Computational Statistics - Spring 2023
Project Report

Team Member Names: Ashish Dhiman, Yibei Hu

Project Title: Forecasting of commodity price structure using HMM

Contents

1 Introduction

Understanding the dynamics of future curves is crucial for traders and investors in the commodity markets. By analyzing the contango or backwardation of future contracts, they can gain insights into market expectation of future supply and demand conditions and adjust their trading strategies accordingly.

In this project, we're going to focus on forecasting the futures curve structure change of a key commodity, the Crude Oil Brent futures. The objective of this project is twofold. Firstly, we aim to use eigenvalue decomposition to identify primary components that can represent the price structure for Crude Oil Brent futures. [1]

Secondly, we plan to employ an out-of-sample HMM structure to anticipate future curve structure changes. By using the identified principal components as inputs, we will develop a predictive model that can generate trading signals. These signals will inform decision-making for traders seeking to navigate the increasingly volatile and unpredictable commodity market.

1.1 Futures

Future contracts are standardized agreements between two parties to buy or sell a specified asset at a predetermined price and date in the future. They are commonly used by producers and consumers of commodities to manage their exposure to price risk. By locking in a future price for a commodity, producers can protect themselves against price declines, while consumers can protect themselves against price increases.

1.2 Future curves: Contango and backwardation

The term "future curve" refers to the relationship between the prices of future contracts with different maturities. In some cases, the future curve is said to be "Contango", which means that future contracts with longer maturities trade at a higher price than those with shorter maturities. This situation can occur when the cost of storing the underlying commodity is high or when there is an expectation of future price increases. In Contango, the future price is higher than the spot price, and the future curve slopes upwards.[4]

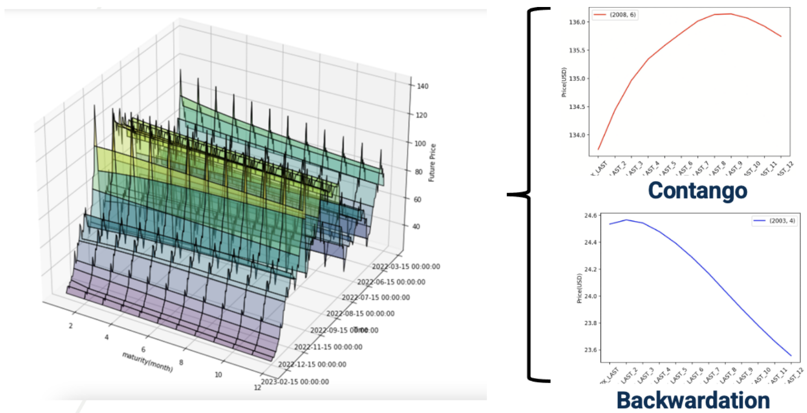

Figure 1: Crude Oil Brent shape

In other cases, the future curve is said to be in "backwardation", which means that future contracts with shorter maturities trade at a higher price than those with longer maturities. This situation can occur when there is a shortage of the underlying commodity or when there is an expectation of future price declines. In Backwardation, the future price is lower than the spot price, and the future curve slopes downwards.

In our study, we're trying to predict changes between the Contango and Backwardation in the Crude Oil Brent futures. We saw that Crude Oil is often in a Contango shape, especially when the economy is down. Also, Crude Oil Brent is constantly in changes of shape and slope as the picture shows:

2 Data Source

The objective of this project is to analyze the structural changes in the Crude Oil Brent futures trading market. To achieve this, we obtained the original dataset from Bloomberg Terminal by collecting the Crude Oil Brent futures trading data from 2003.04.11 to 2023.04.11, with tickers ranging from CO1 to CO12.

The dataset consists of daily futures prices maturing in 1 to 12 months, resulting in a dataset size of 5478 * 12. We chose this range to ensure that our analysis covers a wide range of futures contracts while also capturing the dynamics of the market over time.[3]

3 Methodology

This section details the four steps involved in predicting the transition between shapes of future curves: Contango and Backwardation.

1. **Step 1:** The first step involves calculating the difference $d_{t,i}$ in price p between two futures maturing at month (i) and (i+1) for every timestep t in our data.

$$d_{t,i} = p_{t,i+1} - p_{t,i}, \quad i \in 1, 2, \ldots, 11$$

The 11 price differences created, present high colinearity, with the lowest corelation between 1 month and 12 month maturity future.

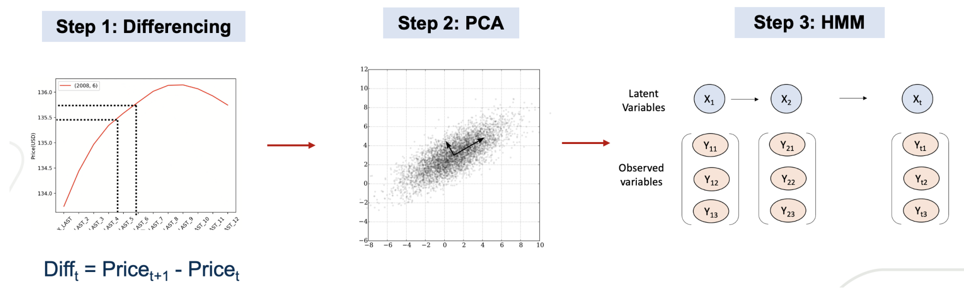

Figure 2: Approach Schematic

2. **Step 2:** Next, Principal Component Analysis (PCA) is used to reduce the dimensionality of the dataset from 11 differences to 2 components. This step helps to simplify the dataset and helps train the downstream HMM model faster.

$$\text{With the } d_{t,i} \text{ defined as above, let:} \tag{1}$$

$$X = \begin{bmatrix} d_{1,1} & \dots & d_{1,11} \\ \dots & \dots & \dots \\ d_{T,1} & \dots & d_{T,11} \end{bmatrix} \tag{2}$$

$$\text{Then PCA reduced data is} \tag{3}$$

$$Z_{T*2} = X_{T*11} \cdot V_{11*2} \tag{4}$$

$$\text{where V are eigenvectors of covariance matrix of X} \tag{5}$$

Evaluation of Eigen vectors, suggests that 1st PC is a diff between the 12 month future and the 1 month future. The second PC, on the other hand had positive weights for middle maturities while negative for the extremes.

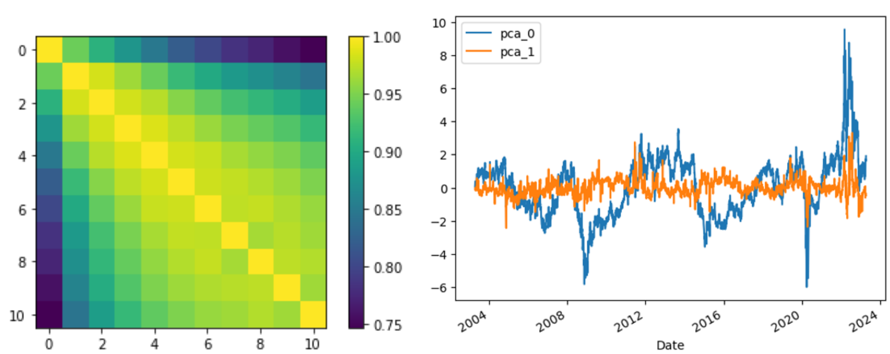

Figure 3: Time series of Z_t

3. **Step 3:** The third step involves applying a Hidden Markov Model (HMM) to the time series of these Principal components with 2 hidden states, namely Contango and Backwardation. Because our observed variable space is continous, we employ a HMM GMM model with 3 components, where emission probability is given by a mixture of 3 bivariate gaussians, whose parameters are governed by the hidden state.[2]

$$P(Z_t|S_t = j) = \sum_{m=1}^{3} \mathcal{N}(Z_t|\mu_j, \Sigma_j)$$

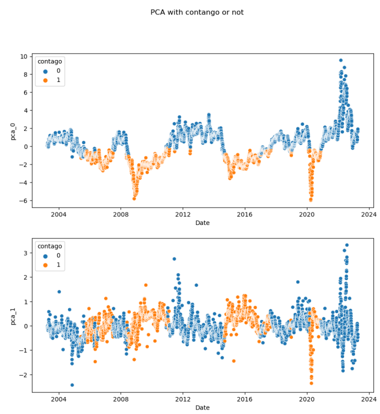

Figure 4: Time series of Z_t with contango status

HMM is a good choice to model the shape of future curve becuase as shown in the figure below, there are significant difference in the PC values depending if the curve is in Contango or not. Consequently the path of hidden states decoded from HMM has great overlap with the true tag of contango, as can bee seen from the KDE estimate below.

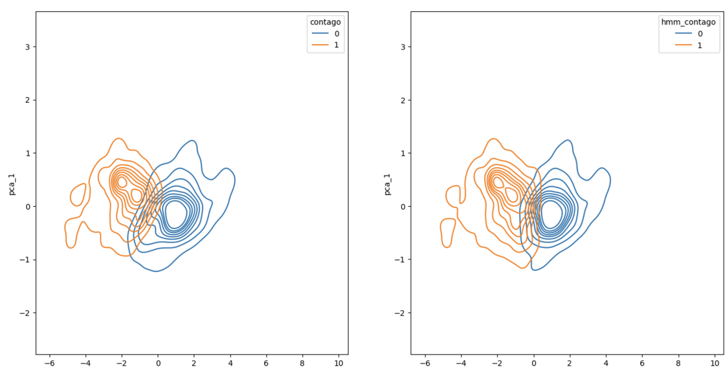

Figure 5: KDE of Z_t with True tag (left) vs HMM tag (right)

4. **Step 4:** Finally the HMM transition matrix A is used to predict the switch between the two hidden states, after k timesteps. Using the Markov property, A^k provides the transition probability between different states after k steps, we sum up the off-diagonal entries of this 2,2 matrix to get the estimate of shape change probabilities.

$$\text{Probability of shape change in next k days, } P_k = \sum_{i \neq j} A^k_{i,j} \tag{6}$$

$$\text{Probability of Contango start in next k days, } P_k = A^k_{1,0} \tag{7}$$

$$\text{Probability of Contango end in next k days, } P_k = A^k_{0,1} \tag{8}$$

$$\tag{9}$$

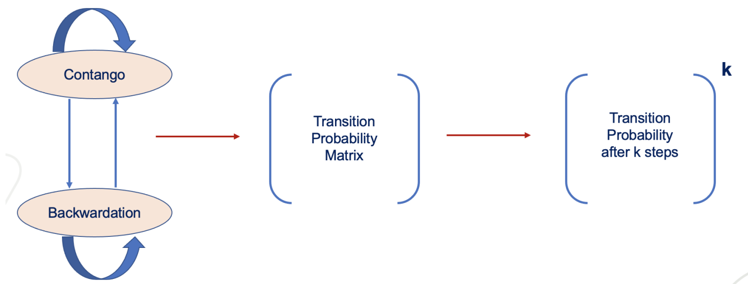

Figure 6: Transition Probablity from HMM parameters

4 Experiments and Results

We were able to generate a chart that displays the probability of transition over the next 10 days. The model provided significant signals of transition, indicated by a surge in probability near actual transition events. Notably, the model identified signals for several important events, such as the end of the prolonged impact of the 2008 financial crisis around 2010.5 and a possible indication of trading behavior before Covid hits before 2020.

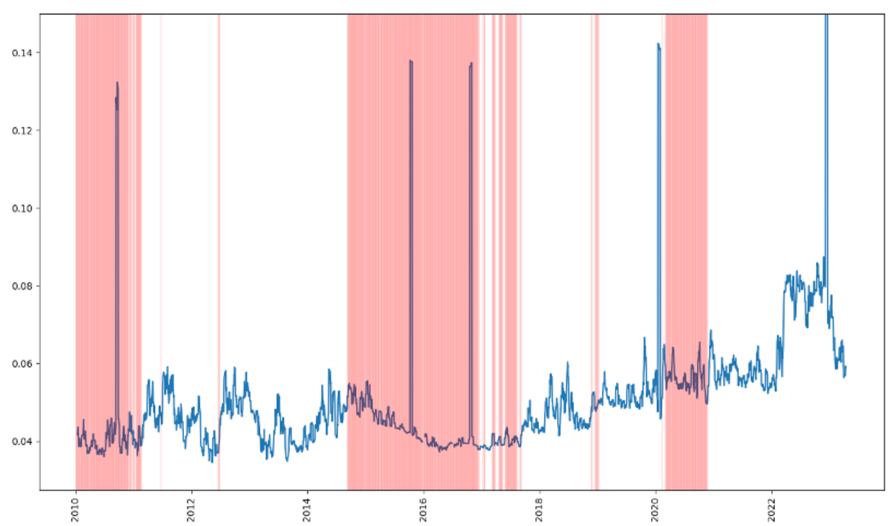

Figure 7: results

However, the model exhibits considerable noise due to a lack of data, resulting in false signals. After smoothing the noise with moving averages, we notice that the short term moving average and long term moving average of transition probabilities divert significantly near the transition points. Hence a more clean signal is created by comparing the 5-day and 150-day moving averages of the transition probabilities.

	Probability of Contango start in next 10 days	
contango_start_clean	*Moving Average 5 days*	*Moving Average 150 days*
0	2.45%	2.43%
1	**2.42%**	**2.19%**

Table 1: Moving Average of Start probabilities

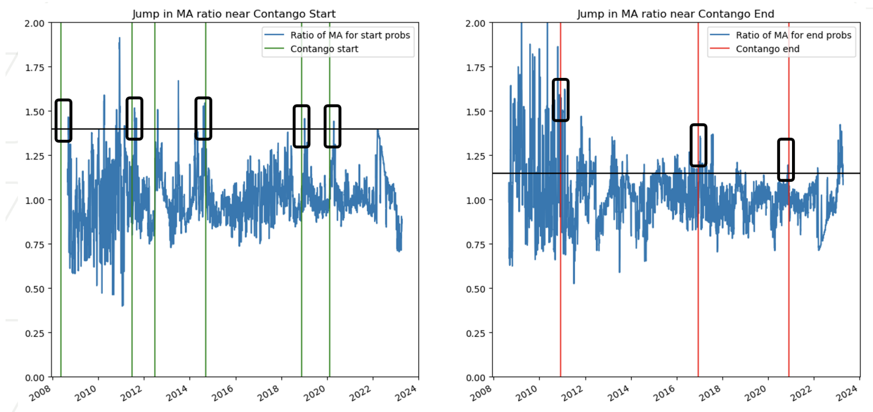

Figure 8: Time series of moving average ratios for Contango start probabilities (left) and Contango end probabilities (right). We can see that the ratio jumps above the threshold (horizontal black line) near the true start and end points.

	Probability of Contango end in next 10 days	
contango_end_clean	*Moving Average 5 days*	*Moving Average 150 days*
0	2.45%	2.42%
1	**2.04%**	**2.23%**

Table 2: Moving Avg of End Probabilities

Hence we are able to create the final signals, Contango-start and Contango-end using the above MA as

$$Contango - start = \mathcal{I}(\frac{\text{5 day MA start probability}}{\text{150 day MA start probability}} > 1.4)$$

$$Contango - end = \mathcal{I}(\frac{\text{150 day MA end probability}}{\text{5 day MA end probability}} > 1.15)$$

These predicted signals for the most part are able to capture the start and end of Contango states as shown in the image below.

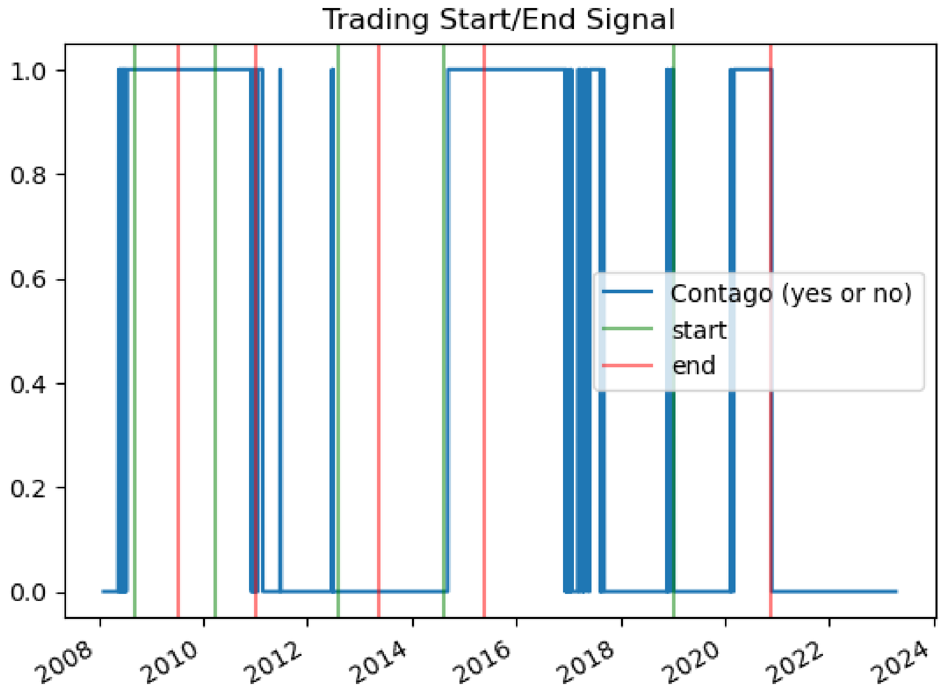

Figure 9: Predicted Start and End signal. Blue line oscillates between 1 and 0 and signifies if the future curve is in Contango (1) state or backward state(0). The green lines: predicted start points appear more or less near the start of contango while red: predicted end points near the end of contango.

5 Portfolio Result

In order to assess the practicality of our model for real-time trading, we applied it to historical data and created a trading portfolio. The portfolio was constructed as follows:

- 1. Since the signals generated by our model are forward-looking, when we detect a "Contango" signal, we take a long position in a short-term crude Brent oil contract with a maturity of one month, while simultaneously taking a short position in a long-term crude Brent oil contract with a maturity of twelve months.
- 2.When we detect a "Backwardation" signal, we take opposite positions as described above, i.e., a short position in a short-term contract and a long position in a long-term contract.
- 3.If we do not detect either of these signals, it indicates that the market is relatively stable and not undergoing a transition stage. Therefore, we close out our positions and calculate profits.

We also take into account situations where the same signals appear continuously. In these cases, we believe it may indicate a prolonged or more severe change in market conditions, so we increase our positions on a daily basis. Additionally, we consider the period from the appearance of the signal to its disappearance as one complete trade, and record the profit or loss accordingly. For simplicity, we're recording the difference in price of the long/ short position as the cost of the trade, instead of calculating for the real margin system. Also, we neglected all transaction cost or tax.

During the testing period of our model, we executed a total of 152 trades, out of which 97 trades were profitable, which indicates a winning rate of 63.8%. The cumulative profit generated from these trades is presented below. Additionally, we observed that the average duration of a complete trade cycle is approximately 3.89 days, with an average cost of $7.96. On average, each trade cycle yielded a profit of $0.41, indicating an average return on investment of approximately 5.08%.

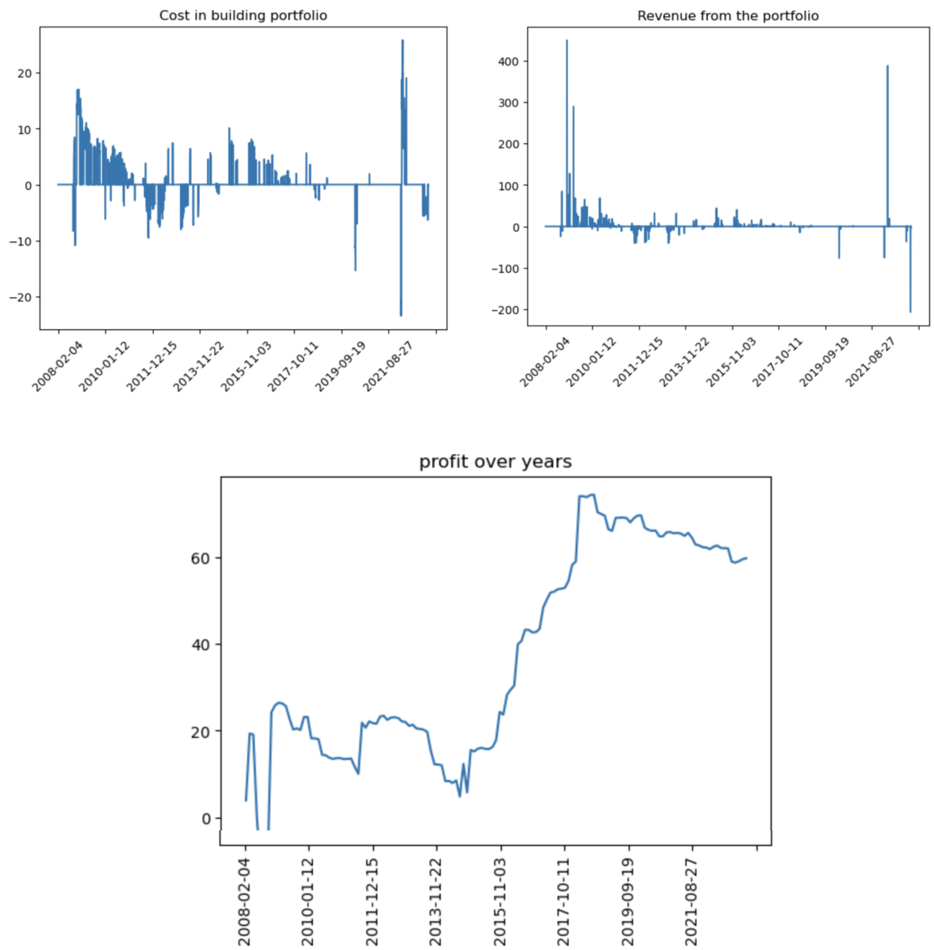

Figure 10: The profit table shows that our model is very stable, the maximum drawdown is shown around 2009, which is resulted from the financial crisis

6 Conclusion and Limitations

The model we developed can be applied in real time trading. According to the transition signal, we can construct a model that informs the possible transition of the futures curve shape. While the approach developed here exhibits good promise in analysing the shape of future curves, we can imporve in the future in following aspects:

- We have not factored the liquidity of futures in this exercise.
- Because the EM algorithm used to fit the HMM provides a local solution, we have used AIC to find the best model. However, we need to increase the number of trials and experiment with BIC.
- We don't have enough data for training. We might need to look into other commodities or for longer period of time.

References

[1] Sergiy Ladokhin and Svetlana Borovkova. Three-factor commodity forward curve model and its joint p and q dynamics. *Energy Economics*, 101:105418, 2021.

[2] Mark Stamp. A revealing introduction to hidden markov models. 2017.

[3] Heng Xiong. *Some Applications of Higher-Order Hidden Markov Models in the Exotic Commodity Markets*. PhD thesis, 2018. - Database copyright ProQuest LLC; ProQuest does not claim copyright in the individual underlying works;- 2023-03-06.

[4] Yingjian Zhang. Prediction of financial time series with hidden markov models. Master's thesis, School of Computing Science, Faculty of Applied Sciences, Simon Fraser University, May 2004. MSc thesis.