**Practice of QCF – Pre-Midterm Presentation Update**

**Group Members**

Johnathan Crist, Walker Hills, Walter Mitchell, Wyatt Nechtman

1. **any update to the main idea of the project**

The scope of our project has been more solidified as we have progressed through the semester. The basis for our analysis is still focused on the paper by Chava, Hsu and Zeng titled Does History Repeat Itself? Business Cycle and Industry Returns, Journal of Monetary Economics. Whereas that paper seeks to understand the difference in equity sector returns based on economic conditions, we looked to analyze the difference in factor performance. We will take the Fama-French five factor model and look at the factor performance across different market cycles. Using a model to predict what part of the economic cycle we are in, we will then create a trading strategy based on taking either long or short positions in the factors.

1. **what data did you use in the analysis so far**

**Fama-French Factors**

We started with the standard Fama-French 5-factor return plus momentum from French’s website and look at their cyclicality and dispersion as well as their correlation with one another. Below is the correlation table and the descriptive statistic of their returns. You can see a lot of negative correlations among the different factors. Also, it seems that the market and momentum have the highest arithmetic average returns but also the largest ranges, min, and max.



|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Mkt-RF** | **SMB** | **HML** | **RMW** | **CMA** | **Mom** |
| **count** | 708 | 708 | 708 | 708 | 708 | 708 |
| **mean** | 1.69% | 0.70% | 0.94% | 0.85% | 0.88% | 1.91% |
| **std** | 0.079 | 0.056 | 0.059 | 0.043 | 0.040 | 0.074 |
| **min** | -31.0% | -22.7% | -22.4% | -29.6% | -11.0% | -49.3% |
| **25%** | -2.5% | -3.0% | -2.6% | -1.3% | -1.6% | -1.2% |
| **50%** | 2.4% | 0.1% | 0.8% | 0.7% | 0.5% | 1.9% |
| **75%** | 6.6% | 4.3% | 3.8% | 2.8% | 3.1% | 5.6% |
| **max** | 26.3% | 30.3% | 27.8% | 27.1% | 18.8% | 36.4% |

**US Economic Indicators**

We also have looked at the following list of indicators when fitting our clustering model.

Graphical user interface, text, email

Description automatically generated

1. **what models did you use in the analysis so far**

**Business Cycle using GDP/CPI**

The first business cycle model we tried to recreate was a model using GDP and inflation to determine where we were in the cycle as follows:

* Recovery (1): Decreasing Inflation, Increasing GDP
* Expansion (2): Increasing Inflation, Increasing GDP
* Slowdown (3): Increasing Inflation, Decreasing GDP
* Contraction (4): Decreasing Inflation, Decreasing GDP

**Business Cycle from High Yield Spreads**

We then looked at using some indicator that was not as lagged and more frequently available than quarterly, so we implemented a model that looks at the high yield spreads to determine where in the cycle we currently are.

* Recovery (1): HY Spread > Median and Decreasing
* Expansion (2): HY Spread < Median and Decreasing
* Slowdown (3): HY Spread < Median and Increasing
* Contraction (4): HY Spread > Median and Increasing

We have looked at a combination of the first two and adding rules to the model based on if NBER has the economy in a recession or not but this again would need to be lagged for any trading strategy.

**Clustering**

We have performed an initial K-Means clustering of technical, economic, valuation and sentiment indicators (USEQ features) over a period from 2015-present. We are trying to see if we can develop some labels from the time series data that will allow us to develop a classification problem where we can apply other ML models to predict the labels.

For the K-Means algorithm we used a Dynamic Time Warping (DTW) Metric to cluster the time series data and evaluated it using the silhouette score. Using a DTW metric will be useful later if we encounter unequal length time series for similar features. This was chosen over the typical Euclidean metric since it ignores the dimension of time that is part of the data. On our initial attempt to cluster we utilized all the indicators in our feature set. The average silhouette score between the clusters was around 0.226, which we think can be improved upon if we more carefully select the features for clustering and implement further algorithms.

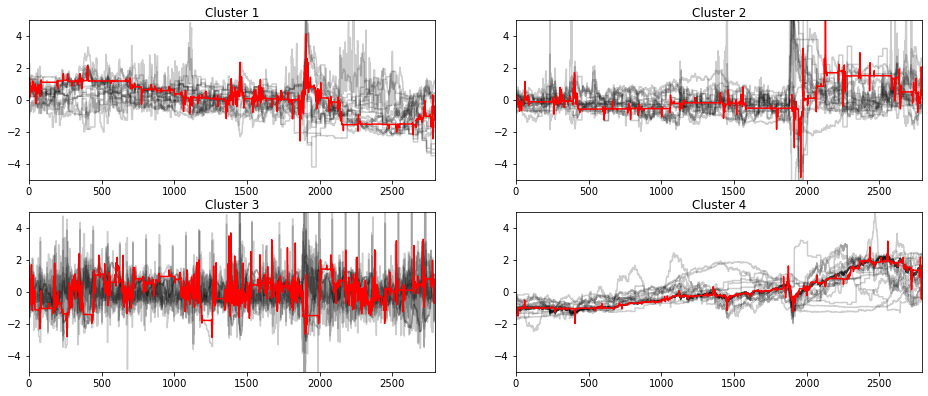


Figure 1: Time Series for each cluster (black) and the barycenter (red)

1. **what are the results and insights that you derived from the project so far**

The first insight is there seems to be sufficient dispersion in factor performance. The opportunity for the strategy is dependent on the volatility of factors and persistence of performance in different market conditions. The low correlation of factors also presents the opportunity to rotate and create long/short strategies between factors. The economic models also are currently not robust enough. Using economic data like GDP and CPI do not seem strong enough due to their lag in reporting and quarterly frequency.

* 1. As explained in the challenges section, our models are not yet robust enough to have the “seasonality” desired for business cycles to occur in sequence. We discuss more appropriate approaches to solve this problem in the next section as well.

1. **what are the challenges that you have experienced with data or analysis or modeling**

The GDP/CPI model just does not seem to be very informative given that GDP data is so infrequent and late. It is also revised all the time so it would not be a good thing to base our model off, if we are trying to trade on it.

Our High Yield Model does not have that much data given that it only goes back to the late 1990s. This leads to fear that it may be too biased of a model going forward.

**More Thoughts on Clustering**

In a prior project, we used unsupervised learning to cluster yield curve data into both 2 and 4 clusters, thus generating labels for us to use for supervised learning tasks (with a goal of predicting the shape that we labeled). Going into this project, we had a similar idea, that we could just perform the same unsupervised learning task to generate labels. However, this project differs in a few ways.

First, when generating labels for a business cycle, they are sequential in nature. While we can generate labels for yield curves in a not-so-robust manor, ideally our labels should follow a sequence of recovery, expansion, slowdown and contraction. Therefore, we may need to look else ware to generate out business cycle predictions.

Second, our goal in this work is not just to predict the business cycles, we also want to identify what factors are overperforming in each labeled cycle. Therefore, our task is two-fold: not only correctly identify the business cycle, but also the factor performance must be somewhat consistent in each cycle. For example, say we generate an arbitrary label 1 that we think the economy is in. We hypothesize that factor *x* is going to outperform factor *y* in this business cycle, therefore that must be the case in every situation where the economy is in cycle labeled 1. If this is not the case, then there is no alpha to capture by this method.

We have begun to experiment with time series clustering and other time series based analyses. Following the work of Gomez-Cram, we might be able to find a baseline label. His work showed that in the ~six months following a recession, returns are predictably negative.

Along with this, we have an idea of possibly backing out the business cycles by analyzing factor performance across time. If we can identify consistent periods of factor outperformance, we may be able to aggregate these, generating labels for our business cycle forecasting model, allowing us to train it directly on the labeled factor performances. For example, we identify timeframes where factor *x* is performing well and factor *y* is not. We can aggregate the data based on when this holds true, generating a label for the business cycle (whatever it may be). If we can robustly do this for four time periods, as we are following the hypothesis that there are four regimes in the business cycle, we could back out the business cycle from the factor performances.

1. **what is the road map / next steps that you intend to take for the rest of the term**

Roadmap:

* 1. Sector Rotation Paper & Factor Data Analysis
     1. Based on the work done here we want to apply a similar method that rotates through different factors as opposed to sectors.
     2. Potentially look at the addition of other factors if the 5 factor is not sufficient
  2. Art of Creating Business Cycle Model
     1. Can we develop some sort of clustering model that allows us to label periods of the business cycle for label classification.
     2. Once we have labelled data are we able to accurately predict a business cycle given the performance of certain factors during a cycle time horizon.
  3. Trading optimization
     1. Once we have a model that can predict business cycles, we want to use these predictions to optimize the factor rotation given the current business cycle.
     2. Formulating the proper strategy will include trying to design an optimal trading strategy based on maximizing the Sharpe Ratio. We will test both in sample and out of sample strategies to ensure there isn’t over fitting of the trading model.