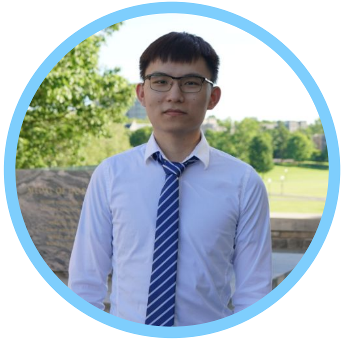
SPINNAKER ANALYTICS

Trading Signals Prediction

Business Advisors: Manish Gupta, Xi Zhao
Faculty Advisor: Professor Peter Wysocki
Team 14 Members: Ziqi Shan, Man Shi, Kangjing Shi, Tzuhua(Agnes) Huang

BU Questrom School of Business

Our Team



Kangjing Shi

Ziqi Shan

Man Shi

Agnes Huang

Agenda

- Introduction
- Hypotheses
- Methods & Models
- Overview of process
- Key findings
- Details
- Conclusion

Introduction

Predict portfolios' appreciation or depreciation and identify 'tradable signals' in financial markets for the investors.

Dataset

Weekly data spans 10 years from 2006 through end-Jan 2017

3 Fund Types

Institutional mutual fund, ETF, retail mutual fund.

20 Asset Classes

20 sectors in each market, including technology, healthcare, financials etc.

4 Features

Funds-flow data including Flow, FlowPct, AssetsEnd, PortfolioChangePct

Hypothesis

#1 %Portfolio Change could be a predictive outcome for trading signals.

#2 Prediction for each sector may be captured by different methods/models.

#3 Explanatory and response variables of ETF and institutional mutual fund are interdependent.

#4 Advanced models are expected to enhance model predictability.

Methods & Models

#1 VAR (Vector Autoregression) Model

- Capture the relationship between multiple time series variables.
- Comprise one equation per response variable in the system.

#2 Long Short-Term Memory (LSTM) Networks

- Store previous information and use it for processing the current input.
- Able to learn long-term order dependencies.

Overview of Process

VAR Model 1 → **VAR Model 2** → **VAR Model 3** → **LSTM**

Cross-market without MA(4)

Cross-market with MA(4)

Cross-market with MA(4) & market indices variable

Cross-market with MA(4) variable

Evaluate Model Predictability

- Return(t) - Return(t-1) > 0.1 → 1: Increase in portfolio return
- Return(t) - Return(t-1) < - 0.1 → -1: Decrease in portfolio return
- -0.1 ≤ Return(t) - Return(t-1) ≤ 0.1 → 0: Unchanged

Key Findings

01

Converting actual values into classification labels helps identify tradable signals.

02

Applying moving average removes volatility and increases prediction accuracy.

03

Adding market index variables improves model predictability.

04

Compared to LSTM, VAR models demonstrate desirable prediction in several sectors.

More Details

- VAR M1: Cross-market
- VAR M2: Cross-market MA(4)
- VAR M3: Cross-market MA(4) & market index
- LSTM Network

		Consumer Goods	Commo-dities	Energy	Finance	Health/ Biotech	Industrials	Large Cap Blend	Large Cap Growth	Large Cap Value
VAR M1	P_e	0.6	0.4	0.7	0.4	0.7	0.5	0.6	0.3	0.5
	P_i	0.2	0.7	0.8	0.5	0.5	0.5	0.7	0.5	0.5
VAR M2	P_e	0.3	0.6	0.6	0.4	0.9	0.5	0.5	0.6	0.4
	P_i	0.2	0.4	0.5	0.7	0.4	0.6	0.3	0.4	0.4
VAR M3	P_e	0.4	0.3	0.6	0.3	0.8	0.7	0.6	0.6	0.4
	P_i	0.5	0.4	0.3	0.6	0.5	0.6	0.5	0.6	0.4
LSTM	P_e	0.5	0.4	0.6	0.7	0.6	0.6	0.5	0.6	0.4
	P_i	0.5	0.4	0.6	0.7	0.7	0.6	0.6	0.5	0.5

P_e: %Portofolio Change in ETF P_i: %Portofolio Change in Institutional Mutual Fund

More Details

■ VAR M1: Cross-market ■ VAR M2: Cross-market MA(4)

■ VAR M3: Cross-market MA(4) & market index ■ LSTM Network

		Consumer Goods	Commo-dities	Energy	Finance	Health/ Biotech	Industrials	Large Cap Blend	Large Cap Growth	Large Cap Value
VAR M1	P_e	0.6	0.4	0.7	0.4	0.7	0.5	0.6	0.3	0.5
	P_i	0.2	0.7	0.8	0.5	0.5	0.5	0.7	0.5	0.5
VAR M2	P_e	0.3	0.6	0.6	0.4	0.9	0.5	0.5	0.6	0.4
	P_i	0.2	0.4	0.5	0.7	0.4	0.6	0.3	0.4	0.4
VAR M3	P_e	0.4	0.3	0.6	0.3	0.8	0.7	0.6	0.6	0.4
	P_i	0.5	0.4	0.3	0.6	0.5	0.6	0.5	0.6	0.4
LSTM	P_e	0.5	0.4	0.6	0.7	0.6	0.6	0.5	0.6	0.4
	P_i	0.5	0.4	0.6	0.7	0.7	0.6	0.6	0.5	0.5

P_e: %Portofolio Change in ETF P_i: %Portofolio Change in Institutional Mutual Fund

More Details (cont.)

- VAR M1: Cross-market
- VAR M2: Cross-market MA(4)
- VAR M3: Cross-market MA(4) & market index
- LSTM Network

		Mid Cap Blend	Mid Cap Growth	Mid Cap Value	Real Estate	Small Cap Blend	Small Cap Growth	Small Cap Value	Techno-logy	Telecom	Utility
VAR M1	P_e	0.4	0.6	0.5	0.8	0.4	0.6	0.4	0.6	0.6	0.5
	P_i	0.3	0.3	0.3	0.3	0.5	0.2	0.4	0.5	0.7	0.5
VAR M2	P_e	0.7	0.5	0.6	0.4	0.5	0.5	0.7	0.2	0.4	0.5
	P_i	0.5	0.3	0.5	0.3	0.6	0.5	0.4	0.3	0.3	0.6
VAR M3	P_e	0.4	0.4	0.8	0.4	0.7	0.6	0.7	0.4	0.4	0.4
	P_i	0.6	0.3	0.5	0.3	0.7	0.5	0.4	0.2	0.3	0.6
LSTM	P_e	0.5	0.4	0.7	0.4	0.5	0.6	0.5	0.5	0.5	0.6
	P_i	0.6	0.6	0.6	0.5	0.5	0.6	0.6	0.5	0.5	0.6

P_e: %Portofolio Change in ETF P_i: %Portofolio Change in Institutional Mutual Fund

Key Results

“

#1

- Simple model (VAR) may perform better than the advanced one (LSTM).
- Predictions with VAR perform well in several sectors’ portfolio changes.

#2

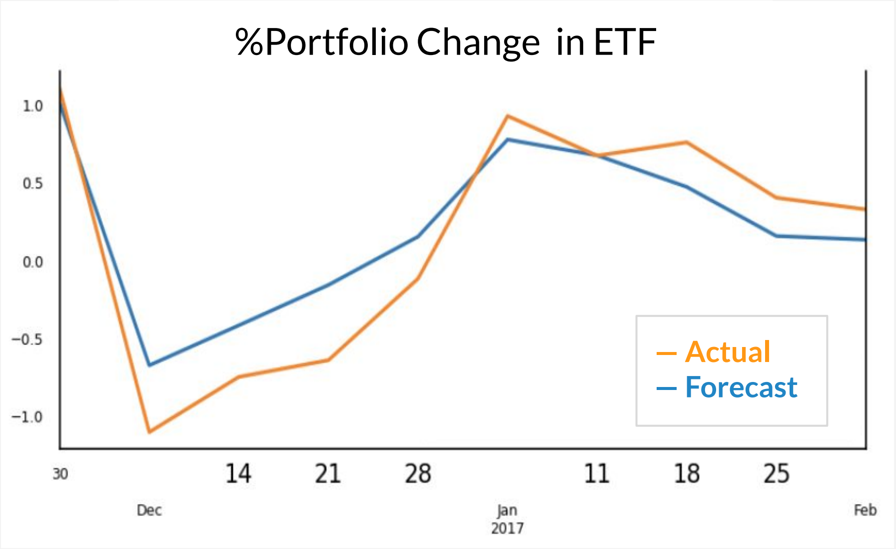
- The three VAR models show different performance on different sectors (higher accuracy: 0.8/0.9, lower accuracy: 0.2/0.3)

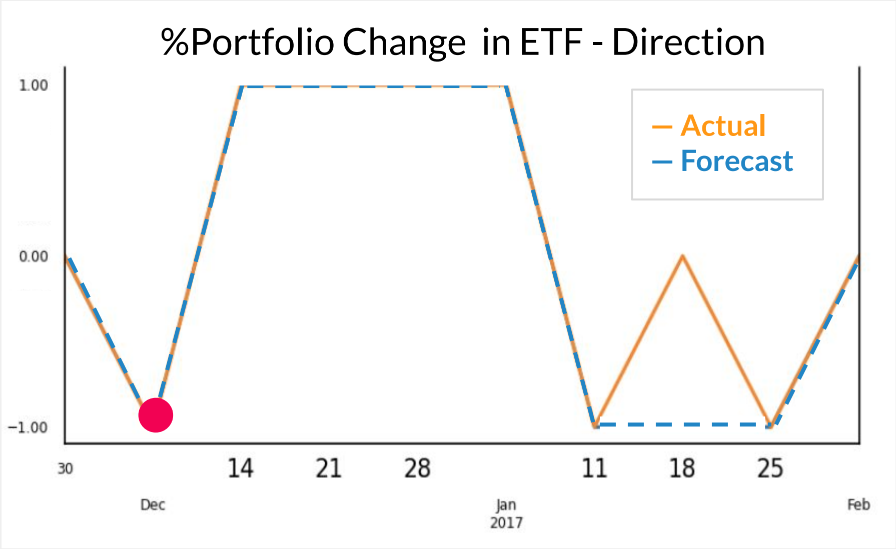
#3

- LSTMs demonstrate balanced accuracy across all the sectors (accuracy rate: 0.4-0.7).

Identify Tradable Signal

Model 2: Cross-market MA(4) - Health/Biotech





Questrom School of Business

Thanks!

Any questions?

Appendix

5 features

- **Flow:** amount of inflow and outflow in Millions of USD
- **FlowPct:** flow as percent of assets at beginning of the week
- **AssetsEnd:** assets at end of the week in Millions of USD
- **PortfolioChangePct:** percent change in overall portfolio during the week
- **ClosePct**: percentage change in close price of 4 US major indices

20 asset classes

- Consumer Goods, Energy, Financials, Health Care, Industrials, Technology, Telecom and Utilities Industries
- Combinations of large, mid and small companies in different phases like value, growth and blend .

Appendix

VAR model (vector autoregressive model):

- A statistical **model** used to capture the relationship between multiple quantities as they change over time. **VAR** is a type of stochastic process **model**. **VAR models** generalize the single-variable (univariate) autoregressive **model** by allowing for multivariate time series.

Cross-markets:

- P(ETF, k, t) = a + b1*F(ETF, k, t-1) + b2*F(instit, k, t-1) + b3*P(instit, k, t-1) + b4*P(ETF, k, t-1)+...+ e

Cross-markets with indices:

- v = related market index (SP 500, Dow Jones, Nasdaq, Russell 2000): ClosePCT
- P(ETF, k, t) = a + b1*F(ETF, k, t-1) + b2*F(instit, k, t-1) + b3*P(instit, k, t-1) + b4*P(ETF, k, t-1)+...+v + e

VAR Model Equations

- %Portfolio_etf(t) = a + b1* %Portfolio_etf(t-1) + b2* %Flow_etf(t-1) + b3* %Portfolio_ins(t-1) + b4* %Flow_ins(t-1)
- %Portfolio_ins(t) = a + b1* %Portfolio_etf(t-1) + b2* %Flow_etf(t-1) + b3* %Portfolio_ins(t-1) + b4* %Flow_ins(t-1)

Suitable Models In ETF Market			
VAR M1	Energy	Health	Real estate
VAR M2	Health	Mid Cap Blend	Small Cap value
VAR M3	Heath	Industrials	Mid Cap Value
	Small Cap Blend	Small Cap Value	
LSTM	Finance	Mid Cap Value	

Suitable Models In Institutional Market		
VAR M1	Commodities	Energy
	Telecom	Large Cap Blend
VAR M2	Finance	
VAR M3	Small Cap Blend	
LSTM	Finance	Health

Highest predictive accuracy Models for each sector

sector	Consumer Goods	Commo-dities	Energy	Finance	Health/ Biotech	Industrials	Large Cap Blend	Large Cap Growth	Large Cap Value
model	VAR M1	VAR M1	VAR M1	VAR M2 LSTM	VAR M2	VAR M3	VAR M1	VAR M2 VAR M3 LSTM	VAR M1

sector	Mid Cap Blend	Mid Cap Growth	Mid Cap Value	Real Estate	Small Cap Blend	Small Cap Growth	Small Cap Value	Techno-logy	Telecom	Utility
model	VAR M2	LSTM	VAR M3	VAR M1	VAR M3	LSTM	VAR M2 VAR M3	VAR M1	VAR M1	LSTM