**Practice of QCF – Project Topic Update**

**New format**

**Group Members:**

Johnathan Crist, Walker Hills, Walter Mitchell, Wyatt Nechtman

(1) any update to the main idea of the project

(2) what data did you use in the analysis so far

(3) what models did you use in the analysis so far

K-Means clustering for features using a time series analysis approach for defining the clusters.

(4) what are the results and insights that you derived from the project so far

(5) what are the challenges that you have experienced with data or analysis or modeling

(6) what is the road map / next steps that you intend to take for the rest of the term

**Old format**

**1. Group Members**

Johnathan Crist, Walker Hills, Walter Mitchell, Wyatt Nechtman

**2. Main idea of the project**

**Project Scope:** the objective is to analyze the cross section return of equity markets and factors dependent on the changing business cycles. Factor returns that have demonstrated statistical significance in the past do not produce steady and consistent returns (i.e. the underperformance of the value factor over the last decade). The project will seek to determine if evaluating macroeconomic variables and the business cycle can be used to create a factor timing strategy. The initial scope of the project will be to evaluate the factors provided in the Fama & French Five-Factor model but can grow in scope to evaluate other known factors. The Chava, Hsu and Zeng paper, Does History Repeat Itself, and the Verdad Capital paper, Countercyclical Asset Allocation, offer potential ways to measure the economic cycle and macroeconomic signals. Other signals and models will also be tested to attempt to create a more robust measurement of the economic cycle. Our expectation is to be able to find some historical relationship between factors and the business cycle and a viable strategy that would profit off of sector rotation. The Invesco strategy suggests that there has been historical alpha in this type of strategy. We would expect that most of the alpha would come in highly volatile markets when there is a large discrepancy in the returns of cyclical and defensive assets.

**3. Do you plan to follow any papers listed or that are not listed, background research**

* **Chava, Hsu and Zeng, Does History Repeat Itself? Business Cycle and Industry Returns, Journal of Monetary Economics**
* Fama, Eugene F, and Kenneth R French. 2015. “A Five-Factor Asset Pricing Model.” Journal of Financial Economics 116 (1): 1–22.
* Invesco Russell 1000 Dynamic Multifactor Fund (<https://www.invesco.com/us/financial-products/etfs/product-detail?audienceType=Investor&ticker=OMFL>)

**4. The modeling and programming choices that you are making**

We plan to model the business cycle with a combination of GDP Growth when it is above or below the trend and what direction it is going and similarly with the high yield spread. We will also look at adding inflation into this model to see if we can get a more accurate picture of what regime of the business cycle we are currently in.

There are a variety of methods we can use in order to determine the current cycle. Ideally, we take some sort of labeled data and a basket of lagged indicators and train a classifier to give us probabilities of which cycle we are in or going to be in. For instance, if we take data like the High Yield Spread (BAMLH0A0HYM2), we can generate labels of 4 business cycles: spread above 10-year median and 3-month return is positive, spread below 10-year median and 3-month return is positive, etc. We can now use those labels to train OneVsRest classifiers with our other data to give probabilities that we are in certain cycles. Another option is via unsupervised learning. We can take a basket of indicators and reduce dimensionality with PCA, use clustering algorithms to find 4 clusters (or optimal amount), and use each cluster as a label. We have data from another project using this to cluster yield curve into shapes and would do something similar.

After we have labeled business cycles, we can group factors based on cycle and compare performance. We would also have the ability to explore other abstract ideas. For example, we could generate more than 5 factors and then cluster the business cycle data into 5 clusters and then see which performs the best. There are several ways we can explore ideas with ML models in this part beyond the initial scope

**5. What data are you planning to use and the feasibility**

* For Modeling the business cycle, we are using public GDP, Inflation, and High Yield Spreads from Fred. Everything will be publicly available and have sufficient history.
* For original factor analysis we will be using fama-french factor return data to see how we can best optimize this return given which business cycle our model predicts.
* For improving our conviction with machine learning we are currently looking about 60 public indices/yields/indicators/etc. May add more depending on how exactly the models start to fit with more testing.

**6. Expected outcome and the main insight you expect to derive from the project**

* The Invesco model has shown an active return on their Dynamic Multifactor ETF when compared to their benchmark (Russell 1000).
* Expect varying results depending on the period we choose to implement the back test while hopefully providing some small alpha.
* Ability to identify business cycles based on our model, which will allow us to better weigh our factors.
* Perhaps a better understanding into the Fama-French factors and their relationship with each other and the business cycle.

**7. Any descriptive stats of the data. Any initial results?**

The initial pass through the data shows a large dispersion of factor performance. This high level of dispersion creates opportunity to trade either long or short and create value. The factors are also lowly correlated and even lean negatively correlated. This also demonstrates the ability to create value through a trading strategy that seeks to time the relative performance of these factors. The true value add in this analysis will be creating a robust enough model to predict when different factors will perform well, but the initial look at the data demonstrates that there is sufficient volatility and dispersion for performance to be generated. The next step is building out the predictive element of the model, starting with a similar analysis to the paper, “Does History Repeat itself?”. The main deviation will be looking at factor performance not sector performance. Once this has been replicated, we will seek to improve the economic cycle identification through the methods listed in section 4.

