

HMM-LSTM Fusion Model for Economic Forecasting

Guhan Sivakumar

May 8, 2024

Abstract

This paper explores the application of Hidden Markov Models (HMM) and Long Short-Term Memory (LSTM) neural networks for economic forecasting, focusing on predicting CPI inflation rates. The study explores a new approach that integrates HMM-derived hidden states and means as additional features for LSTM modeling, aiming to enhance the interpretability and predictive performance of the models. The research begins with data collection and preprocessing, followed by the implementation of the HMM to identify hidden states representing distinct economic conditions. Subsequently, LSTM models are trained using the original and augmented data sets, allowing for comparative analysis and evaluation. The results demonstrate that incorporating HMM-derived data improves the predictive accuracy of LSTM models, particularly in capturing complex temporal patterns and mitigating the impact of volatile economic conditions. Additionally, the paper discusses the implementation of Integrated Gradients for model interpretability and provides insights into the economic dynamics reflected in the forecasting outcomes.

Keywords: Inflation, Recession, HMM, LSTM, RNN, Forecasting, Means, Hidden States, Fusion, Machine Learning, Economics, Interpretability

Table of Contents

1	Introduction	2
2	Literature Review	4
2.1	Machine Learning Methods in Economics	4
2.2	Forecasting China's GDP Using LSTM and HMM	4
2.3	Deep Learning Applications in Economic Forecasting	5
2.4	Economic State Prediction Using Hidden Markov Models	5
2.5	Interpretable Neural Networks for Panel Data Analysis	6
2.6	Forecasting Economics and Financial Time Series with LSTM	6
3	Data Collection and Cleaning	7
3.1	Feature Engineering	7
3.2	Correlations and Stationarity	8
4	Hidden Markov Models	9
4.1	Expectation Maximization and Viterbi Decoding	9
4.2	Markov Chain Probabilities	10
4.3	Determining the Hidden States and Intuition	11
4.4	HMM with Oil Price Data	13
5	Neural Networks	15
5.1	Artificial and Recurrent Neural Networks	15
5.2	Long Short-Term Memory (LSTM) Models	16
5.3	In-Sample Forecasting	17
5.4	Feature Importance using Integrated Gradients	18
5.5	Time Series Cross Validation	20
6	Out-of-Sample Forecasting	22
6.1	1-year Forecast	22
6.2	2.5-year Forecast	24
7	Future Work	26
8	Conclusion	27
References		29
Appendix A: Tables		30
Appendix B: Figures		31

1 Introduction

Forecasting CPI Inflation has been a challenge for several years due to the complexity of economic dynamics. In recent years, advanced time series machine learning models have been applied to solve economic problems, and they have been shown to perform better than traditional forecasting methods; however, the major drawback is the interpretability of the models which is crucial for economic analysis. This paper explores the integration of Hidden Markov Models (HMM) and Long Short-Term Memory (LSTM) neural networks to forecast CPI Inflation, as well as implements Integrated Gradients for feature explainability.

Major economic indicators were selected as features, and the selected economic indicators provide valuable insights into different aspects of the economy, ranging from monetary policy to market sentiment. The forecasting model gains a comprehensive understanding of the various factors influencing inflation rates and mitigates the risk of relying on just one variable. With multiple features, the model can learn and identify different economic trends and adjust the forecasts accordingly. Appropriate data cleaning was done to handle the different variables and ensure consistency throughout the modeling process.

An HMM analysis was first done to determine the validity of the features vs National Bureau of Economic Research (NBER) recession indicators, and subsequently the hidden states and means derived from the HMM were used as additional features in the LSTM network. LSTM networks work similarly to the human brain and perform significantly better than artificial neural networks when dealing with time-dependent data. By augmenting the features, the LSTM is given more information regarding structural trends present in the data. Different feature combinations were tested on the LSTM and validated thoroughly to ensure optimal performance. Finally, a similar model architecture was then used to get 1-year and 2.5-year forecasts and compared against official forecasts.

To overcome the interpretability issue, traditional methods such as Shapley Values and

Local Interpretable Model-Agnostic Explanations (LIME) were tested; however, they have not been adapted yet to work with all custom models. So, a less commonly used technique in time series forecasting, Integrated Gradients, was used to get an approximation of feature importance. Though not perfect, it can provide insights into the LSTM network which is generally considered a “Black-Box” model. Identifying the important economic features gives insights into how the different features affect the forecast, which can be crucial to economists and stakeholders.

This research aims to solve some of the major issues of economic forecasting with advanced machine learning techniques. By creating a fusion model, the forecasts are more robust and the addition of integrated gradients provides interpretability needed for economic analysis. Bridging the gap between traditional economic modeling and advanced machine learning offers a promising avenue for more robust and insightful economic forecasts. As technology continues to evolve, the integration of innovative methodologies becomes essential for policymakers, analysts, and researchers.

2 Literature Review

2.1 Machine Learning Methods in Economics

Athey and Imbens' paper discusses the effectiveness of machine learning methods and their application in economics. Specifically, “[neural networks] have been found to be very successful in complex settings, with extremely large number of features” ([Athey and Imbens, 2019](#)). The addition of multiple hidden layers allows the models to capture non-linear trends between the features, and they mention that the main interest is in “the properties of the predictions from these specifications” ([Athey and Imbens, 2019](#)), and not the parameter estimates. Traditional methods to solve for the non-linear weights would lead to poor results, so different algorithms have been implemented, such as back-propagation. The growing success of machine learning methods in economics, highlighted by the effectiveness of neural networks in capturing complex relationships, signifies their potential to enhance forecasting.

2.2 Forecasting China’s GDP Using LSTM and HMM

In this paper, the authors discuss the advantages of using an LSTM-HMM approach to forecast GDP fluctuation states for China’s GDP. Compared to a standard HMM model, “[they] select the predicted real time CPI fluctuation state by LSTM to be the observable state when we forecast the real time GDP fluctuation state, instead of the observed CPI fluctuation states with time lag” ([Zheng et al., 2023](#)). This approach led to better performance and more consistent results in their long-term window testing, most notably “within the 10-year time window, ..., the LSTM-HMM monthly univariate input has the best performance in prediction” ([Zheng et al., 2023](#)). The integration of LSTM and HMM models in forecasting China’s GDP offers improved accuracy and stability, particularly evident in long-term predictions, showcasing the efficacy of this combined approach for economic forecasting.

2.3 Deep Learning Applications in Economic Forecasting

With the rise in deep learning, the authors explore various architectures and determine which models are preferred and successful in economics. Since RNNs only consider the most recent state, “the cellular state of LSTM determines which states should be left behind and which states should be forgotten. Hence, LSTM plays an important role in many fields of economic research” ([Zhang et al., 2022](#)). In a list of papers that utilize RNNs. specifically LSTMs, they noticed that the LSTM model outperformed traditional econometric approaches such as ARIMA, MA, vanilla neural networks, etc. Some of the drawbacks they mention for RNNs are, “there are many parameters that need to be trained, which are prone to gradient dissipation or gradient explosion; [and they are] without feature learning ability” ([Zhang et al., 2022](#)). The authors highlight the superiority of LSTMs over traditional econometric methods, despite noted challenges like parameter tuning and feature explainability.

2.4 Economic State Prediction Using Hidden Markov Models

The author explores the effectiveness of HMMs to predict the economic state of various countries such as the US, UK, Japan, etc. Various features relevant to the economy were selected to further enhance the capability of these models. Rather than just testing 2-state models, the paper explores using more states, as well as various transformations on the series, to determine if the performance of HMMs improves. From the results, “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” ([Rikken, 2022](#)). Though the 3-state model performs well, the downside is that it is not easily classifiable as the 2-state model, so meaningful results become harder to derive. Despite these drawbacks, HMMs can still be used to identify new economic periods and provide information that may not be captured in traditional classification models.

2.5 Interpretable Neural Networks for Panel Data Analysis

Despite the popularity of advanced machine learning methods, “Economists are still nervous about using more advanced tools like neural networks, since they mostly deliver outcomes from a complicated black box without transparency or interpretability” ([Yang et al., 2020](#)). To solve this issue, the authors created a modified neural network that achieves both high accuracy and interpretability. Their results demonstrate the power of an interpretable neural network, and that “the increase in interpretability may allow models to achieve higher accuracy in the test set than unrestricted models” ([Yang et al., 2020](#)). This paper highlights a new approach to the importance of transparency in black-box models, and the rise in Explainable AI has led to model agnostic interpretability, such as Shapley values, a concept first introduced to solve a problem in cooperative game theory.

2.6 Forecasting Economics and Financial Time Series with LSTM

To emphasize the viability of machine learning methods for forecasting, this paper focuses on comparing the traditional ARIMA model to an LSTM network. The two models were tested on financial data, such as IXIC, GSPC, DJI, etc., and economic data such as housing, medical care, exchange rate, etc., and “the study shows that LSTM outperforms ARIMA. The average reduction in error rates obtained by LSTM is between 84 - 87 percent when compared to ARIMA indicating the superiority of LSTM to ARIMA” ([Siami-Namini and Namin, 2018](#)). The paper also discusses the importance of hyper-parameter tuning and overfitting with regards to the LSTM, and they noticed that there was “no improvement when the number of epochs is changed” ([Siami-Namini and Namin, 2018](#)). The results from this paper demonstrate the predictive power of LSTMs and their application in finance and economics. Advancements in deep learning will eventually lead to better models and even outperform LSTMs.

3 Data Collection and Cleaning

3.1 Feature Engineering

The data was collected from ([St. Louis Fed, 1991](#)) and ([Yahoo Finance, 2024](#)). To cover various aspects of the economy, several key features were selected, such as Federal Funds Rate, Unemployment Rate, GDP growth rate, PPI Rate, etc. Since some of the data was quarterly and daily, interpolation and aggregation were done respectively to make the data into monthly time steps. The features were converted to growth rates to ensure consistency with the target variable, the CPI Inflation Rate. The data ranges from 1970-01-01 to 2023-01-01, and [Figure 1](#) is a plot that provides an overview of the features.

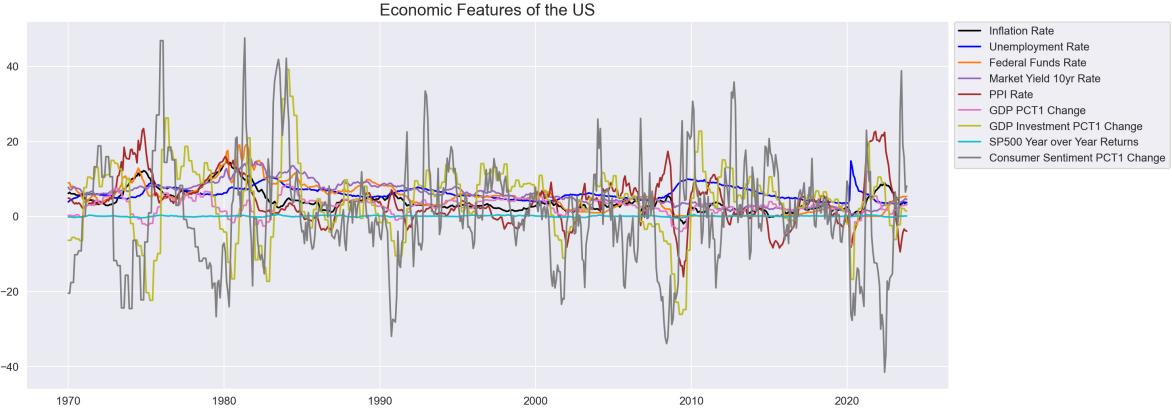


Figure 1: Economic Features of the US

Descriptive statistics are provided in [Table 1](#). There are a total of 647 monthly observations, and the table gives us historical information about the economy.

	inflation_rate	unemployment_rate	federal_funds_rate	market_yield_10_rate	ppi_rate	gdp_pc1	gdp_invest_pc1	sp500_yoy	consumer_senti_pc1	oil_price_pc1	recessions
count	647.0000	647.0000	647.0000	647.0000	647.0000	647.0000	647.0000	647.0000	647.0000	647.0000	647.0000
mean	4.0399	6.1306	4.9066	6.0302	3.9061	2.7262	3.9420	0.0853	0.3952	12.4316	0.1314
std	2.9384	1.7038	3.9166	3.1398	6.0956	2.3174	9.5545	0.1616	13.8638	41.6073	0.3381
min	-1.9588	3.4000	0.0500	0.6200	-16.0584	-7.5285	-26.0881	-0.4251	-41.5205	-74.0839	0.0000
25%	2.1086	4.9000	1.4250	3.5450	0.4937	1.6650	-1.4419	-0.0057	-7.3773	-8.8633	0.0000
50%	3.1940	5.8000	4.9900	5.9800	3.5683	2.8589	4.7050	0.1015	0.0000	5.1300	0.0000
75%	5.0354	7.2000	6.9100	7.8850	6.3669	4.1878	9.6544	0.1922	7.3677	24.4842	0.0000
max	14.5923	14.8000	19.1000	15.3200	23.4409	11.9503	39.1907	0.5264	47.5822	272.9305	1.0000

Table 1: Summary Statistics

Without accounting for breaks, the average inflation rate has been around 4%, and the current inflation rate as of March 2024 is 3.475%. Unemployment rate is below its historical

average, and due to the dual mandate described in (Thorbecke, 2002), there has not been any pressure to decrease interest rates from 5.33%. West Texas Intermediate (WTI) Crude Oil price was considered as a feature, but due to its range and volatility, it was not included in the model. Information regarding oil price and its effects can be found in [Sub-section 4.4](#).

3.2 Correlations and Stationarity

The correlation matrix in [Figure 2](#), shows that there are no unexpected significant correlations with inflation rate. While multicollinearity exists between certain variables like GDP and GDP Investment, each contributes differently to inflation forecasting. GDP Investment acts as a leading indicator for predicting inflation, and GDP offers insights into overall economic growth.

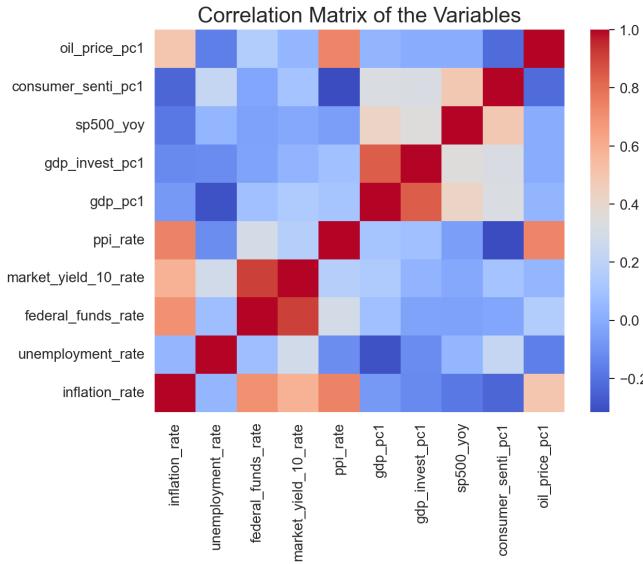


Figure 2: Correlation Matrix of the Variables

The Augmented Dickey-Fuller (ADF) test by ([Dickey and Fuller, 1979](#)) was conducted to test stationarity and the results are found in [Table A1](#). Only two of the series were non-stationary, but HMMs and LSTMs are able to capture non-linear relationships making them more robust to non-stationary trends.

4 Hidden Markov Models

HMMs offer a powerful framework, created by (Rabiner, 1989), for capturing the dynamic nature of economic systems, and they work by determining latent or unobservable states in time series data. It is a probabilistic model and aims to maximize the likelihood of being in a particular state at a given time step. For this paper, four hidden states were used, and the states are indicative of the prevailing economic environment during those periods. There are three fundamental steps to HMMs at each time step, determine initial probabilities, compute the probability of the states, decode the probabilities and determine the latent state.

4.1 Expectation Maximization and Viterbi Decoding

The Expectation Maximization (EM) algorithm is a numerical method used to estimate the Maximum Likelihood Estimator. There are three main probabilities that the HMM computes, initial, transmission, and emission probabilities; however, the implemented method for this paper assumes Gaussian emissions. The algorithm requires prior knowledge, so the Dirichlet distribution, a common distribution used as a prior for Bayesian estimation, is used as the prior. The EM algorithm by (Dempster et al., 1977), is described in Equation 1

$$\begin{aligned} Q(\boldsymbol{\theta} | \boldsymbol{\theta}^{(t)}) &= E_{\mathbf{Z} \sim p(\cdot | \mathbf{X}, \boldsymbol{\theta}^{(t)})} [\log p(\mathbf{X}, \mathbf{Z} | \boldsymbol{\theta})] \\ \boldsymbol{\theta}^{(t+1)} &= \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \, Q(\boldsymbol{\theta} | \boldsymbol{\theta}^{(t)}) \\ \implies \boldsymbol{\theta}^{(t+1)} &= \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \, E_{\mathbf{Z} \sim p(\cdot | \mathbf{X}, \boldsymbol{\theta}^{(t)})} [\log p(\mathbf{X}, \mathbf{Z} | \boldsymbol{\theta})] \end{aligned} \tag{1}$$

where the first equation is the Expectation (E) step and the second is the Maximization (M) step. During the E-step, the algorithm computes the expected values of the latent variables given the observed data and the current parameter estimates. In the M-step, it maximizes the likelihood of the observed data under the model, updating the parameters based on

the computed expected values. This iterative process continues until convergence, yielding parameter estimates that optimize the fit between the model and the observed data.

After calculating the probabilities at each step, the Viterbi Decoding Algorithm is utilized to determine the state at the given time step. The algorithm, created by ([Viterbi, 1967](#)), is given below in [Equation 2](#)

$$\begin{aligned} \max_k V_T^k &= \max_{\{S_t\}_{t=1}^T} p\left(\{S_t\}_{t=1}^T, \{O_t\}_{t=1}^T\right) \\ V_t^k &= \max_{S_1, \dots, S_{t-1}} p(S_t = k, S_1, \dots, S_{t-1}, O_1, \dots, O_t) \\ &= p(O_t | S_t = k) \max_i p(S_t = k | S_{t-1} = i) V_{t-1}^i \end{aligned} \quad (2)$$

where the first equation computes the maximum probability V_t^k of the final state k over all the possible state sequences $\{S_t\}_{t=1}^T$ given the observed sequence $\{O_t\}_{t=1}^T$. The second equation computes the maximum probability V_t^k of being in state k at time step t . It recursively calculates this probability based on the observation O_t given state k and the maximum probability of transitioning from state k from the previous time step $t - 1$. The algorithm provides an optimal approach to determine the most likely sequence of the hidden states based on the observed data and transition probabilities.

4.2 Markov Chain Probabilities

To visualize the transmission matrix generated by the HMM, a Markov chain is provided below in [Figure 3](#). For this paper, four hidden states were used as there are multiple stages of economic recovery and expansion, and two states may not fully capture that. Though there are some probabilities that are 0, there are no absorbed states which means that the states can freely transition from one another and not get stuck in a specific state. The stationary distribution, or equilibrium distribution, of the Markov chain is as follows, State 0: 0.487,

State 1: 0.133, State 2: 0.247, State 3: 0.133.

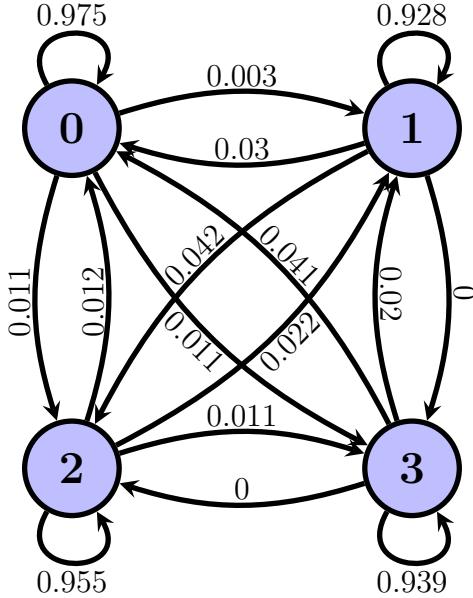


Figure 3: Transmission Probabilities of HMM

For numerical stability, a diagonal covariance was chosen since the correlation matrix determined that there were no major correlations between the target and the features. A diagonal covariance matrix assumes that there are no relationships between the different variables, whereas a full covariance matrix assumes that all the features are interrelated. Although economic theory suggests that the variables are related, the data shows weak relationships between them, and using a full covariance matrix led to absorbed states and overfitting of the model.

4.3 Determining the Hidden States and Intuition

To assess the validity of the HMM, [Figure 4](#) shows the comparison of the hidden states to the NBER recessions. The model is able to capture important recessions such as the 2008 financial crisis, COVID-19, and the oil crisis in the 1970s, and is also able to pick up on smaller economic crises such as the 2015 decline in industrial production and state recessions. To simplify the intuition of the hidden states, we describe them as follows, state 0 represents

economic stability, while states 1, 2, and 3 represent an economic crisis.

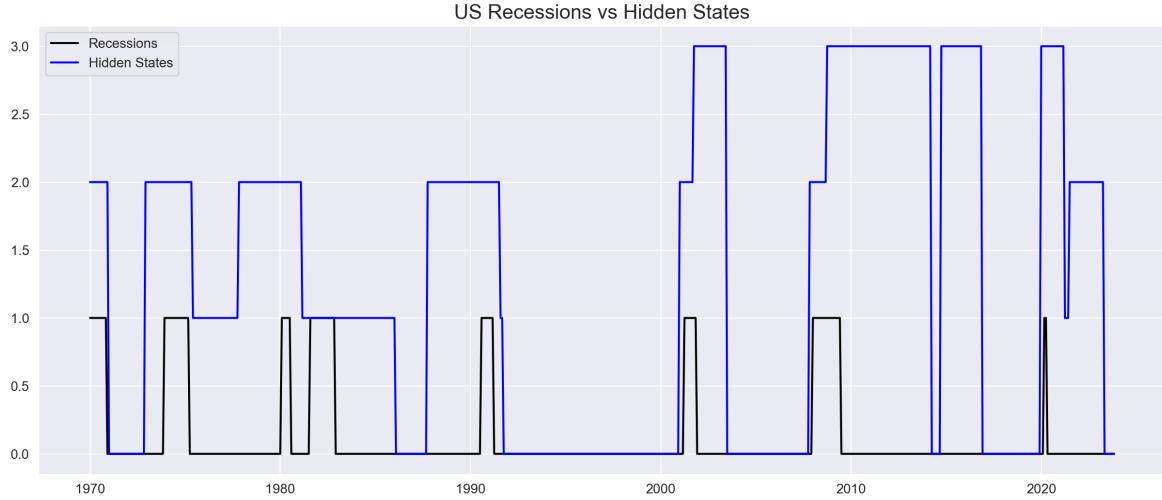


Figure 4: HMM to Recessions

Since the NBER recessions are either 0 or 1 a transformation was done to bring the maximum of the hidden states down to one, and can be seen below in Figure 5.

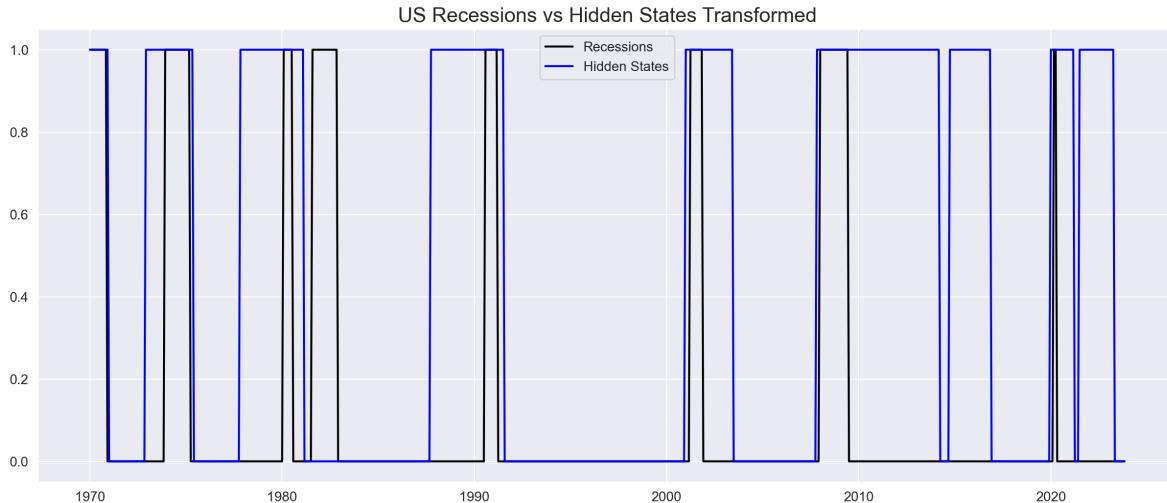


Figure 5: HMM to Recessions Transformed

The hidden states match the recessions relatively well and Table 2 is the classification report. The model attains a 62% accuracy given the severe class imbalance, which suggests good performance, and the precision and recall scores suggest that the model is not just predicting one class and is able to learn how the economy is functioning. Unfortunately,

standard transformations such as down or upsampling cannot be done to fix the class imbalance, since they would not be an accurate representation of the US economy. The HMM returns the hidden states and the means for each of those states by feature, and [Figure B1](#) is available in the appendix and visualizes the hidden states and means at each state for each variable.

	precision	recall	f1-score	support
0.0	0.95	0.59	0.73	562
1.0	0.23	0.81	0.36	85
accuracy			0.62	647
macro avg	0.59	0.70	0.55	647
weighted avg	0.86	0.62	0.68	647

Table 2: Classification Report

4.4 HMM with Oil Price Data

To ensure that all factors of the economy were covered, WTI crude oil price was considered as a feature. [Figure 6](#) shows a comparison of the CPI Inflation Rate and percent change in oil price, where the right y-axis is the oil price, and the left y-axis is the CPI inflation rate.

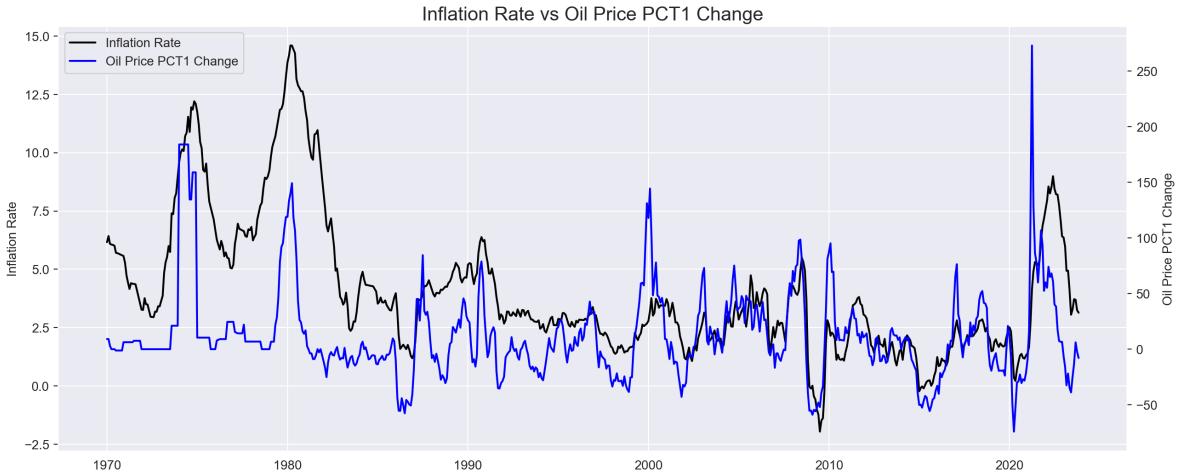


Figure 6: Inflation Rate vs. Oil Price

Oil price is extremely volatile as seen above, and though it follows the same trend as inflation,

due to its range it is given more weight and importance in the selected models. [Figure 7](#), shows the performance of the HMM with the oil price data as a feature. A diagonal covariance matrix was used for this as well to ensure consistency, and visually the HMM does not capture the recessions as effectively.

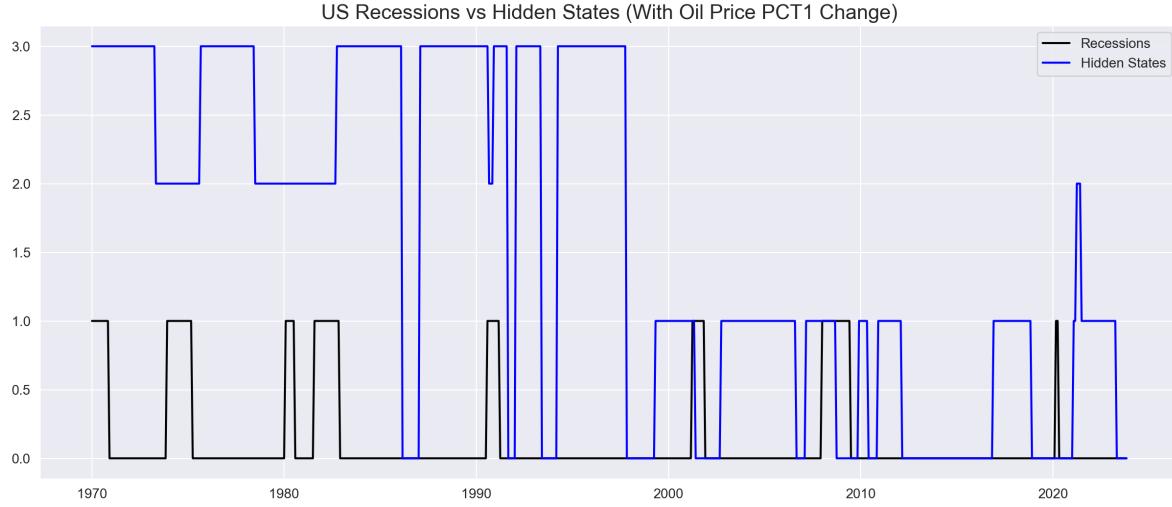


Figure 7: HMM (With Oil) to Recessions

The decision to omit oil price data was made to streamline the modeling process, prioritizing factors with more stable and consistent effects on inflation. Focusing on more stable indicators enhances the robustness of the forecasting models, making them better equipped to handle various economic conditions and changes in market dynamics over time.

Since oil price was excluded, the hidden states and the means derived by the HMM, from [Figure 4](#), without oil price, were added as features to the original dataset creating three new datasets: original data with hidden states as a new feature, original data with means as new features, and original data with hidden states and means as new features. Integrating these derived features into our modeling approach, can enhance the predictive capabilities of other models and gain a deeper understanding of the economic factors influencing inflation rates.

5 Neural Networks

Neural networks are brain-inspired supervised learning models made with interconnected nodes and layers to learn patterns and make predictions from data. Unlike linear models, they are able to capture non-linear relationships between the target and features; however, this leads to their biggest drawback, interpretability. Neural networks are black-box models, which means that the user only sees what goes in and what comes out, which makes them difficult to interpret compared to traditional models. Neural networks have seen a rise in usage in Economics, but due to their lack of interpretability, they get overshadowed by simpler models. There are different types of neural networks, and this paper focuses on Recurrent Neural Networks (RNN), specifically Long Short-Term Memory (LSTM) networks.

5.1 Artificial and Recurrent Neural Networks

Artificial neural networks (ANN) are the simplest type of neural networks. They consist of multiple layers and neurons and can be used for regression and classification tasks. Each node has an activation function that transforms the incoming input and passes it on to the next layer. Common activation functions are sigmoid, tanh, ReLU, and linear, and depending on the task different functions are picked. Training an ANN involves adjusting weights using algorithms like backpropagation to minimize the difference between predicted and actual outputs.

RNNs introduce recurrent connections, allowing information to persist and be shared across different time steps in sequential data, where temporal dependencies are crucial. These models are able to handle variable-length sequences, allowing them to adapt dynamically to different lengths of input data. Their architectures can be extended and optimized with variations such as Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU), which improve the model's ability to retain and utilize information over longer sequences.

5.2 Long Short-Term Memory (LSTM) Models

LSTMs are a type of recurrent neural network and function very similarly to the human brain. They are designed to address the vanishing gradient problem in traditional RNNs, allowing them to better capture time dependencies in sequential data, and they achieve this through the use of gated units called memory cells that can maintain information over time.

[Figure 8](#) represents a single LSTM neuron, and shows how the information is processed.

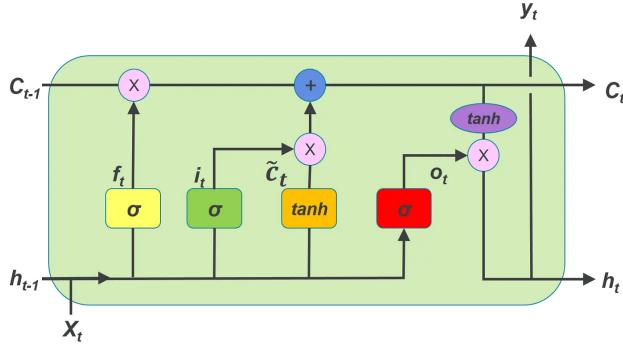


Figure 8: LSTM Neuron at any time step t ([Thakur, 2018](#))

Memory cells with input, forget, and output gates regulate the flow of information, allowing the model to selectively remember or forget information. They are trained using backpropagation through time (BPTT), adjusting the weights to minimize the error in the predictions.

[Equation 3](#) from ([Hochreiter and Schmidhuber, 1997](#)) is the matrix representation for the LSTM cell, and unlike regular neurons in a neural network, LSTM neurons use tanh as the main activation function and sigmoid as the recurrent activation function.

$$\begin{pmatrix} i \\ f \\ o \\ g \end{pmatrix} = \begin{pmatrix} \sigma \\ \sigma \\ \sigma \\ \tanh \end{pmatrix} W \begin{pmatrix} h_{t-1} \\ x_t \end{pmatrix} \quad (3)$$

$$h_t = o_t \odot \tanh(f \odot c_{t-1} + i \odot g)$$

For this research, the LSTM architecture was defined as follows: Input layer, 2 LSTM layers (tanh and sigmoid activation), 1 Dense Layer (linear activation), and the output layer.

5.3 In-Sample Forecasting

With the defined architecture, the model was first tested on a 70:30 split of the data. This particular split was chosen as it allowed for one recession and one expansion to be in the training and testing set. The training set ends at the bottom of the 2008 recession, and the test set starts at the bottom of the 2008 recession to the present. Before training the model, preprocessing of the data was done to ensure numerical stability. Since LSTMs are sensitive to the input data, MinMaxScaler, defined in [Equation 4](#), was used to scale the values in the columns to be between 0 and 1.

$$x' = a + \frac{(x - \min(x))(b - a)}{\max(x) - \min(x)}, a = 0 \text{ and } b = 1 \quad (4)$$

The scaler ensured that all features were uniformly scaled, preventing any single variable from dominating the model's learning process due to differences in scale. To capture the time-dependent information, 24 lags were used on all the variables and transformed into a 3D matrix (tensor) for the LSTM.

After preprocessing, the model was tested on these 4 datasets, original data, original data with hidden states as a feature, original data with means as features, and original data with hidden states and means as features. Their performance is shown below in [Table 3](#) and we can see a significant improvement in all 3 metrics by adding the means as features

	Original Data	Hidden State Data	Means Data	All HMM Data
MSE	1.992	3.548	0.818	0.861
MAE	0.964	1.290	0.694	0.705
R ²	0.519	0.144	0.803	0.792

Table 3: Metrics Report

to the dataset. This improvement indicates that the LSTM model effectively leveraged the additional information provided by the means, allowing it to better capture the underlying time dependencies present in the data. Visually, we can see the improvement in [Figure 9](#),

where the data with the means is able to predict the rise and fall in 2022.

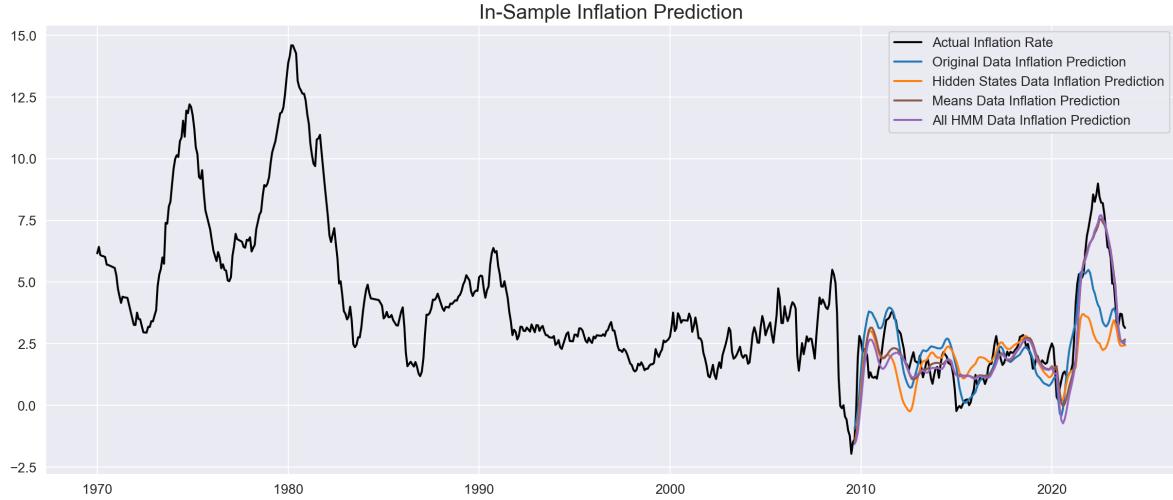


Figure 9: In-Sample Inflation Prediction

Unfortunately, due to the minimal rows, only a train-test split was done, and the training loss curves for the different predictions can be found in [Figure B2](#); we can see a smooth decrease with no sharp turns, suggesting that the model is not overfitting. Compared to traditional methods, the LSTM forecast is able to capture the temporal dependencies and predict CPI inflation rates.

5.4 Feature Importance using Integrated Gradients

With the increase in popularity of Explainable AI, Integrated Gradients (IG) have become popular a technique for feature importance for complex models such as neural networks. They are commonly used in image classification tasks, but since they work for any differentiable model, it has been adapted to interpret time series data. IG works by computing the integral of the gradient of a baseline input and the actual input, a Riemann approximation is used and is given below in [Equation 5](#) from ([Sundararajan et al., 2017](#)):

$$IG_i(x) \approx (x_i - x'_i) \times \sum_{k=1}^m \frac{\partial F\left(x' + \frac{k}{m} \times (x - x')\right)}{\partial x_i} \times \frac{1}{m} \quad (5)$$

Since the IG computation can get very computationally expensive, m was set to 50, so the values are just an approximation. However, they still provide insights into how the LSTM is able to learn and determine the relationships between the economic variables. [Figure 10](#) is the integrated gradients, averaged across the test set and lags, for the different features. The LSTM is able to correctly determine the inverse relationship between the inflation rate and unemployment rate, and the relationship with the other features follows the economic theory. By adding the means as features, it increases the importance of the lags of inflation and the other variables which could be leading to the prediction improvements.

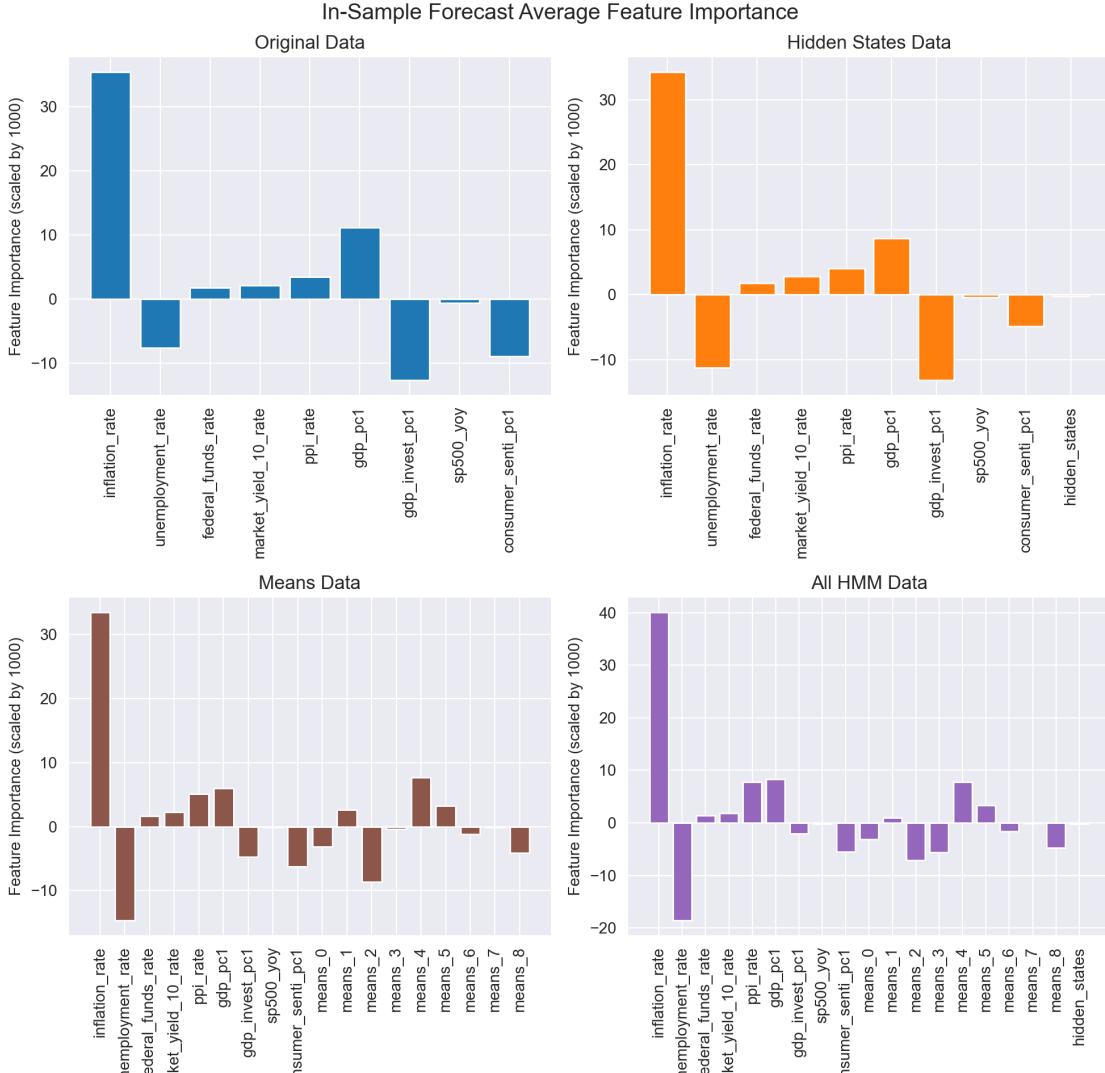


Figure 10: In-Sample Forecast Average Feature Importance

5.5 Time Series Cross Validation

To assess the LSTM model across different time periods, forward chaining cross-validation was conducted. Since there are only around 650 rows of data, and neural networks require a lot of data, 5 folds were chosen. [Table 4](#) provides a summary of the metrics across the folds

		Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Original Data	MSE	11.580	0.486	0.490	1.080	1.631
	MAE	2.930	0.597	0.549	0.740	1.008
	R ²	-0.911	0.590	0.330	0.509	0.711
Hidden State Data	MSE	11.895	1.046	0.623	0.862	0.799
	MAE	3.080	0.884	0.629	0.672	0.677
	R ²	-0.963	0.118	0.149	0.608	0.858
Means Data	MSE	10.044	1.125	1.152	1.223	0.903
	MAE	2.728	0.891	0.786	0.794	0.716
	R ²	-0.657	0.052	-0.574	0.444	0.840
All HMM Data	MSE	8.694	1.522	1.390	1.360	0.694
	MAE	2.566	1.018	0.823	0.838	0.632
	R ²	-0.435	-0.283	-0.901	0.382	0.877

Table 4: Metrics Report (5 Fold CV)

We can see an improvement in the metrics with each fold which makes sense as the model has more data to learn from. [Figure 11](#) is shows the predicted inflation rates

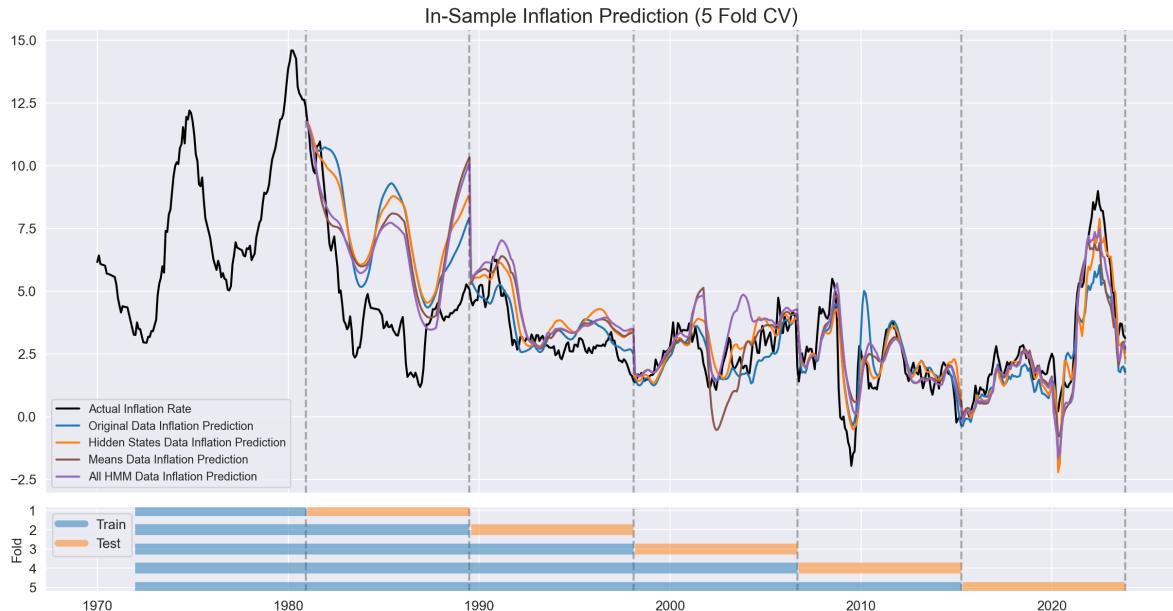


Figure 11: In-Sample Inflation Prediction (5 Fold CV)

With 5 folds, COVID and the 2008 financial crisis were given their own fold, allowing for a comprehensive assessment of the model’s generalizability during economic crises. We can see in the plot that the model is able to predict the recession, and over time the model performs better. The plot of the loss curves are in [Figure B3](#), and across folds demonstrated consistent performance, although fold 4 exhibited slightly higher variability, possibly indicating that the 2008 recession was difficult for the LSTM to capture. Unfortunately, Integrated Gradients were not feasible for cross-validation (CV) due to computational constraints. Additionally, while hyperparameter tuning and regularization were conducted, they were not extensively explored for the same reason.

LSTMs prove to be powerful tools for time series forecasting, leveraging recurrent connections to capture temporal dependencies effectively. The observed improvement emphasizes the synergy between HMM-derived features and LSTM modeling, highlighting the value of integrating diverse modeling approaches. The fusion enhances predictive accuracy and also offers valuable insights into economic dynamics, allowing us to extend this approach to forecast future inflation rates.

6 Out-of-Sample Forecasting

6.1 1-year Forecast

Following the in-sample forecast, future forecasts were conducted to predict the inflation rate. The output layer of LSTM architecture was slightly modified to account for the forecast horizon. For the 1 year forecast, the output layer consisted of 12 neurons, one for each month. LSTMs are able to handle multiple features and account for them in the forecast, which makes them valuable in economics. To ensure that enough past data was captured, without overfitting, a total of 48 lags were used. The forecast with their confidence intervals is shown in [Figure 12](#)

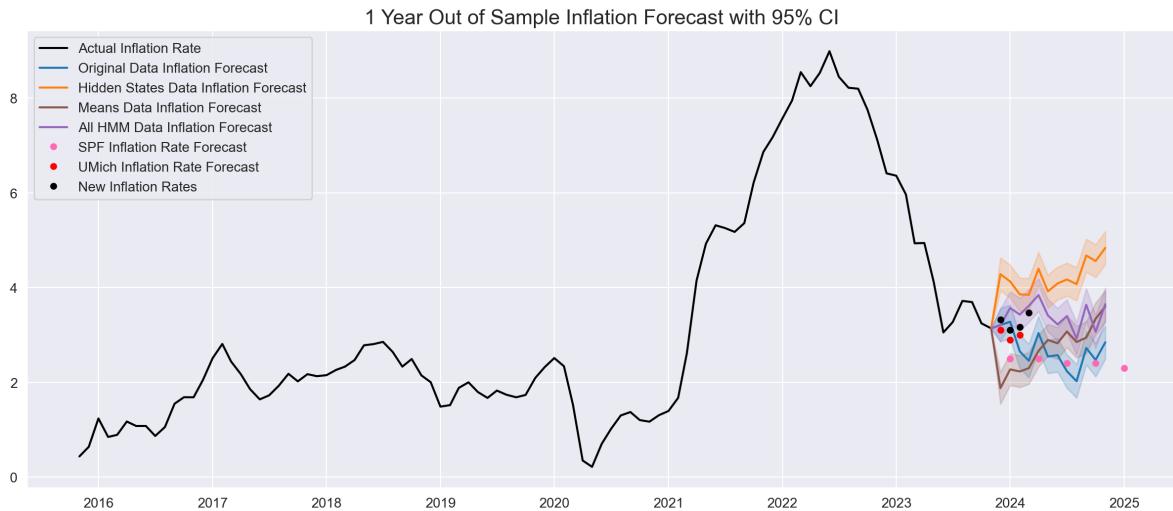


Figure 12: 1 Year Out-of-Sample Forecast

Based on the in-sample performance, the model with all the data was expected to perform the best, and it is seen when comparing it to new inflation rate values, as they fall in the confidence interval. Since that data was collected up until November 2023, the following inflation rates were used as a benchmark. The model was able to predict the March 2024 inflation rate while official forecasts had predicted a decline. Incorporating the HMM-derived features provided valuable insights into economic conditions, enhancing the predictive capa-

bility of the model and contributing to its accurate forecasting performance. To visualize the feature importance, the IG is provided below in [Figure 13](#)

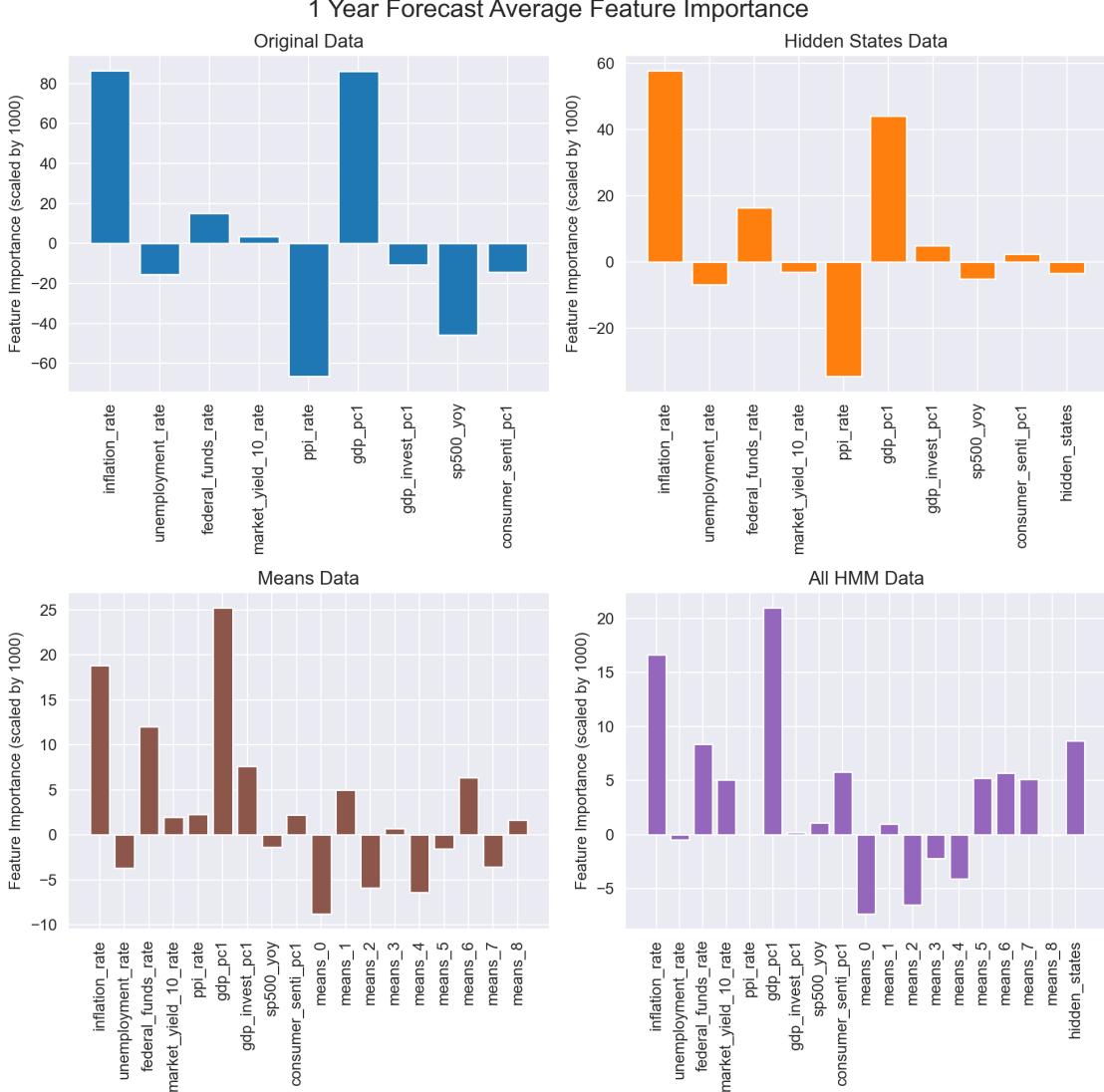


Figure 13: 1 Year Forecast Average Feature Importance

Following economic theory, GDP growth has the most importance in forecasting future inflation rate, followed by the lags of inflation. The means tend to have an inverse relationship with the forecast, and this could be leading to its improved performance. Comparing it to using the original data, the importance of the features is reduced significantly, but the means are adding valuable information that is otherwise not captured.

6.2 2.5-year Forecast

To test the effectiveness of the model, a longer-term forecast was conducted. Similar to the 1-year forecast, the output layer of the LSTM was adjusted to 30 neurons, and 120 lags were used since enough information needed to be captured. Unfortunately, the forecast was not as good as the 1-year forecast, seen in [Figure 14](#), but the addition of the HMM features led to a less volatile forecast when compared to the original data.

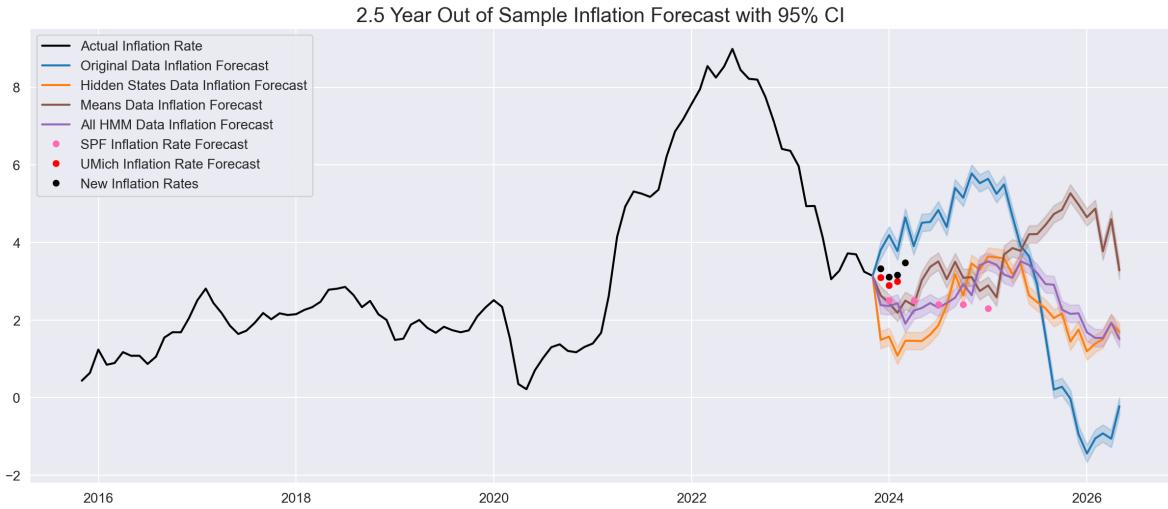


Figure 14: 2.5 Year Out-of-Sample Forecast

The decrease in volatility could be explained by the fact that averaging is a common technique used to reduce variance. By incorporating the HMM means, the inflation rate forecast adjusts accordingly and does not predict a severe recession as indicated by the original data. Though none of the curves match the new inflation rate values, the overall average across the 30 months was around 3%, which has been the average for the last 10-20 years. To further enhance the forecast accuracy, various strategies were explored, including experimenting with different lag lengths and conducting thorough hyperparameter tuning. Despite these efforts, the model's performance for the 2.5-year forecast fell short compared to the 1-year forecast. One of the interesting observations from the IG, in [Figure 15](#), was that GDP investment was given significantly more importance compared to the 1-year forecast.

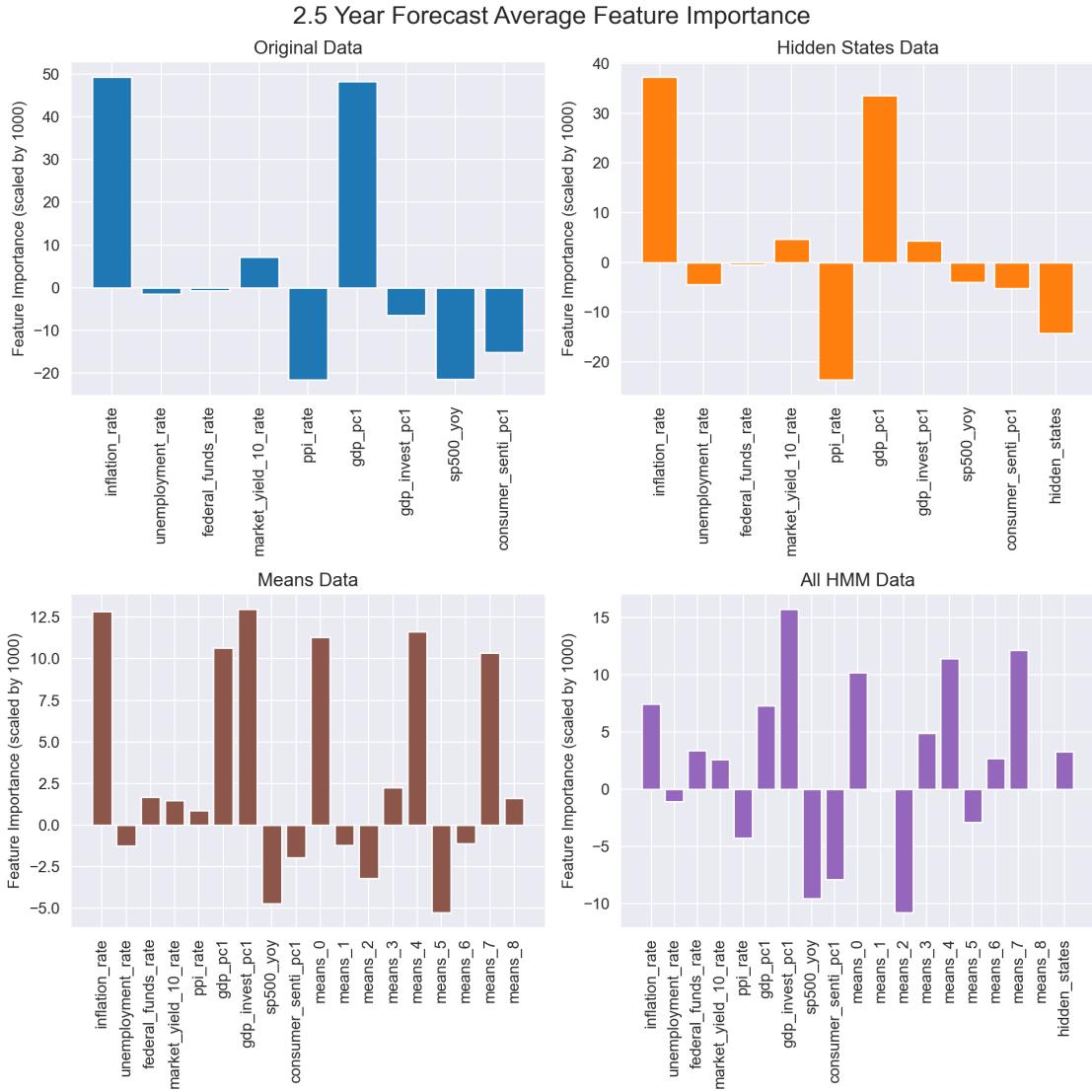


Figure 15: 2.5 Year Forecast Average Feature Importance

GDP investment growth is a leading indicator, so it tends to react before the inflation rate changes. The model could potentially suggest that for longer-term forecasts, more weight should be given to GDP investment compared to the other variables. Additionally the year-over-year returns of SP500 were also significant, though inversely related, while the means of SP500 year-over-year returns have a positive relationship. While the 2.5-year forecast did not achieve the same level of performance as the 1-year forecast, insights gained from feature importance analysis shed light on potential adjustments to improve future forecasting.

7 Future Work

The results from this paper provide promising results; however, there are several avenues for future research. Particularly, GDP and GDP Investment were interpolated, so utilizing the exact values would be more appropriate and could potentially lead to better results. As mentioned earlier, the Integrated Gradients are just an approximation, so with a more powerful machine, more steps could lead to providing more detail into the feature importance. In addition, methods such as Shapley and LIME could be modified to extract insights into recurrent neural networks since they have been proven to perform better than integrated gradients. Similarly, extensive hyperparameter tuning could lead to improved performance in both the HMM and LSTM.

With recent advancements in time series forecasting, different models could be tested such as Temporal Fusion Transformers (TFT), models that use a transformer architecture to obtain an interpretable multivariate forecast. Despite their current computational demands and dataset-specific applicability, advancements in hardware and algorithmic efficiency may soon render TFTs viable for forecasting a wide array of economic variables. Investigating the adaptability and scalability of TFTs within the context of economic forecasting will lead to new insights and predictive capabilities.

Furthermore, exploring ensemble methods that combine the strengths of different forecasting approaches, such as HMMs and LSTMs, could offer further improvements in predictive accuracy and robustness. A Markov switching approach, similar to Markov Switching Autoregressions, could be applied to Neural Networks to create Markov Switching Neural Networks. Instead of just augmenting the features to the dataset, this approach would involve integrating a Markov layer within the neural network architecture to autonomously capture regime switches. By leveraging the complementary strengths of multiple models, the forecasts can navigate the complexities of economic dynamics.

8 Conclusion

This paper presented an interpretable inflation forecasting with a Hidden Markov Model (HMM) and a Long Short-Term Memory (LSTM) fusion model with integrated gradients. The study aimed to explore the effectiveness of incorporating HMM features into LSTM models for inflation prediction, leveraging the dynamic nature of economic data. The fusion model provided valuable insights into the predictive capabilities to capture complex relationships within various economic variables.

The integration of HMM features, such as hidden states and means, significantly enhanced the predictive performance of LSTM models for forecasting. By utilizing the latent state information provided by HMMs, the LSTM models were able to capture subtle shifts in the economy, resulting in more accurate and robust predictions. While the models demonstrated strong performance in short-term predictions, their effectiveness over longer forecasting horizons was somewhat limited. Further modeling, refinement, and experimentation with different economic features could be done to create stable long-term forecasts.

The additional insights gained from the integrated gradients suggest that the model is able to capture the complex dynamics found in economic theory. The augmentation of the HMM-derived data further enhanced the model to determine those relationships and produce reliable forecasts. Compared to the official forecasts and expectations, the fusion model can accurately capture new trends in the 1-year forecast.

With probabilistic modeling, the model was able to capture the inherent uncertainty in economic data, enabling a more robust analysis of inflation trends. Meanwhile, the integration of deep learning with integrated gradients allowed for the extraction of complex patterns and relationships, facilitating deeper insights into the drivers of inflation dynamics. In conclusion, both researchers and policymakers can gain a more nuanced understanding of the US economy to make more informed decisions.

References

- Abadi, Martin, Ashish Agarwal, Paul Barham et al. (2015) “TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems,” <https://www.tensorflow.org/>, Version: 2.15.0.
- Athey, Susan and Guido W. Imbens (2019) “Machine Learning Methods Economists Should Know About,” *Annual Review of Economics*, 11 (Volume 11, 2019), 685–725, [10.1146/annurev-economics-080217-053433](https://doi.org/10.1146/annurev-economics-080217-053433).
- Dempster, A.P., N.M. Laird, and D.B. Rubin (1977) “Maximum Likelihood from Incomplete Data via the EM Algorithm,” *Journal of the Royal Statistical Society. Series B (Methodological)*, 39 (1), 1–38, <http://www.jstor.org/stable/2984875>.
- Dickey, D. and Wayne Fuller (1979) “Distribution of the Estimators for Autoregressive Time Series With a Unit Root,” *JASA. Journal of the American Statistical Association*, 74, [10.2307/2286348](https://doi.org/10.2307/2286348).
- Federal Reserve Bank of Philadelphia (2024) “SPF Q1 2024 Forecast,” <https://www.philadelphiafed.org/surveys-and-data/real-time-data-research/spf-q1-2024>.
- Hochreiter, Sepp and Jürgen Schmidhuber (1997) “Long Short-term Memory,” *Neural computation*, 9, 1735–80, [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735).
- Pedregosa, F., G. Varoquaux, A. Gramfort et al. (2011) “Scikit-learn: Machine Learning in Python,” *Journal of Machine Learning Research*, 12, 2825–2830, <https://scikit-learn.org/stable>, Version: 1.3.2.
- Rabiner, L.R. (1989) “A tutorial on hidden Markov models and selected applications in speech recognition,” *Proceedings of the IEEE*, 77 (2), 257–286, [10.1109/5.18626](https://doi.org/10.1109/5.18626).
- Rikken, T.W. (2022) “The application of hidden Markov models for economic state prediction and the calculation of economic state transition probabilities,” https://essay.utwente.nl/92468/1/Rikken_MA_BMS.pdf.
- Siami-Namini, Sima and Akbar Siami Namin (2018) “Forecasting Economics and Financial Time Series: ARIMA vs. LSTM,” *CoRR*, abs/1803.06386, <http://arxiv.org/abs/1803.06386>.
- Sivakumar, Guhan (2024) “HMM-RNN Fusion Economic Forecasting,” <https://github.com/gs1803/HMM-RNN-Fusion-Economic-Forecasting>, Code used for this paper.
- St. Louis Fed (1991) “Federal Reserve Economic Data,” <https://fred.stlouisfed.org>, Series Used: CPIAUCSL, UNRATE, DFF, DGS10, PPIACO, GDPC1, GDPIC1, UMCSENT, WTISPLC, USREC.
- Sundararajan, Mukund, Ankur Taly, and Qiqi Yan (2017) “Axiomatic Attribution for Deep Networks,” *CoRR*, abs/1703.01365, <https://arxiv.org/abs/1703.01365>.
- Thakur, Divyanshu (2018) “LSTM and its equations,” <https://medium.com/@divyanshu132/lstm-and-its-equations-5ee9246d04af>.

Thorbecke, Willem (2002) “A Dual Mandate for the Federal Reserve: The Pursuit of Price Stability and Full Employment,” *Eastern Economic Journal*, 28 (2), 255–268, <http://www.jstor.org/stable/40326100>.

Viterbi, A. (1967) “Error bounds for convolutional codes and an asymptotically optimum decoding algorithm,” *IEEE Transactions on Information Theory*, 13 (2), 260–269, [10.1109/TIT.1967.1054010](https://doi.org/10.1109/TIT.1967.1054010).

Yahoo Finance (2024) “S&P500 Historical Data,” <https://finance.yahoo.com/quote/%5EGSPC/history>.

Yang, Yucheng, Zhong Zheng, and Weinan E (2020) “Interpretable Neural Networks for Panel Data Analysis in Economics,” <https://arxiv.org/abs/2010.05311>.

Zhang, Junhuan, Jiaqi Wen, and Zhen Yang (2022) “China’s GDP forecasting using Long Short Term Memory Recurrent Neural Network and Hidden Markov Model,” *PLOS One*, 17 (6), 1–26, [10.1371/journal.pone.0269529](https://doi.org/10.1371/journal.pone.0269529).

Zheng, Yuanhang, Zeshui Xu, and Anran Xiao (2023) “Deep learning in economics: a systematic and critical review,” *Artificial Intelligence Review*, 56 (9), 9497–9539, [10.1007/s10462-022-10272-8](https://doi.org/10.1007/s10462-022-10272-8).

Appendix A: Tables

Variable	ADF-Statistic	P-Value	Used Lag	Nobs	1%	5%	10%	Stationary
inflation_rate	-2.9505	0.0398	15	631	-3.4408	-2.8661	-2.5692	Yes
unemployment_rate	-3.3240	0.0138	2	644	-3.4405	-2.8660	-2.5692	Yes
federal_funds_rate	-2.5860	0.0959	17	629	-3.4408	-2.8661	-2.5692	No
market_yield_10_rate	-1.4700	0.5483	12	634	-3.4407	-2.8661	-2.5692	No
ppi_rate	-5.1161	0.0000	15	631	-3.4408	-2.8661	-2.5692	Yes
gdp_pc1	-4.5632	0.0002	12	634	-3.4407	-2.8661	-2.5692	Yes
gdp_invest_pc1	-4.7711	0.0001	12	634	-3.4407	-2.8661	-2.5692	Yes
sp500_yoy	-4.9281	0.0000	13	633	-3.4407	-2.8661	-2.5692	Yes
consumer_senti_pc1	-5.4568	0.0000	16	630	-3.4408	-2.8661	-2.5692	Yes
oil_price_pc1	-5.3638	0.0000	12	634	-3.4407	-2.8661	-2.5692	Yes

Table A1: ADF Test Results

Original Data Model			Hidden States Data Model		
Layer (type)	Input Shape	Output Shape	Layer (type)	Input Shape	Output Shape
lstm_input	(None, 24, 9)	(None, 24, 9)	lstm_input	(None, 24, 10)	(None, 24, 10)
lstm (LSTM)	(None, 24, 9)	(None, 24, 50)	lstm (LSTM)	(None, 24, 10)	(None, 24, 50)
lstm (LSTM)	(None, 24, 50)	(None, 50)	lstm (LSTM)	(None, 24, 50)	(None, 50)
dense (Dense)	(None, 50)	(None, 25)	dense (Dense)	(None, 50)	(None, 25)
dense (Dense)	(None, 25)	(None, 1)	dense (Dense)	(None, 25)	(None, 1)

Means Data Model			All HMM Data Model		
Layer (type)	Input Shape	Output Shape	Layer (type)	Input Shape	Output Shape
lstm_input	(None, 24, 18)	(None, 24, 18)	lstm_input	(None, 24, 19)	(None, 24, 19)
lstm (LSTM)	(None, 24, 18)	(None, 24, 50)	lstm (LSTM)	(None, 24, 19)	(None, 24, 50)
lstm (LSTM)	(None, 24, 50)	(None, 50)	lstm (LSTM)	(None, 24, 50)	(None, 50)
dense (Dense)	(None, 50)	(None, 25)	dense (Dense)	(None, 50)	(None, 25)
dense (Dense)	(None, 25)	(None, 1)	dense (Dense)	(None, 25)	(None, 1)

1 Year Forecast Model			2.5 Year Forecast Model		
Layer (type)	Input Shape	Output Shape	Layer (type)	Input Shape	Output Shape
lstm_input	(None, 48, 19)	(None, 48, 19)	lstm_input	(None, 120, 19)	(None, 120, 19)
lstm (LSTM)	(None, 48, 19)	(None, 48, 50)	lstm (LSTM)	(None, 120, 19)	(None, 120, 50)
lstm (LSTM)	(None, 48, 50)	(None, 50)	lstm (LSTM)	(None, 120, 50)	(None, 50)
dense (Dense)	(None, 50)	(None, 25)	dense (Dense)	(None, 50)	(None, 25)
dense (Dense)	(None, 25)	(None, 12)	dense (Dense)	(None, 25)	(None, 30)

Table A2: LSTM Model Summaries

Appendix B: Figures

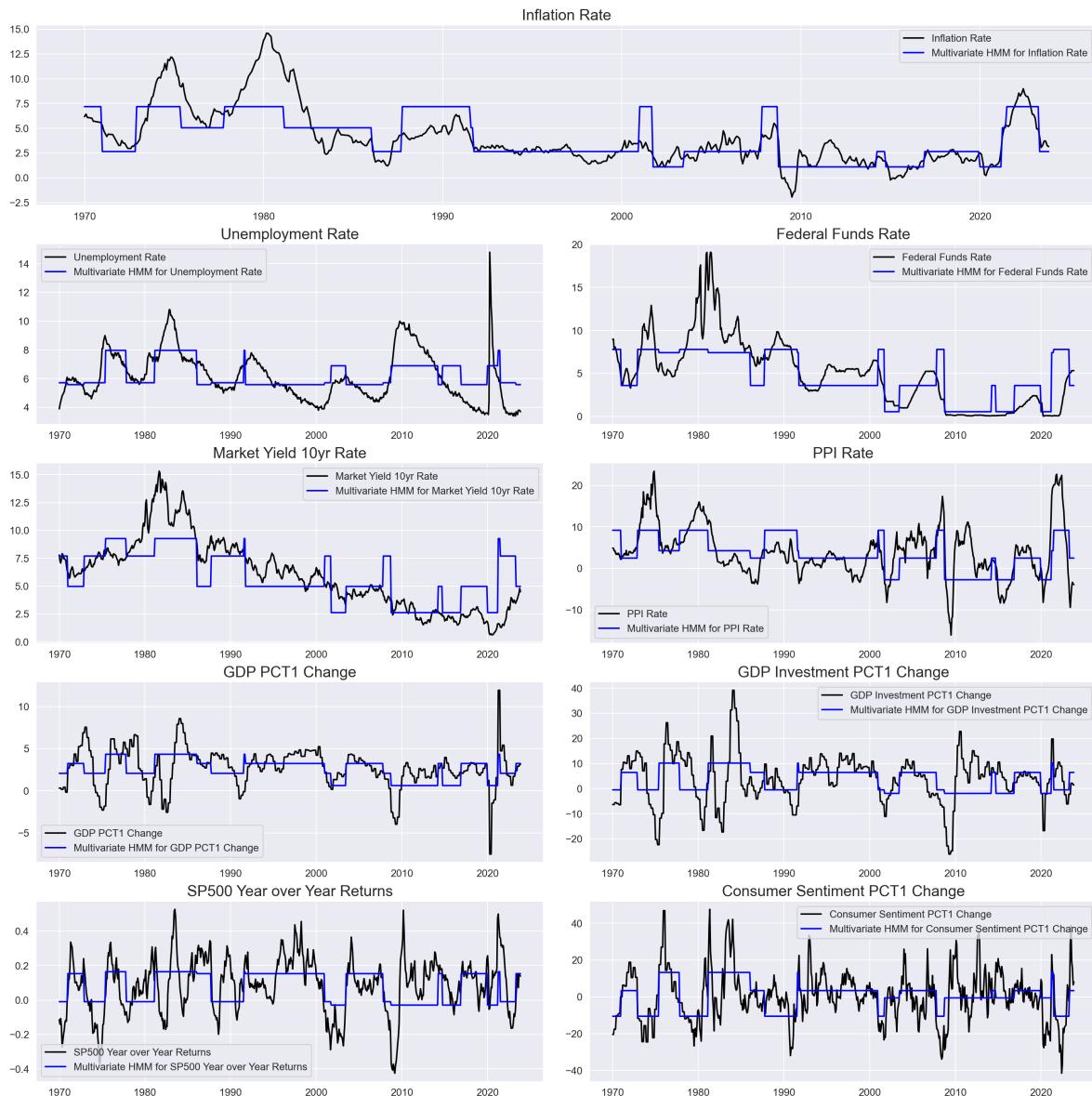


Figure B1: Averages Across Hidden States of the Features

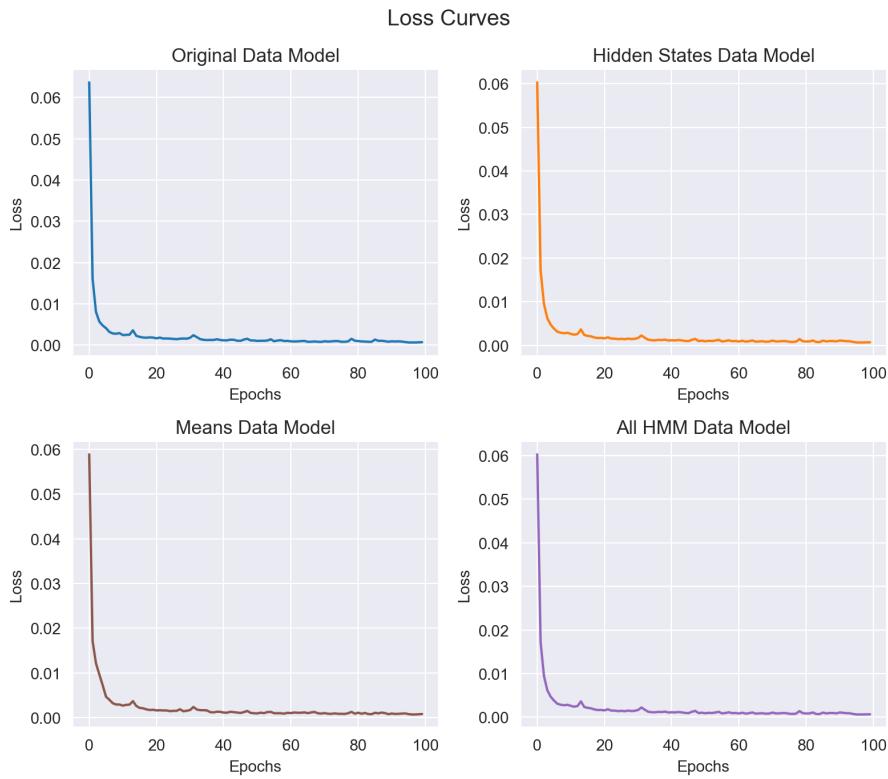


Figure B2: Training Loss Curves

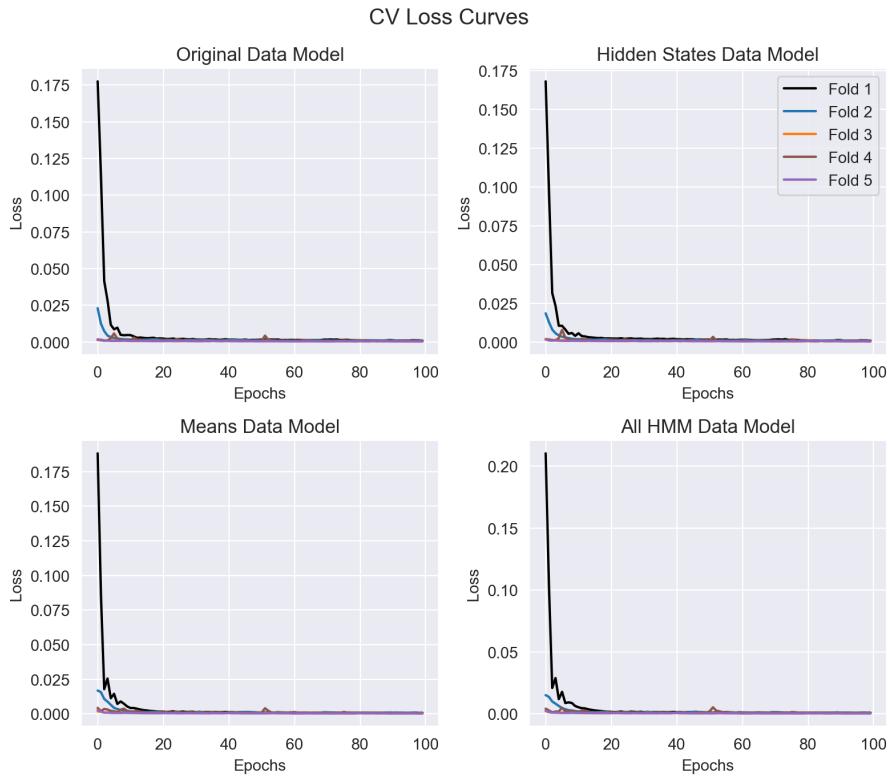


Figure B3: Training Loss Curves CV