Conlan Scientific: Financial ML
Progress Report

DS6011 Capstone Project
Alex Cathcart, Ellen Yu and Jamie Oh

University of Virginia

Completed so far

- Defined scope of the project

- Set up recurring meetings with the sponsor

- Drafted and signed team contract

- Set up recurring team meetings

- Set up Git repository

- Set up Kanban

- Set up Rivanna (Thanks, Jeremy!)

Completed so far (cont'd)

- Completed foundational literature review
  - *Advances in Financial Machine Learning* by Marcos Lopez De Prado
  - *Algorithmic Trading with Python* by Chris Conlan
- Received the data
- Created SQL database
- Received metadata and data walkthrough from sponsor

Items in progress

- **Review the data**
  - Metadata, documentation for fundamentals
  - EDA and Data Validation

- **Literature Review - ongoing**
  - *Fast Python* by Chris Conlan

- **Sponsor will conduct code review**

The Data

- Two primary sets - funds and prices

  - Funds - fund information, returns, revenue, cash flow etc

  - Prices - price of stocks, at start of business, close of business, high/low etc

- Derivative set - Daily

- Industry and metadata steps to map other sets

```
prices.shape, funds.shape, daily.shape, assets.shape

((12055241, 8), (133240, 109), (8352196, 9), (6297, 24))
```

Next Steps

- **Have finite research questions**
  - Lock down objective statement
  - Determine which Bayesian models we will be implementing

- **Feature engineering**

Questions?