**Submission Element**

1. **Title and Abstract**
2. **Introduction**

* 开头点题：傅立叶（亮点）说明有周期性，42个月主导
* 识别算法：PS（低频），高盛（中频），未来20days平均收益（高频）
* 预测模型：计量，ML，HMM

1. **Data Selection【统一数据的内容、格式、时间范围和对应period、含义等】**

* 数据提取：Russell 3000, Macro Data, 技术量价因子等
* 数据清理：差分环比同比处理等
* Corresponding variable list做表 (name, explanation, time period, frequency)

1. **Statistical Analysis & 模型假设**
   * 描述性统计：每年或者每个月的波动箱形图，每年一条线的季节性趋势对比图
   * 测试：所有变量的ADF结果图，正态分布，共线性，vif？
   * Granger Causality：热力图（因果检验？）
   * KS Test & Spearman Correlation（Correlation plots of each feature）
   * OLS Results（系数和显著性）
   * Auto Regression（PACF&ACF）and Lags（自相关，异方差？）
   * 通过计量方法确定所用因子，而后传给机器学习
2. **Proposed Methodology / Models**

* Regime Recognition
  + 傅里叶变换分析Russell频谱，得到周期和相位，从而证明指数存在周期性，周期决定了宏观上的牛熊状态（前提：平稳和正态检验？）；多个指数画图，说明周期存在，算出参数和三角函数图？
  + LT算法, PS算法识别：42个月作为参数，放到识别算法里，得到label
  + 20day平均收益？高盛标签？
* Regime Prediction
  + 1. 计量模型（inference model）：
    - ARX-GARCH, VAR, Ridge, Logit
    - 通过计量方法确定所用因子，而后传给机器学习
  + 2. ML模型：运用label和features进行训练和预测（2018之前作为训练数据；2018-2021作为测试集并执行策略，train和test的结果都放对比？）
    - ML, DL, RL
    - PCA&LDA, KNN, DTR, RFR, GBR, SVC
    - LSTM？
  + 3. HMM模型：return作为x，三个regime的预测，回测收益
  + Implication / Interpretation结果分析
  + Extra Feature Description

1. **Investment Strategy and Results**
   * 第一步：每个模型后面跟一个简单的、用于对比模型之间收益差别的回测图（图内三个策略对比：buy and hold, buy and short, buy and 3m T-bill）
   * 第二步：可以给予前面的模型，构建复合/复杂的高收益策略（例如，信号叠加）
2. **Conclusion (Brief)**
3. **References**
4. **Appendix**

其他：

一页500字，9页内容，1页摘要和引用

如果out of market，用三个月的T-bill rate.

考虑罗素指数的自相关，将过去的价格/收益作为ML的features？

Future model improve：即时计算feature？

图片包含 图形用户界面

描述已自动生成

Russel3000 按照行业划分后 看各个行业 景气 比重？

表格

描述已自动生成

https://www.bloomberg.com/quote/RAY:IND

需要成分股？指数价格是根据成交还是个股

动量？自回归？lag？

市场状态分类权威

手机屏幕截图

描述已自动生成

表格

描述已自动生成

表格

描述已自动生成

表格

描述已自动生成

图表

中度可信度描述已自动生成文本, 信件

描述已自动生成图片包含 文本

描述已自动生成图形用户界面, 文本, 应用程序, 电子邮件

描述已自动生成文本

描述已自动生成

表格

描述已自动生成手机屏幕截图

描述已自动生成图形用户界面, 文本, 应用程序, 电子邮件

描述已自动生成

文本

描述已自动生成

文本

描述已自动生成

文本, 信件

描述已自动生成表格

描述已自动生成