:** Dynamic-factor Markov-switching model

The vector represents the first difference of the log of the endogenous observable indicator. The parameter relate to the factor loadings of the observable macroeconomic variables, which measure the sensitivity to changes in the business cycle andis the growth rate of the common factor. Finally, the last component of the first formula represents the idiosyncratic movement which can be further decomposed into the intercept of each observable indicator and the error term which is modelled as a process with autoregressive representation.

Thus, each observable indicator consists of an individual idiosyncratic component and a linear combination of current and lagged values of the common factor delta (Kim and Yoo (1995)). The common factor is assumed to be generated by an autoregressive process with a long-run growth and Markov switching deviations from the long-run growth. The first part of the second formula depends on whether the economy is in a recession () or in expansion (). Transitions between regimes or states of the business cycle are then assumed to follow a Markov process. Lastly, the parameter represents common shocks to the change in the common factor. Through this common shocks regime shifts are generated.

With regard to the topic of this study the model can be interpreted as follows: refers to the long-run growth rate of the unobserved common factor . The parameter produces deviations from that long-run growth rate according to whether the economy is in a recession, which corresponds to , or in an expansion (). Thus, business cycles are generated from the co-movements among the observable economic variables and from common shocks to the dynamic factor.

By putting the equations into state space form, the parameters of the dynamic factor and the Markov regime switching probabilities can be simultaneously estimated by maximizing its likelihood function. In order to estimate the model it is necessary to consider both the unobserved nonlinear common factor of the coincident economic variables and the latent Markov state. Since the dynamic factor model can be estimated by applying the Kalman filter, the estimation procedure for the dynamic-factor Markov-switching model consists of a combination of Hamilton’s filter and a nonlinear discrete version of the Kalman filter, as developed by Kim (1994).

## **LSTM Neural Network**

In a basic neural network, information is mapped from inputs to a target output through a sequence of simple data transformations (layers). Those layers represent operations on the data as it moves from input to output. The basic architecture upon which deep learning models are built is through three distinct set of layer: an input layer, a hidden layer of computational nodes, and a final layer representing model outputs (Chollet (2018)).

The neural network model does not, as in the case of Markov-switching models, employ maximum likelihood estimation (MLE) for estimation. Instead, neural networks rely on the backpropagation training algorithm, which is essentially non-parametric. Through the training process, the model will identify which features and parameters (i.e. computational nodes) are relevant for prediction. This enables the circumstance to be less selective about what data to supply to a model and less concerned with how to pre-process the data (Cook (2017)).

For the study, a neural network was selected which is built on a recurrent network architecture. The recurrent neural network (RNN) draws information from the temporal structure of the input data (Hochreiter, Schmidhuber (1997)). In fact, these types of networks take the sequence of data into account in which the input data is presented to the model. Recurrent neural networks do this by accepting input only from the current input in a sequence but from the state of the network that arose when considering previous inputs in that sequence.

As there are many variants of RNN architectures, this study focused on a long short-term memory (LSTM) network. This type of RNN architecture is used as it solves the vanishing gradient problem from which RNN architectures and other neural network architectures with gradient-based learning methods and backpropagation suffer (Hochreiter, Schmidhuber (1997)). In those models, each of the neural network’s weights receive updated information proportional to the partial derivative of the error function with respect to the current weight in each iteration of training . In case, the gradient becomes very small, this circumstance might stop the neural network from further training. This problem is not that apparent when using LSTM neural networks due to their structure (Cook (2017)).

The basic architecture of a LSTM neural network model is composed of a cell, which refers to the memory part of the LSTM unit and three regulators, which are referred to as gates. The first gate is an input gate. This gate controls the extent to which new information flow into the cell. The second is the so called forget gate. As the name suggests, the forget gate controls the extent to which information remain in the cell. The last gate is the output gate, which control the extent to which the information in the cell is used to compute the output activation of the LSTM unit. Finally, there are connections into and out of the LSTM gates, of which a few are recurrent. The weights assigned to these connections eventually determine how the LSTM gates operate (Hochreiter, Schmidhuber (1997)).

# **Results**

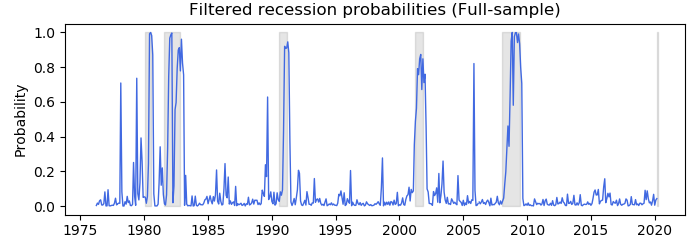
## **Evaluation of dynamic-factor Markov-switching models**

In the following the performance of the dynamic-factor Markov-switching model applied to coincident economic indicators and yield curve data is evaluated. As a first step, the parameters of the dynamic-factor Markov-switching models are estimated using the Kim filter. For the estimation process, standardized starting parameters as proposed by Kim and Yoo (1995) are used as a proxy in order to decrease the time of iteration.

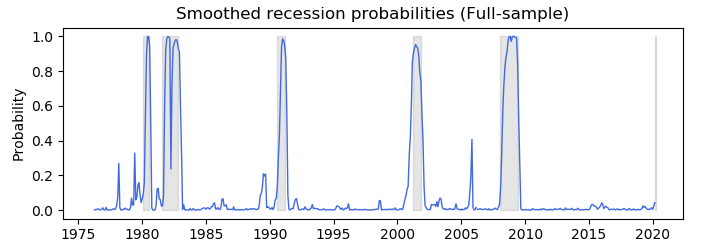
The parameter estimates obtained through numerical maximization of the maximum likelihood function are portrayed in Tables 1 and 2 in the appendix, for the leading economic indicators and the yield curve parameters, respectively. As for the approach with leading indicators, the estimated parameters provide an overall support to the dynamic-factor Markov-switching model. The asymmetries in the length of the business cycle regimes (expansion and recession) are well characterized through the estimated transition probabilities p and q. That is, the probability of staying in expansion (p) is higher than the probability of staying in contraction (q). This duration dependence based on the state of the business cycle was widely investigated in related literature (see e.g. de Bondt, Vermeulen, 2018) and also confirmed by this study. The expected duration for recession and expansion implied by the transition probabilities is 29.41 and 7.18 months, respectively. With respect to the factor loadings 𝛾 , industrial production (‘IP’) has the highest coefficient, indicating that this specific indicator is the most sensitive to changes in business cycles. This observation was also confirmed by other studies (see e.g. Chauvet (1989)).

As for the approach with yield curve data, the estimated parameters also provide support to the underlying model framework. As in the case of leading indicators, the transition probabilities calculated using yield curve data confirm the asymmetry in the length of business cycles. The expected duration for a recession implied by the probabilities amounts to 26.04 months and for an expansion to 6.99 months. These expected durations estimates are fairly similar to the expected duration as estimated by using coincident economic indicators.

Furthermore, the factor loading of the slope parameter of the yield curve is negative. As the slope of the yield is supposed to be negatively correlated to business cycle movements, this result seem to be appropriate.[[3]](#footnote-3) Furthermore, the level of the yield curve corresponds to the highest coefficient. However, it is surprising that the factor loading of the economic variable industrial production is relatively small. One possible explanation for this observation might be that the interrelationship between the leading yield curve parameters and the lagging economic variable was relatively imbalanced.

In the second step, the estimated model parameters were used to determine the smoothed and filtered recession probabilities as inferred by the model. Those inferences of the recession probabilities can be used to identify peaks and troughs as determined by the NBER. The figures 5 and 6 plot the estimated smoothed and filtered probability that the economy is in a recession for the full-sample of coincident economic indicators. The blue shaded graph refers to the estimated recession probability for full-sample data that the economy is in the recession

**Figure 5:** Filtered recession probabilities conditional on the full-sample with coincident indicators as underlying input data.



**Figure 6:** Smoothed recession probabilities conditional on the full-sample with coincident indicators as underlying input data.

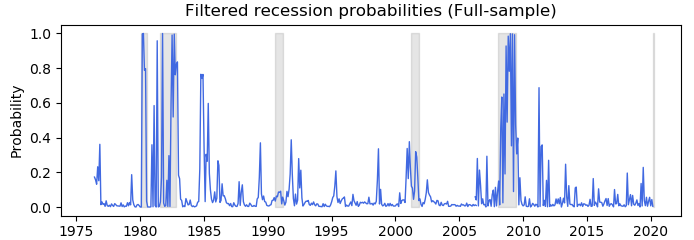
state at time t, based on information up to time t.[[4]](#footnote-4) Both plots show that the probabilities of the recession states are similar to the recessions as dated by the NBER (grey shaded area). For an interpretation of the inferred recession probabilities two factors are important. First, the ability to capture the start date of a recession and , second, the ability to capture the duration of the recession. Within this framework, all recessions conditional on the full-sample are well characterized by the smoothed and filtered recession probabilities. At around the time where the NBER recessions start, the model inferred recession probabilities rise substantially and remain high until around the end of the recession period.

In order to assess the performance of the inferred probability in predicting recession dates as implied by the NBER, the in-sample and out-of-sample forecasts were evaluated with respect to their accuracy and precision. For this purpose, Hamilton’s (1989) approach to characterize peaks and troughs was used in order to create a dataset of binary classifiers. That is, following Hamilton’s method, a ‘1’ was assigned to a date variable if the corresponding smoothed probability of a recession is greater than 0.5 and ‘0’ otherwise. The table down below summarizes the results for this evaluation of binary classifiers. In this approach four classification outcomes are possible, ‘true positive (TP)’, ‘true negative (TN)’, ‘false positive (FP)’, and ‘false negative (FN)’. The accuracy rate (AR) measures the fraction of all instances that are correctly categorized in relation to the total number of correct and incorrect classifications.[[5]](#footnote-5) The precision rate (PR) relates to the ratio of the number of true positive (true negative) classifications to the total number of positive (negative) events.[[6]](#footnote-6) Finally, the root-mean-square error (RMSE) measures the differences between the values predicted by the model and the values observed.

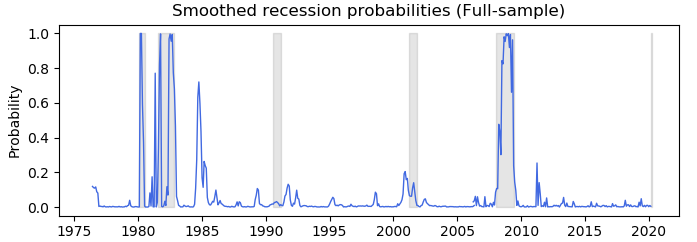
|  |  |  |
| --- | --- | --- |
| **Turning Point** | **Coincident indicators** | |
| **Evaluation** | **In-sample** | **Out-of-sample** |
|  |  |  |
| *True Positive* | *27* | *14* |
| *False Positive* | *11* | *2* |
| *True Negative* | *311* | *146* |
| *False Negative* | *11* | *6* |
| ***Total periods*** | ***360*** | ***168*** |
| *Accuracy Rate* | *94%* | *95%* |
| *Precision Rate* | *71%* | *88%* |
| *RMSE* | *6%* | *5%* |

**Table 1:** Performance evaluation of dynamic-factor Markov-switching model using coincident indicators as underlying input data

The idea is to compare and evaluate not only the model performance on ex-post in-sample data but also on out-of-sample data in order to assess the quality of the model. As shown in the table 1, the calculated accuracy and precision rates provide support to the dynamic-factor Markov-switching model in the ability to forecast recessions. The out-of-sample performance is slightly better in terms of the accuracy rate and the precision rate than the in-sample results.

With regard to the application of the dynamic-factor Markov-switching model with yield curve parameters and one proxy for economic activity as the underlying input data , the estimated smoothed and filtered recession probabilities conditional on the full-sample are plotted down below.[[7]](#footnote-7) The inferred recession states also capture fairly well the NBER business cycle dating, except for the recession in 1990 and 2001.

**Figure 7:** Filtered recession probabilities conditional on the full-sample with yield curve parameters and one proxy for economic activity as underlying input data.



**Figure 8:** Smoothed recession probabilities conditional on the full-sample with yield curve parameters and one proxy for economic activity as underlying input data.

As already described in previous sections, the predictive content of the yield is not stable over time and that the forecasting ability of the yield curve decreased since the mid 1980s. In fact, it was observed by several other studies that the slope of the yield did not turn negative before the 1990 recession. In addition, researchers were also not able to predict the 2008/2009 recession adequately. However, the intention of the dynamic-factor Markov-switching methodology was to overcome those shortfalls by including the possibility to capture abrupt changes in the underlying input series. Those abrupt changes were well captured in the 2008/2009 recession, whereas the model failed to capture the 1990 and 2001 recession dates.

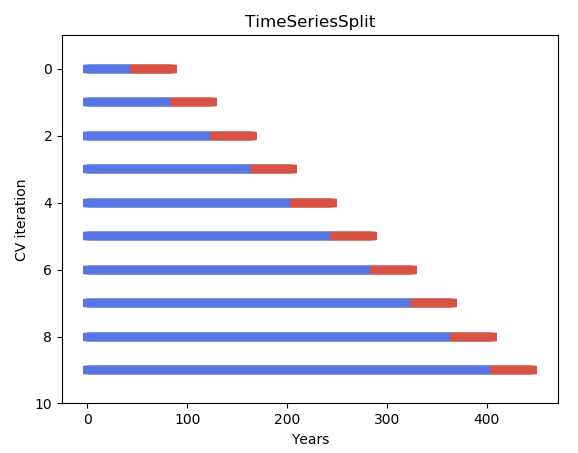
Those observations can be confirmed when inspecting the turning point evaluation in the table 2 down below. The in-sample performance of the model is less satisfactory than the out-of-sample performance. The inability to capture the 1990 and 2001 recession correctly results in a precision rate of 73% and an accuracy rate of 91% with respect to the in-sample performance. The out-of-sample performance is determined by a striking forecast of recession dates, with a precision rate of 100% and an accuracy rate of 95%.

|  |  |  |
| --- | --- | --- |
| **Turning Point** | **Yield curve** | |
| **Evaluation** | **In-sample** | **Out-of-sample** |
|  |  |  |
| *True Positive* | *11* | *11* |
| *False Positive* | *4* | *0* |
| *True Negative* | *318* | *148* |
| *False Negative* | *27* | *9* |
| ***Total periods*** | ***360*** | ***168*** |
| *Accuracy Rate* | *91%* | *95%* |
| *Precision Rate* | *73%* | *100%* |
| *RMSE* | *9%* | *5%* |
|  |  |  |

**Table 2:** Performance evaluation of dynamic-factor Markov-switching model using yield curve parameters and one proxy for economic activity as underlying input data

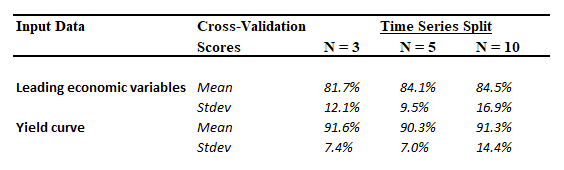
All in all, the outlined results provide an overall support to the dynamic-factor model with Markov-switching for both approaches, using coincident economic variables, and yield curve parameters and one proxy for economic activity as underlying input data.

## **Evaluation of LSTM Neural Network models**

Before assessing the quality of the LSTM neural network model, a cross validation run is executed to derive a more accurate estimate of the model’s prediction performance. Due to the temporal components inherent to time series data, traditional cross validation techniques such as k-fold cross validation cannot be directly used as they assume no direct relationship between the observations, i.e. that each observation is independent. Thus, for the LSTM neural network models applied in this study, a cross validation technique called “time series split” is used in order to respect the temporal order in which the values are observed. This approach will result in a more accurate estimate of the performance of the LSTM models on unseen data. For both datasets of this study 3 different cross validation runs were executed, with n=3, n=5 and n=10 splits, in order to assess the robustness of the model.[[8]](#footnote-8) In the table below, the methodology of the time series split cross validation technique is visualized (with the example of n=10 splits).

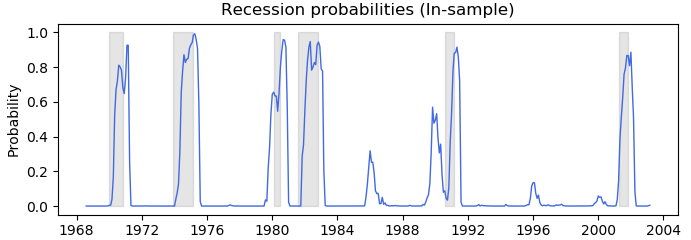
**Figure 9:** Cross validation technique “time series split”

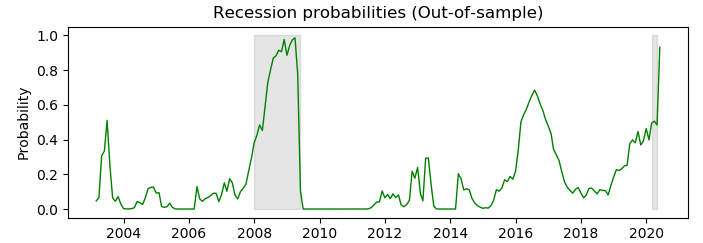
The table below presents the results of time series split cross validation across both underlying datasets used in this study. For this analysis, the metric accuracy was chosen to score each approach. The table shows the mean accuracy estimates and standard errors for the LSTM neural network model with coincident economic indicators and yield curve data as the underlying input data.



**Table 3:** Cross validation results using “time series split” for n=3, n=5 and n=10 runs

For both input data, the cross validation lead to satisfactory results using LSTM neural networks with a simple two hidden-layer architecture. In order to prevent overfitting, a LSTM model with only two dense layers with 16 nodes each was implemented. For the those hidden layers, the rectified linear (“relu”) activation function was used. The “relu” function will directly forward positive inputs from the first layer to the second layer and output a zero for every negative input. For the output layer a sigmoid function was used, which generates a result between zero and one, which is adequate for a binary classification problem. The loss function implemented in the model refers to “binary-crossentropy” as indicated by the binary target variable to determine recessions (1) and expansions (0). Finally, in order to assess the quality of the LSTM neural networks in forecasting recessions, three different time horizons (1 month, 3 months and 12 months) were predicted by using the last 12 months of observation in case of the application with coincident economic indicators and 24 months of observations in case of the application with the yield curve. The duration of those timesteps was chosen based on the respective nature of the coincident and leading variables. The quality of those predictions is, in a next step, evaluated using accuracy rates and precision rates.

To begin with, the performance of the LSTM neural network using coincident economic variables as input data is evaluated. The figure below plots the estimated recession probabilities for the full-sample and the out-of-sample using the last 12 months as observed data points in order to predict a 1 month forecast horizon.[[9]](#footnote-9)



**Figure 10:** Recession probabilities conditional on the full-sample with coincident economic variables as underlying input data.

**Figure 11:** Recession probabilities conditional on the out-of-sample with coincident economic variables as underlying input data.

The visual result is satisfactory as all historic recessions are captured fairly well regarding start date as well as duration of recessions.

The results of the out-of-sample exercise is illustrated in the table down below. The probability forecasts of the 1 month, 3 months and 12 months forecast horizons are evaluated with respect to their accuracy and precision. The accuracy rates are fairly high for all forecast horizons. However, since it is more likely to capture true negative events than true positive events, that is, due to the asymmetries in the duration of business cycles, the precision rate is the more accurate metric to assess the quality of the model’s performance. For a 1 month forecast horizon the precision rate is 88%. As the forecast horizon increases the precision rate decreases.

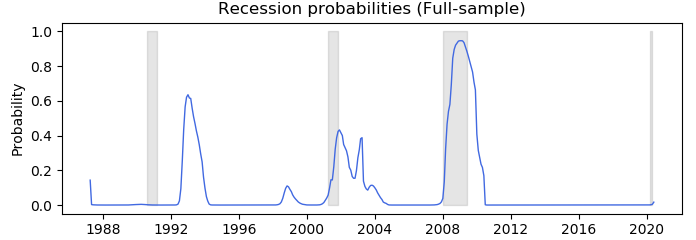
|  |  |  |  |
| --- | --- | --- | --- |
| **Turning Point** | **Coincident indicators** | | |
| **Evaluation** | **1 month** | **3 months** | **12 months** |
|  |  |  |  |
| *True Positive* | *15* | *8* | *0* |
| *False Positive* | *2* | *9* | *0* |
| *True Negative* | *186* | *177* | *183* |
| *False Negative* | *5* | *13* | *21* |
| ***Total periods*** | ***208*** | ***207*** | ***204*** |
| *Accuracy Rate* | *97%* | *89%* | *90%* |
| *Precision Rate* | *88%* | *47%* | *n/a* |
| *RMSE* | *3%* | *11%* | *10%* |

**Table 4:** Performance evaluation of the LSTM neural network model using coincident economic variables as underlying input data for 1 month, 3 months and 12 months out-of-sample exercise

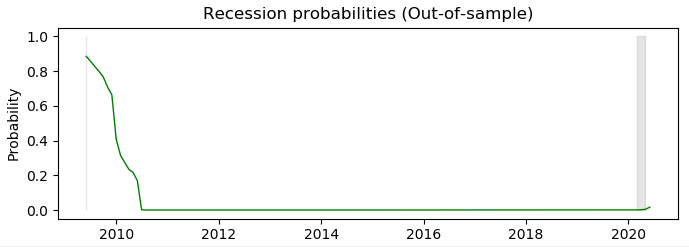
data.

For a 12 month forecast horizon the model was not able to capture any recession dates. This observation confirms the expectation that coincident indicators do not anticipate movements in the economy very well but rather measure current economic performance and indicate whether the economy is currently experiencing a recession or a slowdown, a recovery or a boom. In this regard, they can be used to identify and date the peaks and troughs in the business cycle at a point in time. However, this result give support to the LSTM neural network model at least for short-term predictions.

In the following, the out-of-sample performance of the LSTM neural network using the slope of the yield curve as input data for a 1 month, 3 months and 12 months forecast horizon is evaluated.[[10]](#footnote-10) For this, the length of the sequence within the LSTM model was set to 24 months. From a first visual analysis of figures 12 and 13, the LSTM model performs well in forecasting the 2008/2009 recession. However, the 2001 recession is not captured very well and the 1990 is not predicted at all. In addition, the model output falsely signals a peak and tough in the years 1992/1993. If this model were to be used to predict recession at that time, the result would provide large uncertainty as to whether and when the economy would be heading to a recession within this time frame, since prior to this period the term spread predicted recessions with an average of two years.



**Figure 12:** Recession probabilities conditional on the full-sample with the slope of the yield curve as underlying input data.



**Figure 13:** Recession probabilities conditional on the out-of-sample with the slope of the yield curve as underlying input data.

In applying Hamilton’s (1989) approach to characterize peaks and troughs, the result as shown in table 5 is satisfactory in terms of accuracy rate. The model’s forecast for the 2008/2009 recession is perfect in terms of the predicted duration and the predicted start of the recession. However, in terms of the precision rate only the 3 month forecast horizon is very good.

|  |  |  |  |
| --- | --- | --- | --- |
| **Turning Point** | **Yield curve data** | | |
| **Evaluation** | **1 month** | **3 months** | **12 months** |
|  |  |  |  |
| *True Positive* | *0* | *2* | *2* |
| *False Positive* | *0* | *0* | *10* |
| *True Negative* | *122* | *122* | *118* |
| *False Negative* | *12* | *9* | *2* |
| ***Total periods*** | ***134*** | ***133*** | ***132*** |
| *Accuracy Rate* | *91%* | *93%* | *91%* |
| *Precision Rate* | *n/a* | *100%* | *17%* |
| *RMSE* | *9%* | *7%* | *9%* |

**Table 5:** Performance evaluation of the LSTM neural network model using the slope of the yield curve as underlying input data for 1 month, 3 months and 12 months out-of-sample exercise

data.

All in all, the outlined results provide an overall support to the LSTM model, but especially for approach using coincident economic variables.

# **Conclusion**

The results of this study provide overall support to the recession forecasting ability of both, the dynamic-factor Markov-switching model and the LSTM neural network model.

With regard to the out-of-sample performance using coincident economic variables, the results suggest that the LSTM neural network model tends to slightly outperform the dynamic-factor Markov-switching model in terms of overall accuracy and in the error metric in the short-term, whereas in the long-term the latter tends to better capture abrupt changes in the underlying input data series. The accuracy rate can be interpreted in this scenario as a measure of how well the start of a recession is predicted, i.e. it only provides information about the state of economy at specific point in time, but it does not provide an indication of the duration of the predicted recession. In this regard, the neural network seems to be slightly more sensitive with respect to changes in economic activity in the short-term than the underlying regime shift methodology of the dynamic-factor Markov-switching model. In terms of the precision rate, the dynamic-factor Markov model clearly outperforms the neural network. An explanation for this result is the tendency of the Markov model to better capture the discrete shift in the underlying economic factors, which display the asymmetric nature of business cycle phases.

With respect to the out-of-sample performance using yield curve data, the overall result suggests that the dynamic-factor Markov-switching model slightly outperforms the LSTM neural network approach in terms of accuracy and clearly outperforms the LSTM model in terms of precision, at least for the 1 month and the 12 month forecast horizon.

For the purpose of comparison it has to be noted at this point that the dynamic-factor Markov-switching approach was applied using multivariate information extracted from three yield curve components and one proxy for economic activity. For the neural network approach, only the slope of the yield curve was considered. The purpose of this decision to only include univariate information into the neural network was not only to compare the performance of both models in terms of accuracy and precision, but also to investigate both models in terms of flexibility and complexity. As indicated by the results using yield curve data, both model are somewhat well-suited to forecast recessions in a reasonable time horizon of e.g. 3 months. However, whereas the dynamic-factor Markov-switching model needs to introduce additional parameters in order to achieve the results, with the LSTM neural network approach only one variable was considered without the necessity of complex model specifications.

For future investigations, it might be worthwhile to apply an univariate version of Markov switching models, e.g. by applying Hamilton’s (1989) method, to test the performance of the slope of the yield as a predictor of future business cycles within this framework. Also, a hybrid version of both models analyzed in this study, .i.e. a LSTM neural network model that considers a dynamic factor of economic variables as input data, might be worthwhile to investigate.

# **References**

Burns, Arthur F., and Mitchell, Wesley C. (1946), “Measuring Business Cycles.” New York: National Bureau of Economic Research

Chauvet, Marcelle. (1998), “An Economic Characterization of Business Cycle Dynamics with Factor Structure and Regime Switching.” *International Economic Review* 39, no.4: 969-996.

Chauvet, Marcelle., and Senyuz, Zeynep. (2012), “A Dynamic Factor Model of the Yield Curve as a Predictor of the Economy.” Finance and Economics Discussion Series 2012-32, Board of Governors of the Federal Reserve System (U.S.).

Chollet, Francois. (2018), “Deep learning with Python.” Manning Publications Co.

Cook, Thomas R., and Hall, Aaron S. (2017), “Macroeconomic Indicator Forecasting with Deep Neural Networks.” kcFED Research Working Papers 17-11.

De Bondt, Gabe., and Vermeulen, Philip. (2018), “Business cycle duration dependence and foreign recessions.” ECB Working Paper No. 2205.

Diebold, Francis., and Rudebusch, Glenn D. (1996), “Measuring Business Cycles: A Modern Perspective.” *The Review of Economics and Statistics* 78, no.1: 67-77.

Diebold, Francis X., and Li, Canlin. (2006), “Forecasting the term structure of government

bond yields.” *Journal of Econometrics* 130: 337-364.

Dueker, Michael J. (1997), “Strengthening the Case for the Yield Curve as a Predictor of U.S. Recessions.” *Review*, March/April 1997: 41-51.

Estrella, Arturo., and Mishkin, Frederic S. (1996), “How Stable is the Forecasting Performance of the Yield Curve for Output Growth?.” NBER Working Paper 5379.

Giacomini, Raffaella., and Rossi, Barbara. (2006), “Strengthening the Case for the Yield Curve as a Predictor of U.S. Recessions.” *Oxford Bulletin of Economics and Statistics* 68,no.1: 783-795.

Gurkaynak, Refet S., Sack, Brian., and Wright, Jonathan H. (2012), “The U.S. Treasury Yield Curve: 1961 to the Present.” Finance and Economics Discussion Series 2006-28, Divisions of Research & Statistics and Monetary Affairs, Federal Reserve Board.

Hamilton, James D. (1989), “A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle.” *Econometrica* 57, no. 2: 357-384.

Hill, Tim., O’Connor, Marucs., and Remus, William. (1996), “Neural Network Models for Time Series Forecasts.” *Management Science* 42, no.7: 1082-1092.

Hochreiter, Sepp., and Schmidhuber, Jürgen. (1997), “Long Short-Term Memory.” *Neural Computation*, no.9(8): 1735-1780

Kiani, Khurshid. (2008), “On Forecasting Recessions via Neural Nets.” *Economics Bulletin* 3, no. 13: 1-15.

Kim, Chang-Jin. (1994), “Dynamic linear models with Markov-switching.” *Journal of Econometrics* 60, no. 1-2: 1-22.

Kim, Myung-Jig., Yoo, Ji-Sung. (1995), “New index of coincident indicators: A multivariate Markov switching factor model approach.” *Journal of Monetary Economics* 36, no. 3: 607-630.

Kim, Chang-Jin., and Nelson, Charles R. (1998), “Has the U.S. Economy Become More Stable? A Bayesian Approach Based on a Markov-Switching Model of Business Cycle.” *The Review of Economics and Statistics* 81, no.4: 608-616.

Moore, Geoffrey H. (1983), “Why the Leading Indicators Really Lead.” Business Cycles, Inflation, and Forecasting, 2nd edition.

Peter, Johansson., and Meldrum, Andrew. (2018), “Predicting Recession Probabilities Using the Slope of the Yield Curve.” FEDS Notes, Washington: Board of Governors of the Federal Reserve System.

Puglia, Michael., and Tucker, Adam. (2020), “Machine Learning, the Treasury Yield Curve and Recession Forecasting.” Finance and Economics Discussion Series 2020-038, Board of Governors of the Federal Reserve System.

Shen, Lujia., Du, Tianyu., and Ji, Shouling. (2019), “Neural Network Model for Classifying the Economic Recession and Construction of Financial Stress Index.” Data Science - 5th International Conference of Pioneering Computer Scientists, Engineers and Educators.

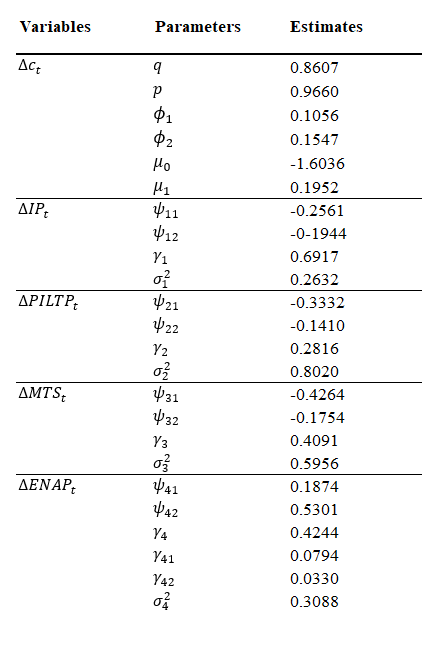
Soybilgen, Baris. (2018), “Identifying US business cycle regimes using dynamic factors and neural network models.” MPRA Paper no. 94715.

Stock, James H., and Watson M. (1993), “A Procedure for Predicting Recessions with Leading Indicators: Econometric Issues and Recent Experience.” NBER Working Paper 4014.

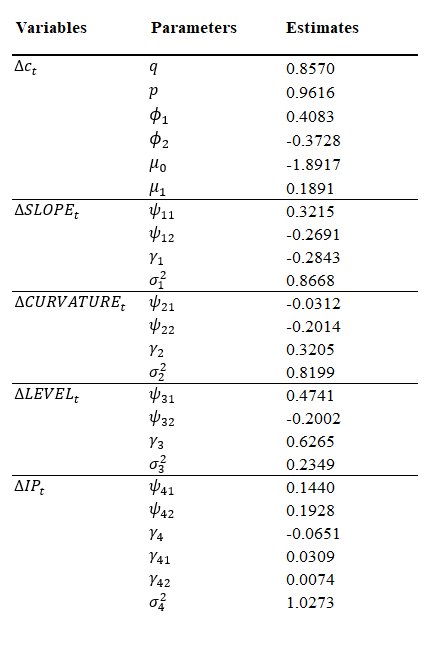
Stock, James H., and Watson M. (1989), “New Indexes of Coincident and Leading Economic Indicators.” *NBER Macroeconomics Annual*, no.4: 351-409.

# **Appendix**

**Table 6:** Maximum Likelihood Estimates of the Dynamic Factor Model with Markov switching of coincident indicators

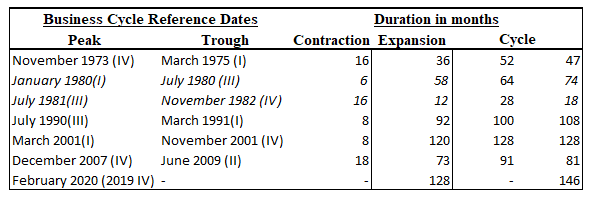


**Table 7:** Maximum Likelihood Estimates of the Dynamic Factor Model with Markov switching of the Yield Curve and the Economy

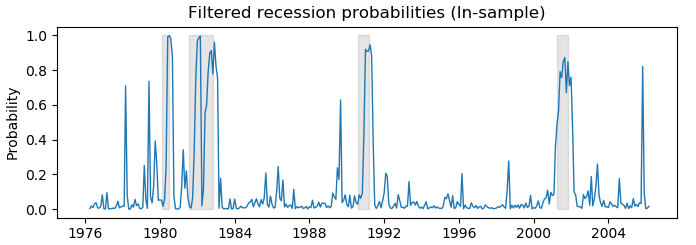


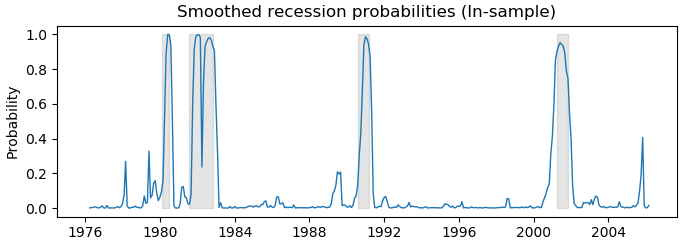
**Table 8:** U.S. Business Cycle Expansions and Contractions as defined by the NBER

In the “Contraction” column the months from Peak to Trough and in the “Expansion” column the months from previous Trough to this Peak are displayed. The left side of the “Cycle” column determines the duration from one Trough to the previous Trough and the right side the duration from one Peak to the previous Peak.

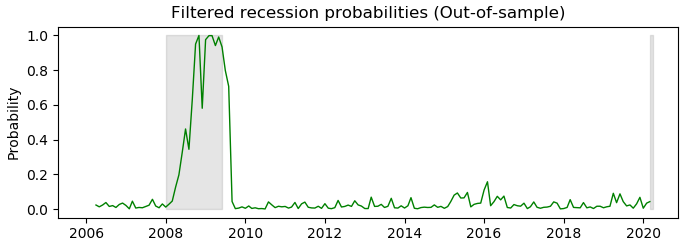


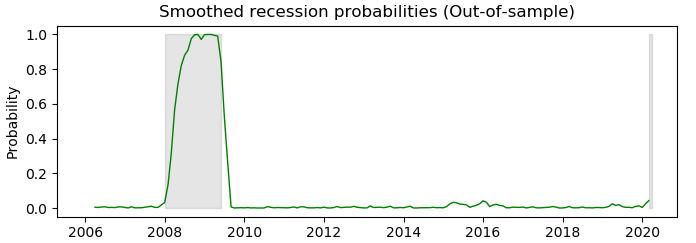
**Figure 14:** In-sample Filtered and Smoothed Probabilities of Recessions from Dynamic Factor Markov Switching Model based on coincident economic indicators



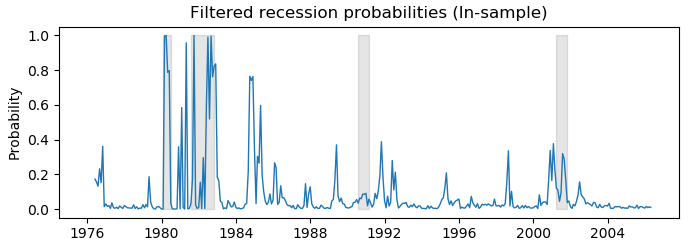


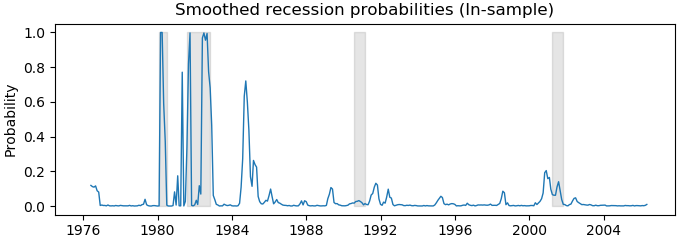
**Figure 15:** Out-of-sample Smoothed and Filtered Probabilities of Recessions from Dynamic Factor Markov Switching Model based on coincident indicators



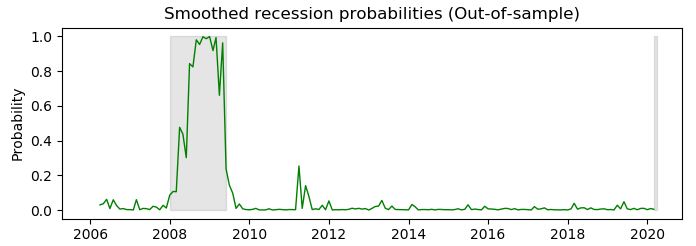
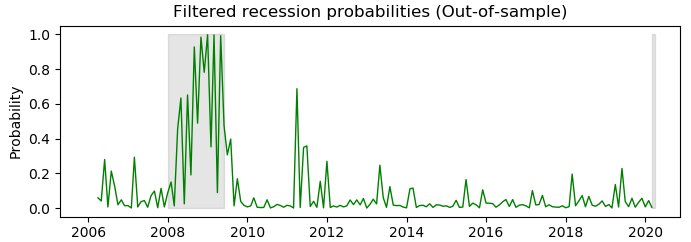


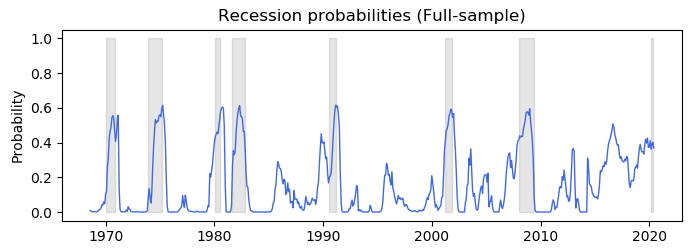
**Figure 16:** In-sample Smoothed and Filtered Probabilities of Recessions from Dynamic Factor Markov Switching Model based on yield curve data

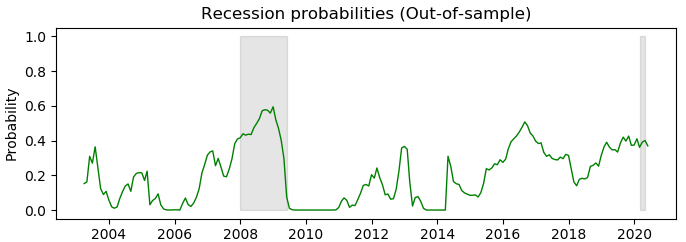


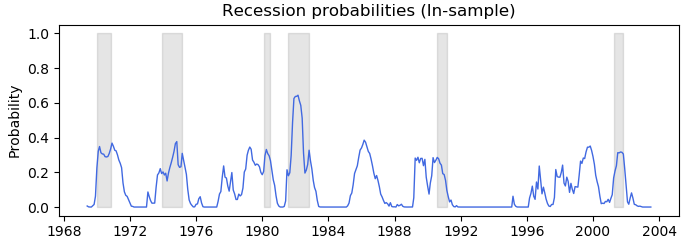


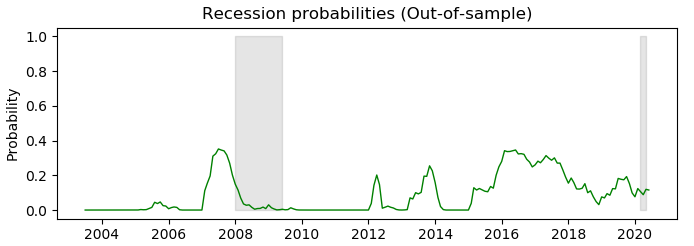
**Figure 17:** Out-of- sample Smoothed and Filtered Probabilities of Recessions from Dynamic Factor Markov Switching Model based on yield curve data

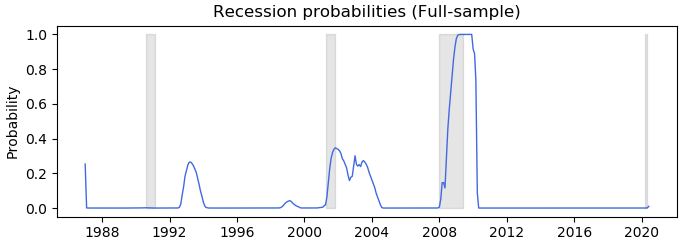


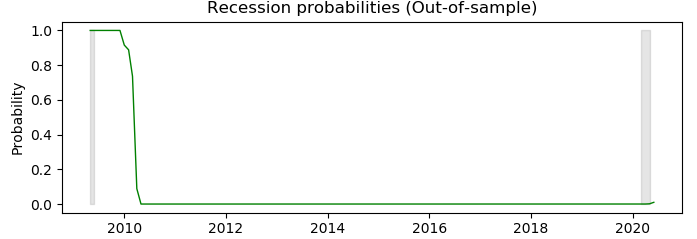
**Figure 18:** Probabilities of Recessions from LSTM Neural Network based on coincident indicators using a 3 month forecast horizon



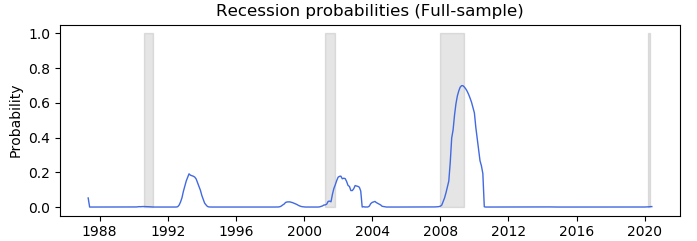
**Figure 19:** Probabilities of Recessions from LSTM Neural Network based on coincident indicators using a 12 month forecast horizon

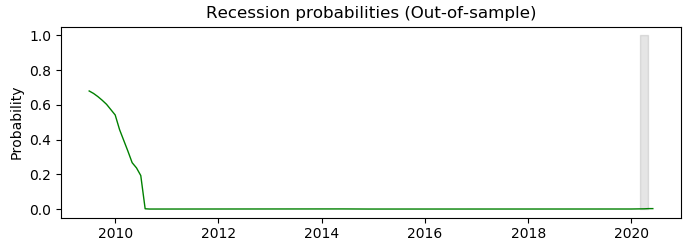


**Figure 20:** Probabilities of Recessions from LSTM Neural Network based on yield curve using a one month forecast horizon



**Figure 21:** Probabilities of Recessions from LSTM Neural Network based on yield curve using a 12 month forecast horizon





1. Please refer to table 1 in the appendix for an overview of the last business cycle dates as identified by the NBER. [↑](#footnote-ref-1)
2. Please refer to the methodology section for further explanation of the vanishing gradient problem. [↑](#footnote-ref-2)
3. This observation can also be confirmed by figure 2. [↑](#footnote-ref-3)
4. Please refer to figures 14 and 15 in the appendix for a separate illustration of in-sample and out-of-sample smoothed and filtered recession probabilities based on the dynamic-factor Markov-switching model applied with coincident economic factors. [↑](#footnote-ref-4)
5. Accuracy rate = (true positive + true negative) / (true positive + false positive + true negative + false negative) [↑](#footnote-ref-5)
6. Precision rate = (true positive) / (true positive + false positive) [↑](#footnote-ref-6)
7. Please refer to figures 16 and 17 in the appendix for a separate illustration of in-sample and out-of-sample smoothed and filtered recession probabilities based on the dynamic-factor Markov-switching model applied with coincident economic factors. [↑](#footnote-ref-7)
8. Please note, that as described in the previous section, the actual split between training and testing data used for the implementation of the model refers to a time series split of n=2 (416 training and 207 testing samples for the application with coincident economic indicators, 298 training and 149 testing samples for the application with yield curve data). This is due to the intention to include at least on historic recession in the testing data. [↑](#footnote-ref-8)
9. Please refer to figures 18 and 19 in the appendix for an illustration recession probabilities based on the LSTM neural network model applied with coincident economic factors for 3 month and 12 months time horizon [↑](#footnote-ref-9)
10. Please refer to figures 20 and 21 in the appendix for an illustration of recession probabilities based on the LSTM neural network model applied with yield curve data for 1 month and 12 months time horizon [↑](#footnote-ref-10)