SML320 Research Project

Sreeta Basu

February 29 2024

1 Introduction

After the lockdowns enacted by governments worldwide in response to the COVID-19 pandemic, many people were not allowed to return to work for months, resulting in a global loss of productivity. Combined with unprecedented supply chain disruptions across industries due to travel restrictions, fears of a global recession were widespread. Supply shortages caused prices of household consumer items to surge. Governments supplied generous cash subsidies to households and firms to ensure that consumer spending did not fall, which induced additional demand-driven inflation. Central banks then implemented diverse monetary policies to tamp down inflation, among which were rapid increases in interest rates, which resulted in the high interest rates in 22 years in December 2023.

The heightened volatility and uncertainty in financial markets created a unique environment for strategic investment approaches. In particular, the U.S. dollar traded at the same value as the euro for the first time since 2002. To benefit from the strength of the U.S. dollar, institutional traders needed to short the dollar and carry the currency of another country. The carry trade is particularly sensitive to liquidity risk, market confidence and other macroeconomic factors, causing sporadic periods of high profitability strongly linked to the global business cycles. Because it relies on the future exchange rate between currencies that continuously fluctuates, the return of the carry trade is uncertain.

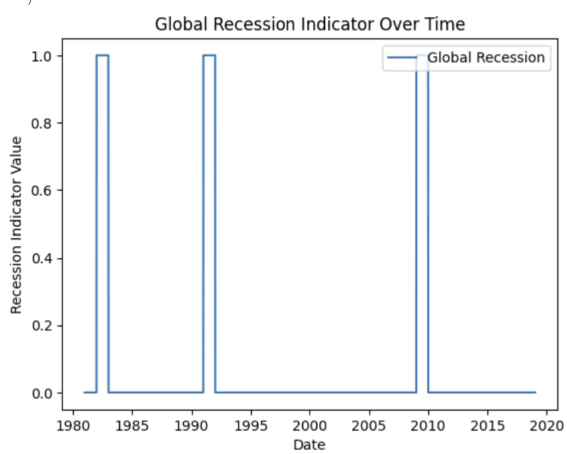
Regime-aware carry trade strategies, which involve adjusting investments based on prevailing economic conditions, have emerged as a profitable approach in this market. By incorporating a regime-aware approach, investors dynamically adapt their carry trades to capitalize on interest-rate differentials between currencies. This flexibility lets investors diversify their holdings and reduce risk in highly uncertain markets. During a time marked by heightened global tensions, adopting a regime-aware carry trade strategy presents a practical and strategic approach to navigate through the prevailing uncertainty.

In this paper, I will use Bayesian analysis to predict profitable carry trades, conditioned on market conditions in a two-step process, where I first use clustering techniques to detect regimes and optimize ideal investment and then use Bayesian updating to predict the current regime.

2 Data

We use spot currency exchange rates (to the U.S. dollar) from the following nations: Australia, Canada, China, Denmark, Hong Kong, India, Japan, Malaysia, New Zealand, Norway, Singapore, South Africa, Sri Lanka, Sweden, Switzerland, Thailand and United Kingdom from the Foreign Exchange Rates tabulated by the Unites States Federal Reserve. The dataset used begins 01/02/1981 and ends 12/28/2019. The prior is based on data until 1/2/2018 and the posterior uses the remaining data as 'new' information.

The global recessions, as defined by the IMF, occured during the following dates: 1975-01-01 to 1975-12-31, 1982-01-01 to 1982-12-31, 1991-01-01 to 1991-12-31, 2009-01-01 to 2009-12-31.



3 Exploratory Data Analysis

We use two approaches to regime clustering: k-means and gaussian mixture model. We choose to use 4 clusters to maintain consistency through the analysis. Below, we include the results of the k-means clustering, using Euclidean distance as the distance metric.

Cluster	**Number of Data Points**	**Center**
1	2032	-0.1577
2	5953	0.0006
3	1929	0.1900
4	2	-0.4545

Table 1: Number of Data Points in Each Cluster and Cluster Centers

The model has a training accuracy of 0.9208, where the 'true' model is the global recessions defined by the IMF. However, we notice immediately the lack

		Predicted	
		Positive	Negative
Actual	Positive	9131	783
	Negative	2	0

Table 2: Confusion Matrix

of true negatives, which implies that the model has low precision. Thus, we search for another clustering mechanism, namely the Gaussian Mixture Model.

Below, we tabulate the results of the Gaussian Mixture Model.

Cluster	**Number of Data Points**	**Center**
1	4664	-0.045
2	3479	-0.0124
3	1057	0.0046
4	716	0.0184

Table 3: Number of Data Points in Each Cluster and Cluster Centers

		Predicted	
		Positive	Negative
Actual	Positive	4664	3479
	Negative	514	202

Table 4: Confusion Matrix

The model has a lower overall accuracy of 0.8896, but has a much better precision of 0.2580.

4 Bayesian Conjugate Prior

In this section, we investigate the individual statistics of Malaysia, whose economy is highly dependent on the U.S. economy. This makes it a great sample case for the regime hypothesis. First, we include some graphics to depict the clusters found. We choose to graph Malaysia against Switzerland, which is relatively indifferent to the U.S. economy because the economies do not overlap much in sectors or consumer demographics.

Now, we tabulate the results of each cluster derived from k-means. We run a t-test to determine if the means are non-zero and reject the null hypothesis with 95% confidence for 3 of the 4 clusters. This implies that the returns are statistically signficant in these regimes. The lower and upper bounds are the bounds of a frequentist 95% confidence interval.

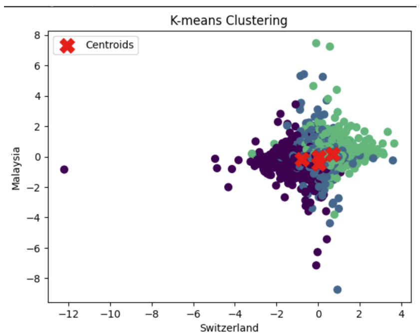


Figure 1: K-means Clusters

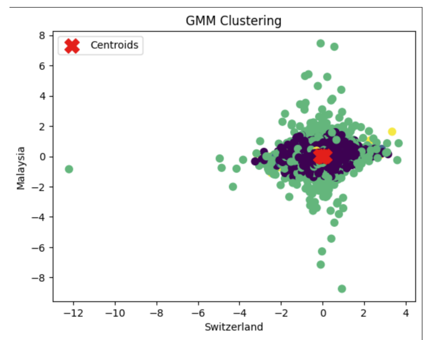


Figure 2: GMM Clusters

Figure 3: Clustering Mechanisms

At the end of this paper, we include graphs of the prior, likelihood and posterior of a normal-normal conjugate pair for each of the four clusters.

We also tabulate the results from the Gaussian Mixture model on page 6. As before, the prior refers to all data before 2018 and the posterior is formed using the remainder of the data. The T-statistic and p-value refer to a null hypothesis test where the null hypothesis is that the cluster mean is zero, and we reject this hypothesis with confidence 0.95. The lower and upper bounds are the bounds of a 95% frequentist interval.

5 Future Directions

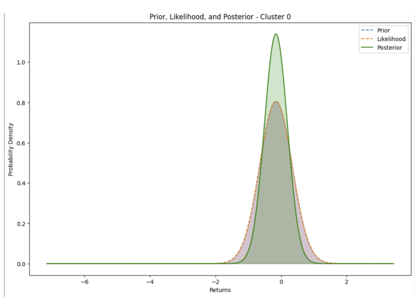
I would like to increase the size of the dataset, including more countries as well as trade-based indices for commodities like gold and oil. I will also redo my test-train split so that I can backtest my results with data I have not trained on, which was a great insight from my presentation.

I would also like to train a Gaussian Mixture Variational auto-encoder, which allows for non-linear transformations in determining clusters. This offers greater adaptability in comprehending emerging recessions and states, thereby facilitating more dynamic adjustments to new regimes.

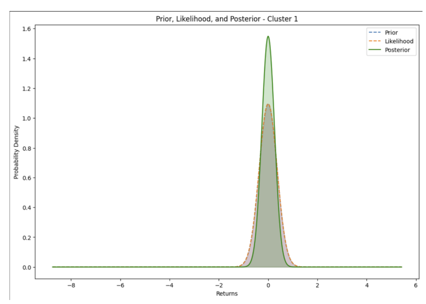
I'd also like to work with Bayesian neural networks to compare the results from different methods of modeling this data.

Parameter	**Cluster 0**	**Cluster 1**	**Cluster 2**	**Cluster 3**
Prior Mean	-0.1594	0.0007	0.1902	-0.4545
Prior Std Dev	0.4954	0.3643	0.4862	0.6428
Posterior Mean	-0.1595	0.0007	0.1902	-0.4545
Posterior Std Dev	0.3502	0.2575	0.3438	0.4545
T-statistic	-23.6116	-21.0328	8.1505	-1.2200
P-value	4.23×10^{-109}	8.57×10^{-95}	6.45×10^{-16}	0.4371
Prior Probability	0.6997	0.6074	0.4264	0.8058
Posterior Probability	0.7706	0.6500	0.3965	0.8888
Prior Odds	3.3598	1.5471	0.7433	4.1504
Posterior Odds	3.3598	1.8574	0.6570	7.9902
Bayes Factor	1.4417	1.2006	0.8838	1.9252
Lower Bound	-0.1810	-0.0082	0.1684	-0.9091
Upper Bound	-0.1388	0.0103	0.2144	0.0000

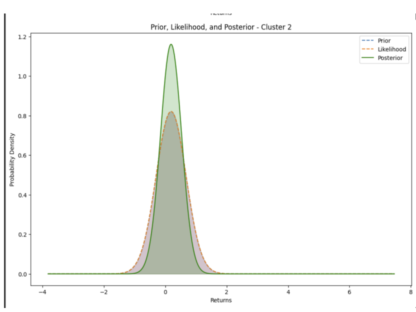
Table 5: Summary of Results for Each K-means Cluster



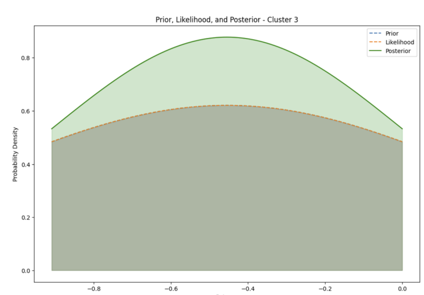
(a) Prior and Posterior for Cluster 0



(b) Prior and Posterior for Cluster 1



(c) Prior and Posterior for Cluster 2

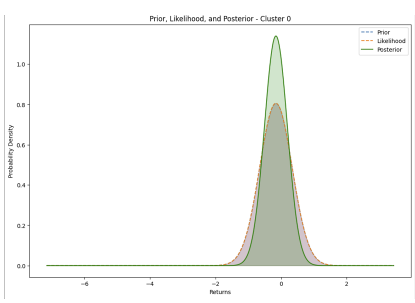


(d) Prior and Posterior for Cluster 4

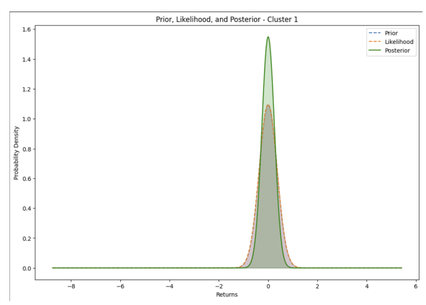
Figure 4: Prior and Posterior Distributions For Each Cluster

Parameter	Cluster 0	Cluster 1	Cluster 2	Cluster 3
Results by Cluster:				
Prior Mean	0.00298	-0.00038	0.02348	0.01379
Prior Std Dev	0.32034	0.11717	1.10093	0.28582
Posterior Mean	0.00292	-0.00038	0.02348	0.01379
Posterior Std Dev	0.22649	0.08285	0.77847	0.20211
T-statistic	-20.71	-50.53	-2.26	-8.07
P-value	3.01×10^{-91}	0.0	0.024	2.96×10^{-15}
Prior Probability	0.619	0.804	0.528	0.619
Posterior Probability	0.666	0.887	0.539	0.665
Prior Odds	1.625	4.107	1.117	1.621
Posterior Odds	1.993	7.862	1.170	1.986
Bayes Factor	1.227	1.914	1.047	1.225
Lower Bound	-0.00605	-0.00426	-0.04199	-0.00639
Upper Bound	0.01292	0.00335	0.08831	0.03479

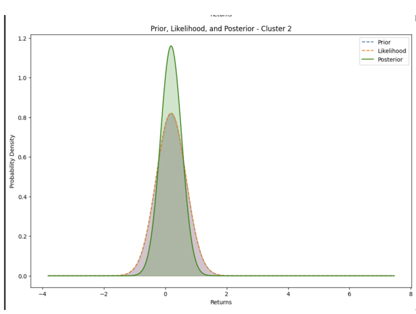
Table 6: Results by Cluster



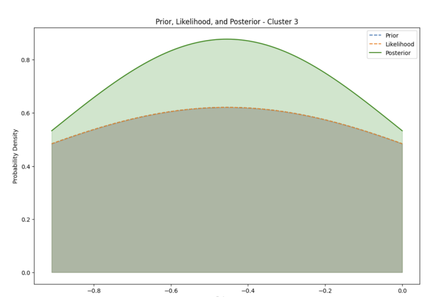
(a) Prior and Posterior for Cluster 0



(b) Prior and Posterior for Cluster 1



(c) Prior and Posterior for Cluster 2



(d) Prior and Posterior for Cluster 4

Figure 5: Prior and Posterior Distributions For Each Cluster