Trading Volume Alpha

Russ Goyenko *McGill*
Bryan Kelly *Yale*
Toby Moskowitz *Yale*
Yinan Su *Hopkins*
Chao Zhang *HKUST-GZ*

EDHEC Speaker Series, October 28, 2025

Motivation

Research in asset pricing primarily focuses on predicting return moments
Challenges:

- diminishing value in return prediction

Motivation

Research in asset pricing primarily focuses on predicting return moments
Challenges:

- diminishing value in return prediction

Predicting non-return variables

- beyond moments of returns:

  predicting non-return variables for asset pricing (fundamentals, volume, etc.)

  how do we assess their value?
- beyond prediction:

  connecting non-return variables to economics

  better investment decisions

Trading volume

- trading volume highly predictable (by ML)
- connect volume to trading costs through "participation rate"
- translate volume prediction into an economic cost/benefit
  - volume prediction can be as valuable as return prediction

Economically important frontier in portfolio research

Volume in after-cost portfolio optimization

- trading costs matter critically but are hard to observe and are investor-specific
- trading volume: widely available data

$$PriceImpact_{a,i,t} \propto ParticipationRate_{a,i,t} := \frac{\$Traded_{a,i,t}}{\$Volume_{i,t}}$$

- volume forecast ⇒ trading cost optimization
- considering trading costs, a key trade-off:
  - cost of trading vs. (opportunity) cost of not trading
    volume ↑ ⇒ price impact ↓ ⇒ trade more aggressively (vice versa)
  - better volume forecast ⇒ better implementation

Volume in after-cost portfolio optimization

- trading costs matter critically but are hard to observe and are investor-specific
- trading volume: widely available data

$$PriceImpact_{a,i,t} \propto ParticipationRate_{a,i,t} := \frac{\$Traded_{a,i,t}}{\$Volume_{i,t}}$$

- volume forecast ⇒ trading cost optimization
- considering trading costs, a key trade-off:

- cost of trading vs. (opportunity) cost of not trading
  volume ↑ ⇒ price impact ↓ ⇒ trade more aggressively (vice versa)
- better volume forecast ⇒ better implementation

* Importantly, *not* trying to estimate best tcost model!

- purposely simple

Volume in after-cost portfolio optimization

- trading costs matter critically but are hard to observe and are investor-specific
- trading volume: widely available data

$$PriceImpact_{a,i,t} \propto ParticipationRate_{a,i,t} := \frac{\$Traded_{a,i,t}}{\$Volume_{i,t}}$$

- volume forecast ⇒ trading cost optimization
- considering trading costs, a key trade-off:
  - cost of trading vs. (opportunity) cost of not trading
    volume ↑ ⇒ price impact ↓ ⇒ trade more aggressively (vice versa)
  - better volume forecast ⇒ better implementation

* Importantly, *not* trying to estimate best tcost model!

- purposely simple

forecasting (neural networks) + finance modeling (transfer learning)

→ **net-of-cost "alpha"**

Side note: Generic tcost model

- Modeling tcosts generically is challenging
- Tcosts are trade and investor-specific
- Require proprietary data

⇒ *Forecasting* costs has received little (if any) attention

- Focusing on volume, while simple, is also general to any tcost model
- Provides a simple, but generic net-of-cost portfolio solution
- This could be a starting point, . . . more to do

Today's Talk

→ Volume prediction from statistical perspective

new dataset, neural network methods, and new benchmarks

– Portfolio strategy

transfer statistical forecasts into finance decision making

– Performance evaluation

substantial economic impact, comparable to return prediction

Daily stock-level dollar trading volume

Data structure

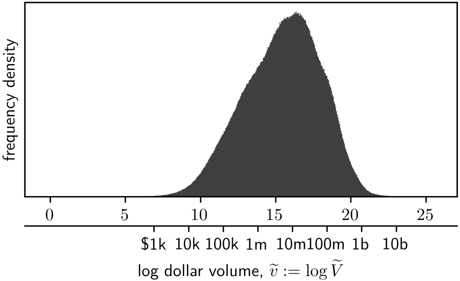
- unit of observation: stock, day (i, t panel)
- sample period from 2018 to 2022, 5 years or 1,200+ days
  3 year train + 2 year testing
  4,700+ stocks in total, avg 3,500 observed stocks per day,
  4,400,000+ observations in total

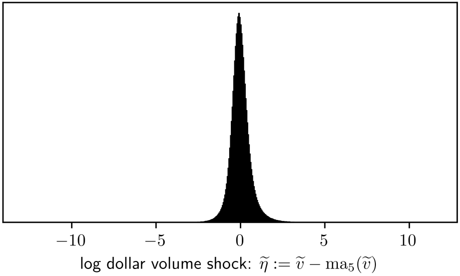
Predictors

Predictors, $X_{i,t}$, 175 in total, 4 categories:

1. technical signals
   return and volume lagged moving averages
2. firm characteristics
   size, b/m, ..., developed for return prediction (JKP dataset)
3. deterministic calendar events
   triple witching, index rebalancing, early closing etc.
4. scheduled earnings announcements

Dollar volume, log dollar volume, and its shock





- log dollar volume is persistent, like prices
  5-day moving average predicts log dollar volume with $R^2 = 93.7\%$
  - higher than one-day lag, moving avg 22, or moving avg 252
- take ma_5 as baseline throughout the paper
  predict log dollar volume shock $\widetilde{\eta} := \widetilde{v} - \mathrm{ma}_5$
  - like predicting return := log price - log yesterday's price.

Prediction results, OOS R^2 (%)

cumulative # of predictors	tech 8	fund-1 14	fund-2 161	calendar 165	earnings 175
A: OOS R^2 (%)					
ma_5	0				
ols	12.09	12.26	12.27	14.85	15.99
nn	14.31	14.90	14.42	17.13	18.45
rnn	15.80	16.25	15.47	18.12	19.86
B: number of parameters					
ma_5	0				
ols	9	15	162	166	176
nn	961	1,153	5,857	5,985	6,305
rnn	6,049	6,817	25,633	26,145	27,425

Prediction results in size groups

	all	nano	micro	small	large	mega
R^2 (%)						
ols_{all}	15.99	13.32	12.60	20.90	25.49	26.16
nn_{all}	18.45	15.80	14.86	23.71	27.76	29.12
rnn_{all}	19.86	16.63	16.14	26.00	30.50	32.02

Today's Talk

- Volume prediction from statistical perspective

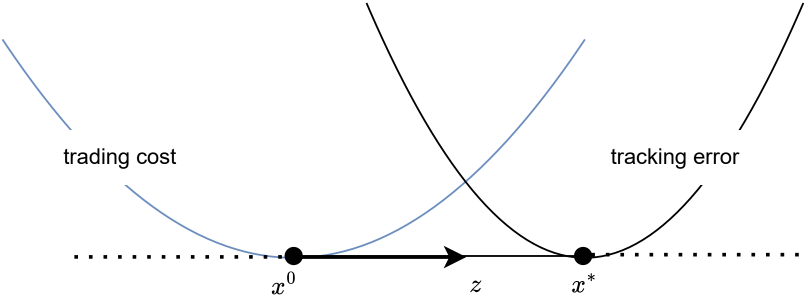
→ Portfolio strategy based on volume prediction

- Investment performance evaluation

Portfolio optimization problem

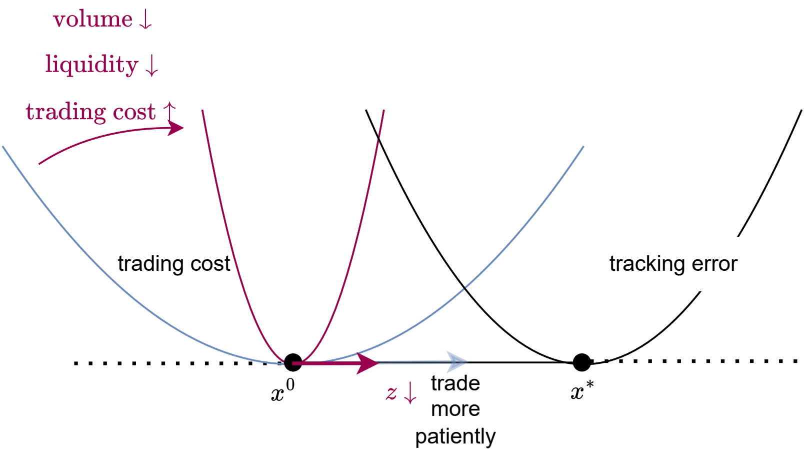
Trade-off: track a target portfolio vs. avoid trading cost:

$$\min_{\{x_{i,t}\}} \sum_{i,t} (TrackError_{i,t} + TradeCost_{i,t})$$



- choose $z \in [0, 1]$, the ratio between x^0 and x^* (stay put vs. trade all the way)

Portfolio optimization problem



- choose $z \in [0, 1]$, the ratio between x^0 and x^*
- low volume ⇒ high trading cost ⇒ trade less aggressively

Tracking error modeling

$$TrackError := \frac{1}{2}\mu(x^* - x)^2$$

- μ controls the penalty for being away from the target
  larger $\mu \Rightarrow$ trade more aggressively in general
- Microfoundation: a mean-variance optimization:
  - x^* embeds return signals and AUM
  - $\mu \propto$ risk aversion / AUM
- Reframing as a tracking error problem avoids the well-trodden ground and usual pitfalls of return prediction, which we do not want confounded with volume prediction

Trading cost modeling

$$TradeCost := \frac{1}{2}\widetilde{\lambda}(x - x^0)^2$$
$$\widetilde{\lambda} := 0.2/\widetilde{V}$$

Microfoundation: $PriceImpact$ linear in participation rate:

$$PriceImpact = 0.1\frac{x - x^0}{\widetilde{V}}, \quad TradeCost = PriceImpact \cdot (x - x^0)$$

E.g., buying 10% of daily volume incurs 1% trading cost
(cf. Frazzini, Israel, and Moskowitz)

Combined objective

Minimize the expectation of $TradeCost + TrackError$ conditional on information

$$\min_x \mathbb{E}\left[\frac{1}{2}\widetilde{\lambda}(x - x^0)^2 + \frac{1}{2}\mu(x^* - x)^2 \middle| \text{conditioning information}\right]$$

Importantly, $TradeCost$ is unknown when choosing x: $\widetilde{\lambda}$ depends on $\widetilde{V}$.

Combined objective

Minimize the expectation of $TradeCost + TrackError$ conditional on information

$$\min_{x} \mathbb{E}\left[\frac{1}{2}\widetilde{\lambda}(x - x^0)^2 + \frac{1}{2}\mu(x^* - x)^2 \middle| \text{conditioning information}\right]$$

Importantly, $TradeCost$ is unknown when choosing x: $\widetilde{\lambda}$ depends on $\widetilde{V}$.

Key point: predicting $\widetilde{V}$ well helps solving this problem.
Intuition: more adaptive to liquidity condition:
trade more aggressively if believe $\widetilde{V}$ high, patiently otherwise

Next: analyze and solve the problem to demonstrate this point and show the economic gains

Objective function analysis

Isolate target trade size:

Define trading rate $z := \frac{x-x^0}{x^*-x^0}$, then the target becomes

$$\begin{aligned}&\frac{1}{2}\widetilde{\lambda}(x-x^0)^2+\frac{1}{2}\mu(x^*-x)^2\\=&\frac{1}{2}\left(x^*-x^0\right)^2\big(\underbrace{\widetilde{\lambda}z^2+\mu(1-z)^2}_{loss^{\mathrm{econ}}(z)}\big)\end{aligned}$$

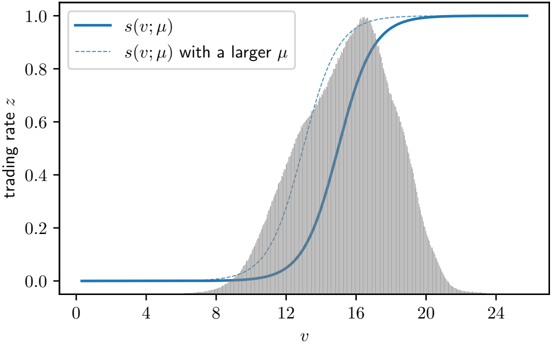
New problem: minimize expected economic loss by choosing z

$$\min_z \mathbb{E}\left[loss^{\mathrm{econ}}(z,\widetilde{v};\mu)|X\right]$$

- same problem regardless of tracking target (will experiment with different targets)
- choose trade rate $z \in [0,1]$

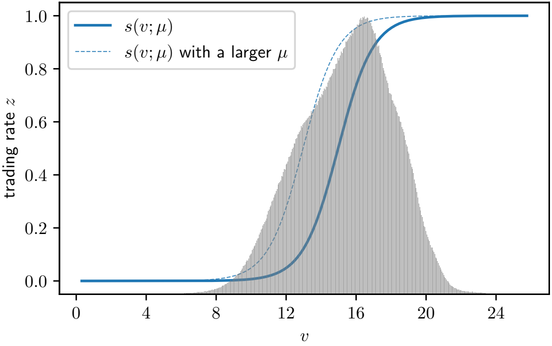
Optimal trading decision if volume is known

- say we believe $\widetilde{v} = v$, optimal policy: $s(v; \mu)$
  high $v \Rightarrow z \to 1$; low $v \Rightarrow z \to 0$



Optimal trading decision if volume is known

- say we believe $\widetilde{v} = v$, optimal policy: $s(v;\mu)$
  high $v \Rightarrow z \to 1$; low $v \Rightarrow z \to 0$



- ...but we do not know v
  how to turn $\widehat{v}$ into z? i.e. turn forecast into decision?

Approach 1, statistical learning plug-in stats forecast $\widehat{v}$, trade $\widehat{z} = s(\widehat{v}; \mu)$

1. ML based on a statistical loss, MSE:

$$\min_{v(\cdot)} \sum_{i,t \in \text{train}} loss^{\text{least squares}}\big(\widetilde{v}_{i,t}, v(X_{i,t})\big)$$

2. plug $\widehat{v}_{i,t} = v(X_{i,t})$ into the optimal policy $s(\,\cdot\,; \mu)$
   choose trading rate $\widehat{z}_{i,t} = s(\widehat{v}_{i,t}; \mu)$

Approach 1, statistical learning
plug-in stats forecast $\widehat{v}$, trade $\widehat{z} = s(\widehat{v}; \mu)$

1. ML based on a statistical loss, MSE:

$$\min_{v(\cdot)} \sum_{i,t \in \text{train}} loss^{\text{least squares}}\left(\widetilde{v}_{i,t}, v(X_{i,t})\right)$$

2. plug $\widehat{v}_{i,t} = v(X_{i,t})$ into the optimal policy $s(\ \cdot\ ; \mu)$
   choose trading rate $\widehat{z}_{i,t} = s(\widehat{v}_{i,t}; \mu)$

- Evaluated by economic loss
  portfolio optimization as a downstream task, do not care about MSE per se

$$\frac{1}{|\text{test}|} \sum_{i,t \in \text{test}} loss^{\text{econ}}\left(\widetilde{v}_{i,t}, \widehat{z}_{i,t}; \mu\right)$$

[$loss^{\text{least squares}}(\widetilde{v}, v) = (\widetilde{v} - v)^2$, $loss^{\text{econ}} = \ldots$ as defined]

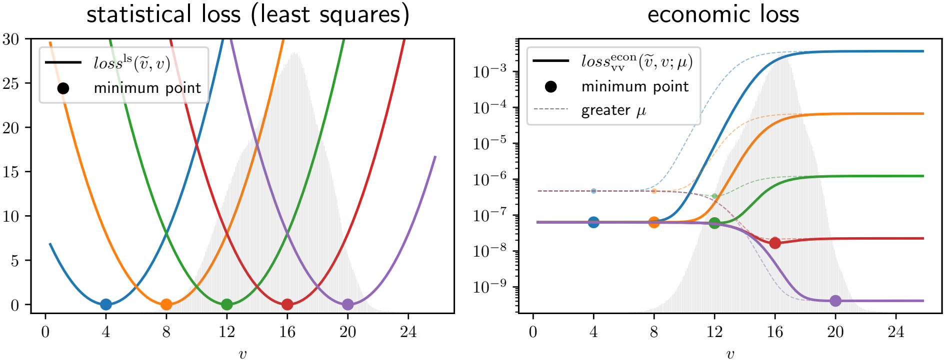
Approach 2, economic learning: directly learn z to optimize economic loss

$$\min_{z(\cdot)} \sum_{i,t \in \text{train}} loss^{\text{econ}}(\widetilde{v}_{i,t}, z(X_{i,t}))$$

Key ideas:

- let nn output z
- let training minimize economic loss
- back out $\hat{v}$ from $\hat{z}$

Comparing statistical and economic loss

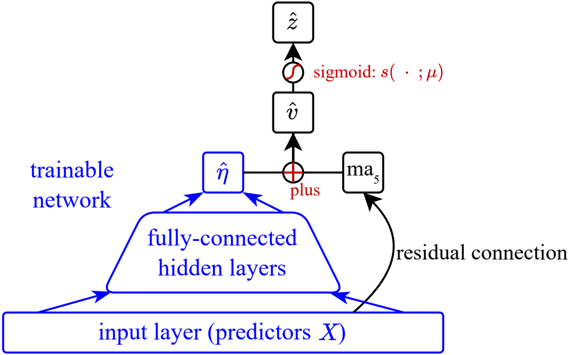


Economic loss is asymmetrical

- huge cost if actual $\widetilde{v}$ is low but predicted as high
  trade a lot when liquidity is low
- in contrast, opportunity cost for not trading is bounded

Economic learning: conservative to avoid overestimating volume

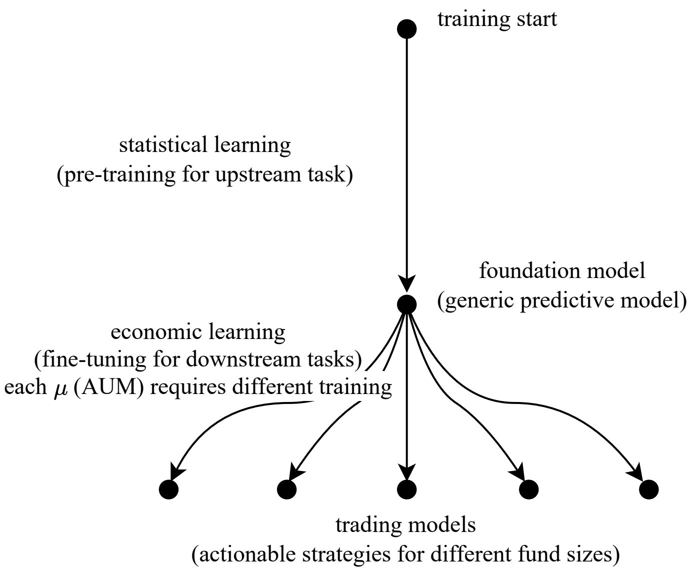
Network architecture, transfer learning



Transfer learning:

- statistical learning fits an nn model, applicable to all μ values
  serve as pre-trained foundation model
- economic learning fine-tunes the nn for each μ
  fine-tuning for downstream tasks, since $loss^{\text{econ}}$ is indexed by μ

Transfer learning routine



Today's Talk

- – Volume prediction from statistical perspective
- – Portfolio strategy based on volume prediction
- → Investment performance evaluation
- Key difference between statistical vs. economic learning
  - Less interested in getting volume "right"
  - More interested in making a better (economic) *decision*

Statistical vs. economic learning performance

	1	2	3	1	2	3
AUM	\$10b	\$1b	\$100m	\$10b	\$1b	\$100m
μ	1.2e-9	6.3e-8	4.7e-7	1.2e-9	6.3e-8	4.7e-7
avg $\widetilde{z}$	0.13	0.57	0.78	0.13	0.57	0.78
R^2 (% reduction in MSE)				% reduction in mean economic loss		
ma_5	———	0	———	———	0	———
ols	———	16.0	———	32.4	27.2	8.1
nn	———	18.4	———	33.3	28.2	6.2
rnn	———	19.9	———	34.8	29.5	11.7
nn.econ	10.0	-26.8	-34.9	39.6	69.2	70.9
rnn.econ	13.9	-0.6	-9.0	43.7	68.8	70.3
oracle	———	100	———	———	100	———

Note: OOS MSE and mean economic loss, i.e. \$1 trade task for each i, t

- Econ learning **sacrifices** MSE to **gain** economic value

Trading experiments, setup

Apply the strategies $(\widehat{z})$ to track various sets of trading targets $\{x^*_{i,t}\}$ that mimic different trading tasks, including

- simulated quantitative trading strategy
  implementing **before-cost** high-Sharpe strategies
- factor zoo portfolios (JKP factors)
  implementing different investment styles

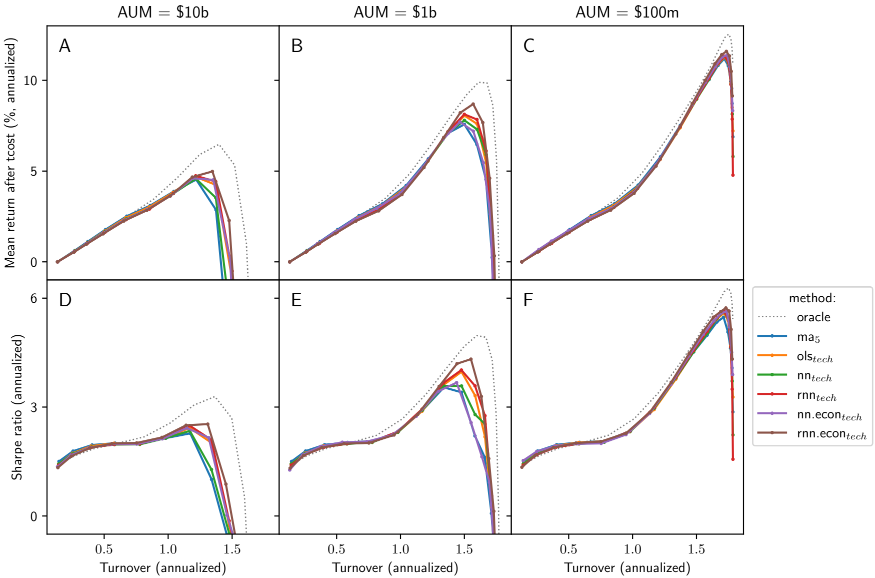
Recursively track each supplied target
Evaluate the investment outcome in the testing sample

Experiment 1, quantitative trading task

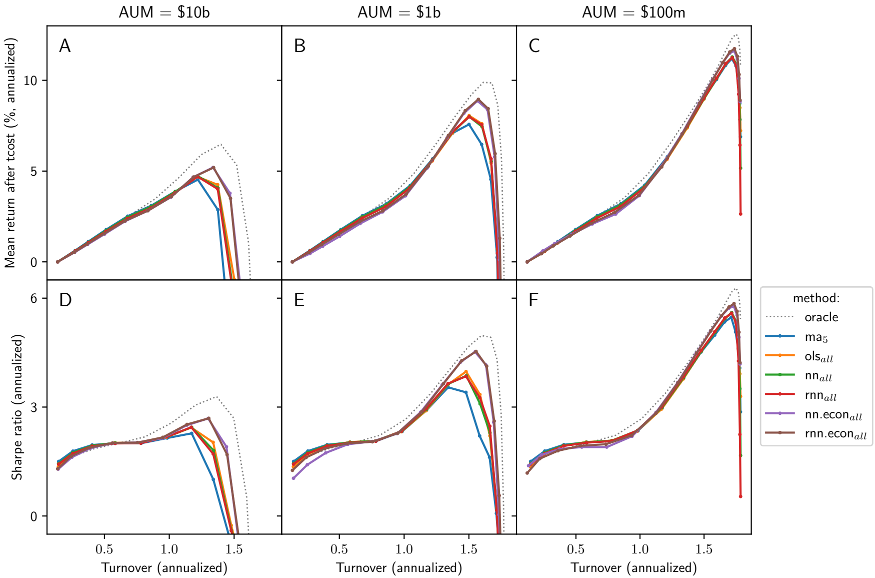
Construct targets $\{x^*_{i,t}\}$ with unrealistically high (before-cost) investment performance

- Sharpe ratio ≈ 7
- blow up to various AUM magnitudes

Performance across μ, only tech predictors



Performance across μ, all 175 predictors



Portfolio performance – quantitative trading task

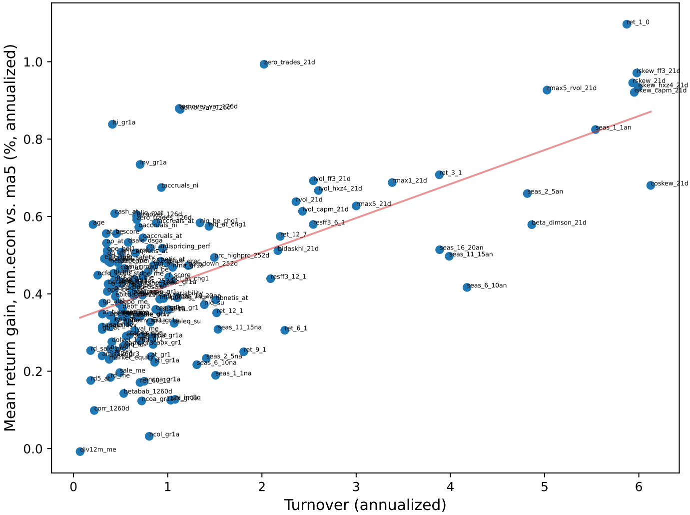
	A. Mean return (%, annualized)			B. Sharpe ratio (annualized)		
AUM	\$10b	\$1b	\$100m	\$10b	\$1b	\$100m
ma_5	3.88	6.47	11.19	2.00	2.21	5.47
ols_{all}	3.82	7.60	11.28	2.17	3.35	5.60
nn_{all}	3.86	7.44	11.28	2.19	3.09	5.60
rnn_{all}	3.79	7.55	11.25	2.18	3.26	5.59
nn.econ_{all}	4.64	8.87	11.61	2.18	4.24	5.74
rnn.econ_{all}	4.68	8.95	11.77	2.50	4.53	5.85
oracle	6.47	9.89	12.54	3.05	4.97	6.28

Experiment 2, implementing factor zoo portfolios

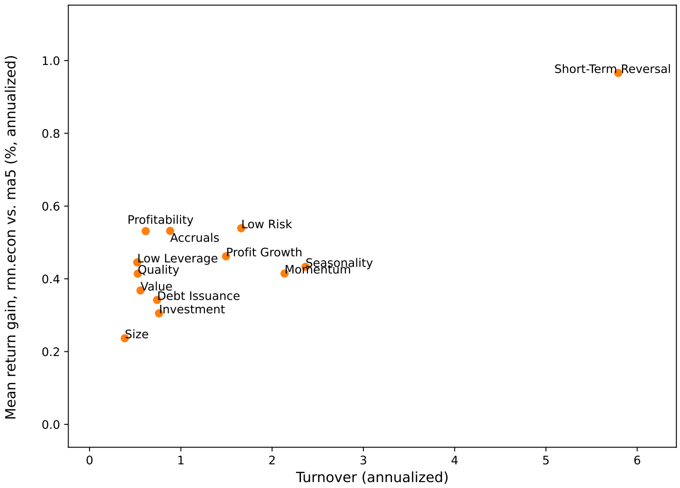
Long-short sorted portfolios for each 153 JKP characteristics

- AUM $10b, leverage 100%

Mean return improvements in implementing factor portfolio



Mean return improvements by theme clusters



Conclusion

- machine learning predicts daily stock-level trading volume
- powerful AI/ML tool turns predictability into economic value
- significant economic gain, comparable to return prediction
- Recasting forecasting of non-return variables into economic inputs
  - to assess economic impact
  - to improve the *value* of predictability

Trading Volume Alpha

Russ Goyenko	Bryan Kelly	Toby Moskowitz	Yinan Su	Chao Zhang
McGill	*Yale*	*Yale*	*Hopkins*	*HKUST-GZ*

EDHEC Speaker Series, October 28, 2025

APPENDIX

Microfoundation of trading model

Quadratic tracking error as the result of mean-variance optimization

$$A(1+r_t^{\mathrm{f}}) + \sum_i x_{i,t} m_{i,t} - \frac{\gamma}{2A} \sum_{i,j} x_{i,t} x_{j,t} \sigma_{ij,t}^2 - \sum_i TradingCost_{i,t},$$

$\Rightarrow$

$$-\frac{\gamma\sigma^2}{2A} \sum_i \left(x_{i,t} - \frac{A}{\gamma\sigma^2} m_{i,t} \right)^2 - \sum_i TradingCost_{i,t} + \left(A(1+r_t^{\mathrm{f}}) + \frac{A}{2\gamma\sigma^2} \sum_i m_{i,t}^2 \right)$$

(assuming homoscedsticity ($\mathbb{V}\mathrm{ar}_t r_{i,t+1} = \sigma^2$) and zero covariances)

Why trading volume?

A general approach to after-cost optimization

- trading costs are hard to predict

Why trading volume?

A general approach to after-cost optimization

- trading costs are hard to predict
- market liquidity proxied by volume, unknown ex ante
  (daily stock-level dollar trading volume)

$$PriceImpact_{a,i,t} \propto ParticipationRate_{a,i,t} := \frac{\$Traded_{a,i,t}}{\$Volume_{i,t}}$$

volume is generic across investors (a), exogenous, observable data

Why trading volume?

A general approach to after-cost optimization

- trading costs are hard to predict

- market liquidity proxied by volume, unknown ex ante
  (daily stock-level dollar trading volume)

$$PriceImpact_{a,i,t} \propto ParticipationRate_{a,i,t} := \frac{\$Traded_{a,i,t}}{\$Volume_{i,t}}$$

  volume is generic across investors (a), exogenous, observable data

- volume forecast informs trading decision
  volume $\uparrow \Rightarrow$ price impact $\downarrow \Rightarrow$ trade more aggressively (vice versa)

- better volume prediction $\Rightarrow$ better implementation

Introduction

Trading volume alpha:

- forecast volume $\rightarrow$ portfolio strategy $\rightarrow$ net-of-cost "alpha"

- new dataset, neural network methods, and new benchmarks
- transfer statistical learning into finance problem solving
- substantial economic gain, comparable to return prediction

Economically important frontier in portfolio research

Prediction results in size groups and “mixture of experts”

size group	jointly	nano	micro	small	large	mega
training obs	2,522,619	300,790	797,880	680,209	479,839	263,901
testing obs	1,893,067	273,792	467,413	552,503	384,819	214,540
A: pooled training evaluated in size groups and jointly						
ols$_{all}$	15.99	13.32	12.60	20.90	25.49	26.16
nn$_{all}$	18.45	15.80	14.86	23.71	27.76	29.12
rnn$_{all}$	19.86	16.63	16.14	26.00	30.50	32.02
B: size group training evaluated in size groups and jointly (mixture of experts)						
ols+moe$_{all}$	16.34	13.68	12.73	21.43	25.93	27.47
nn+moe$_{all}$	17.78	15.29	14.43	22.69	26.57	27.71
rnn+moe$_{all}$	18.26	15.24	14.71	24.76	29.02	30.99

Neural network tools

Run (ols and nn) regressions of $\widetilde{\eta}_{i,t}$ on $X_{i,t}$

Neural network tools

Run (ols and nn) regressions of $\widetilde{\eta}_{i,t}$ on $X_{i,t}$

Transparent and standard machine learning implementation

- nn (neural network):
  3-layer, fully-connected, 32-16-8 ReLU nodes

Neural network tools

Run (ols and nn) regressions of $\widetilde{\eta}_{i,t}$ on $X_{i,t}$

Transparent and standard machine learning implementation

- nn (neural network):
  3-layer, fully-connected, 32-16-8 ReLU nodes
- rnn (recurrent neural network):
  to capture time series predictive dependency,

$$(\widehat{\eta}_{i,t}, state_{i,t}) = rnn(X_{i,t}, state_{i,t-1})$$

  architecture: lstm as the first hidden layer, all else the same

Theoretical analysis

Proposition: least squares optimizer does not maximize economic objective, even in population

- well-known: conditional expectation is the least squares minimizer

$$\mathbb{E}[\widetilde{v}|X] = \min_{v \in \sigma(X)} \mathbb{E}[loss^{\text{least squares}}(\widetilde{v}, v)]$$

- how about other loss functions?
  iff the loss function is in Bregman class (Banerjee, et. al., 2005; Patton, 2020)
  however, $loss^{\text{econ}}_{\text{vv}}$ is not in Bregman class

z learning performance, statistical vs. economic learning

	1	2	3	4	1	2	3	4
μ	1.2e-9	6.3e-8	4.7e-7	9.4e-6	1.2e-9	6.3e-8	4.7e-7	9.4e-6
avg $\tilde{z}$	0.13	0.57	0.78	0.95	0.13	0.57	0.78	0.95
relevant AUM	\$10b	\$1b	\$100m	\$10m	\$10b	\$1b	\$100m	\$10m
A. Mean economic loss (MEL) ($\times 10^{-8}$)					**A′. % reduction in mean economic loss**			
ma$_5$	0.1046	3.163	15.41	93.0	0.0	0.0	0.0	0.0
ols$_{tech}$	0.1041	3.011	14.81	93.2	27.9	29.6	11.1	-0.3
nn$_{tech}$	0.1043	3.100	14.97	99.7	19.2	12.3	8.0	-12.8
rnn$_{tech}$	0.1040	2.955	14.37	102.2	33.6	40.8	19.2	-17.7
nn.econ$_{tech}$	0.1041	2.991	12.35	66.8	31.3	33.6	56.5	50.2
rnn.econ$_{tech}$	0.1039	2.855	11.78	64.2	39.7	60.4	67.0	55.3
ols$_{all}$	0.1040	3.024	14.97	94.7	32.4	27.2	8.1	-3.3
nn$_{all}$	0.1040	3.019	15.07	106.5	33.3	28.2	6.2	-25.9
rnn$_{all}$	0.1040	3.012	14.78	109.8	34.8	29.5	11.7	-32.1
nn.econ$_{all}$	0.1039	2.810	11.56	61.9	39.6	69.2	70.9	59.6
rnn.econ$_{all}$	0.1038	2.812	11.60	66.4	43.7	68.8	70.3	51.0
oracle	0.1029	2.653	9.99	40.8	100	100	100	100
B. Mean squared error (MSE)					**B′. R^2 (% reduction in MSE)**			
ma$_5$	—— 0.437 ——				—— 0.0 ——			
ols$_{tech}$	—— 0.385 ——				—— 12.1 ——			
nn$_{tech}$	—— 0.375 ——				—— 14.3 ——			
rnn$_{tech}$	—— 0.368 ——				—— 15.8 ——			
nn.econ$_{tech}$	0.389	0.449	0.457	0.630	11.2	-2.6	-4.5	-44.1
rnn.econ$_{tech}$	0.392	0.492	0.487	0.481	10.3	-12.4	-11.3	-10.0
ols$_{all}$	—— 0.367 ——				—— 16.0 ——			
nn$_{all}$	—— 0.357 ——				—— 18.4 ——			
rnn$_{all}$	—— 0.350 ——				—— 19.9 ——			
nn.econ$_{all}$	0.394	0.555	0.590	1.979	10.0	-26.8	-34.9	-352.5
rnn.econ$_{all}$	0.377	0.440	0.477	0.785	13.9	-0.6	-9.0	-79.5
oracle	—— 0.00 ——				—— 100 ——			

Performance with μ hyper-parameter tuning (complete table)

	A. Mean return (%, annualized)				B. Sharpe ratio (annualized)			
AUM	\$10b	\$1b	\$100m	\$10m	\$10b	\$1b	\$100m	\$10m
ma$_5$	3.88	6.47	11.19	13.20	2.00	2.21	5.47	6.55
ols$_{tech}$	3.82	7.60	11.28	13.14	2.16	3.32	5.59	6.59
nn$_{tech}$	3.76	7.30	11.32	13.13	2.14	2.79	5.63	6.59
rnn$_{tech}$	3.74	7.84	11.33	13.13	2.18	3.58	5.64	6.59
nn.econ$_{tech}$	4.60	7.20	11.29	13.22	2.13	2.57	5.59	6.63
rnn.econ$_{tech}$	4.67	8.69	11.59	13.30	2.50	4.32	5.73	6.66
ols$_{all}$	3.82	7.60	11.28	13.14	2.17	3.35	5.60	6.59
nn$_{all}$	3.86	7.44	11.28	13.13	2.19	3.09	5.60	6.58
rnn$_{all}$	3.79	7.55	11.25	13.09	2.18	3.26	5.59	6.56
nn.econ$_{all}$	4.64	8.87	11.61	13.29	2.18	4.24	5.74	6.66
rnn.econ$_{all}$	4.68	8.95	11.77	13.30	2.50	4.53	5.85	6.68
oracle	6.47	9.89	12.54	13.56	3.05	4.97	6.28	6.80