Machine Learning in Asset Pricing:

Rise, Breakthroughs, and Limitations

Dacheng Xiu

Chicago Booth and NBER

January 22, 2025

Disclaimer: This presentation provides an overview of select aspects of my work on this topic. For a more comprehensive review, please refer to Financial Machine Learning by Kelly and Xiu, Foundations and Trends in Finance, Vol. 13, No. 3-4, (2023), 205-363, and Factor Models, Machine Learning, and Asset Pricing, by Giglio, Kelly, and Xiu, Annual Review of Financial Economics, Vol. 14, (2022), 337-368.

The Rise of the Machines

"Two Pillars of Asset Pricing," **Fama (AER, 2014)**

- Market Efficiency and Time-varying Expected Stock Returns
- Asset Pricing Models and the Cross Section of Expected Returns

Measurement of an asset's expected return is fundamentally a problem of prediction!

Why ML?

What is ML?

- Historical Perspective: Arthur Samuel coined the term Machine Learning in his 1959 paper: "Some studies in machine learning using the game of checkers."
- Modern view: Machine Learning integrates statistical methods to extract information from data with computational techniques that facilitate efficient implementation on large datasets.

Why ML?

- Lots of explanatory variables with potentially high correlations
- Functional form is unknown and likely complex (nonlinear)

"ML" is Inevitable

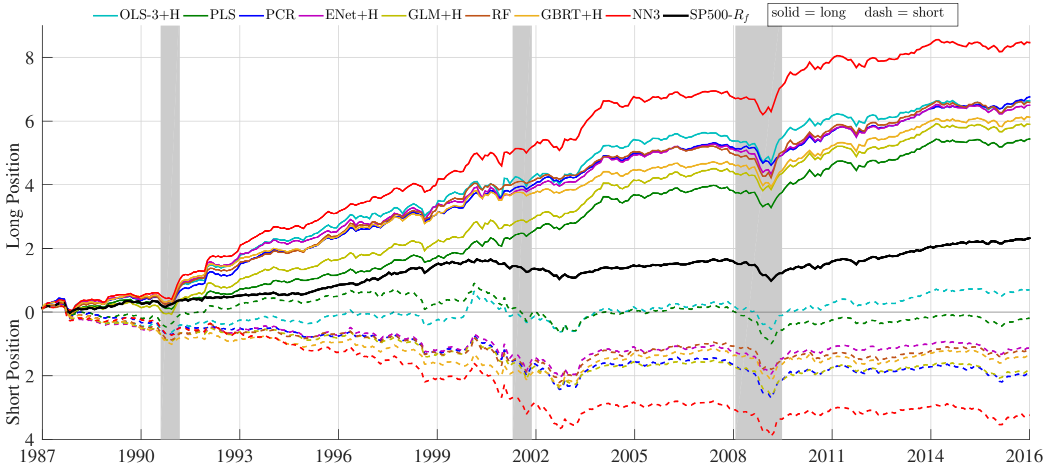
Has ML been used before? **Yes**, perhaps without knowing or emphasizing it.

"In Finance, Humans Were the First Machines," **CBR, Feb 05, 2024**

- **nonlinearity**: sorting returns by characteristics
- **dimension reduction:** employing portfolios instead of individual assets
- **variable selection and factor analysis**: e.g., Fama-French factors
- **priors** (regularization): incorporating economic intuition/theory
- ...

Why not embrace modern ML techniques?

"Empirical Asset Pricing via Machine Learning", **Gu, Kelly, and Xiu, RFS (2020)**



Black Box? Interpretability

- Interpretability is a highly relevant and expansive topic in AI.
  It describes our ability to understand and articulate the decision-making processes of AI models.
  - **Statistical:** What functions can a NN represent? How effectively can it learn?
    Unlike straightforward linear models, NNs are highly non-linear and non-convex.
  - **Algorithmic:** Why do various ad hoc training algorithms succeed?
    For instance, while ordinary least squares are well-understood, NNs use complex mechanisms like batch normalization, dropout, and Adam optimizer.
  - **Economic:** Which variables are important, and why?
    While tools like feature importance and SHAP address which variables matter, explaining 'why' remains challenging, especially in models with many variables.
  - **Ethical, Regulatory, Legal, and more**

Return Prediction Models For "Alternative" Data

"(Re-)Imag(in)ing Price Trends" **Jiang, Kelly, and Xiu, JF (2023)**

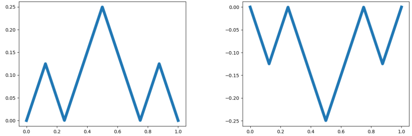
- "Nevertheless, technical analysis has survived through the years, perhaps because its visual mode of analysis is more conducive to human cognition, and because pattern recognition is one of the few repetitive activities for which computers do not have an absolute advantage (yet)."

  Lo, Mamaysky, and Wang (JF, 2000)
- If a human can more readily detect patterns in images, one can imagine a statistical model may as well.
- Computer vision models, e.g., CNN, can turn a price chart into return predictive signals that are more profitable (despite being more complex).

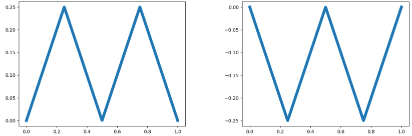
A Technical Chart

Note: OHLC chart for Tesla stock with 20-day moving average price line and daily volume bars. Daily data from January 1, 2020 to August 18, 2020.

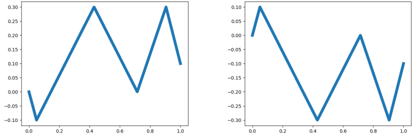
Some Patterns Discovered by "Chartist"



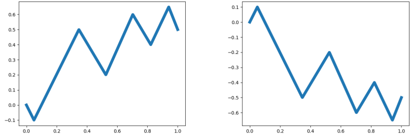
Head and Shoulders Top/Bottom



Double Top/Bottom



Ascending/Descending Triangles



Rising/Falling Wedges

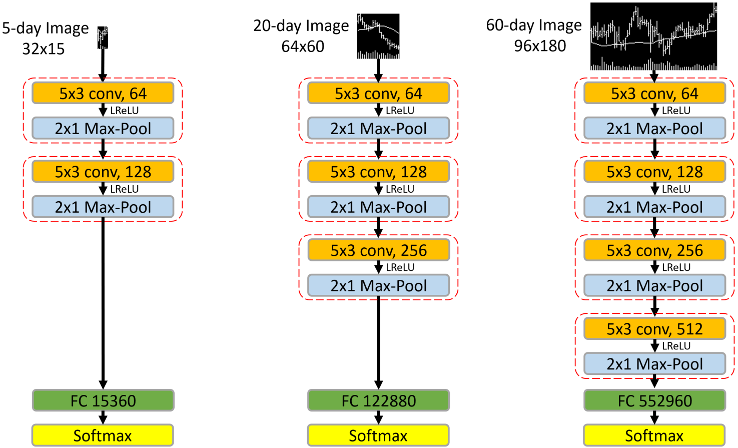
Technical Signals Documented by Financial Economists

- Momentum: Asness (1994), Jegadeesh and Titman (1993)
- Short-term Reversal: Lehmann (1990), Lo and MacKinlay (1990)
- Long-term Reversal: DeBondt and Thaler (1985)
- ...

Man vs Machine

- Man-made signals are simple and transparent
- This does not mean they are necessarily more interpretable
- The economics behind technical signals are often ambiguous, e.g., momentum
  - Da, Gurun, and Warachka (RFS, 2014): investors underreact to information
  - Antoniou, Doukas, and Subrahmanyam (2012, JFQA): investors' cognitive dissonance
  - Hong, Stein, and Yu (2007, JF): investors make persistent forecast errors
  - Grinblatt and Han (2005, JFE): investors' disposition effect
  - ...
  - Nonetheless, a 2014 study "Fact, Fiction and Momentum Investing" by a team led by AQR's Clifford Asness urged momentum-doubters to surrender, already. "The existence of momentum," they wrote, has become "a well-established empirical fact."
- Can machines discover more complex and more profitable signals?

Technical Patterns Recognition via CNNs



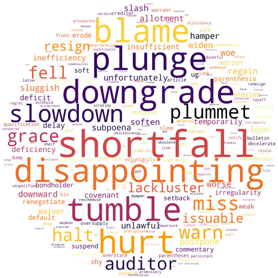
Stock price charts are translated into probabilities of Up (Down) in the next trading period.

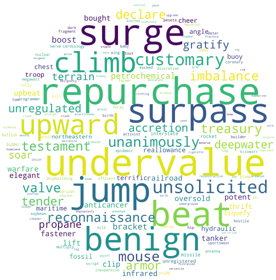
Performance of CNN-based Portfolios

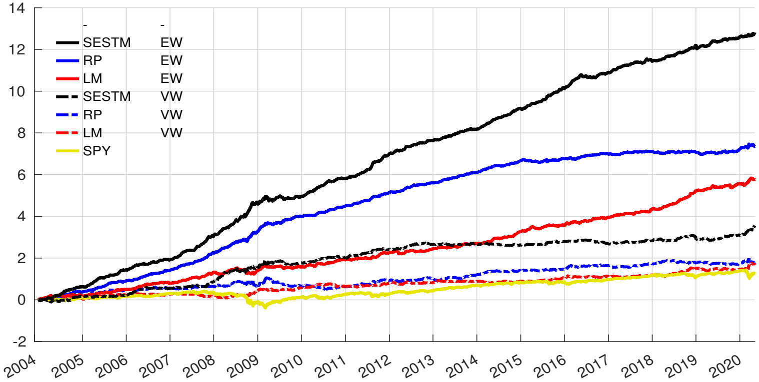
	I5/R5		I20/R5		I60/R5		MOM/R5		STR/R5		WSTR/R5		TREND/R5	
	Ret	SR	Ret	SR	Ret	SR	Ret	SR	Ret	SR	Ret	SR	Ret	SR
Low	-0.28	-1.92	-0.32	-1.94	-0.21	-1.10	0.15	0.44	-0.01	-0.03	-0.08	-0.34	-0.11	-0.46
High	0.54	2.89	0.52	2.76	0.33	1.85	0.16	0.78	0.38	1.19	0.46	1.56	0.48	1.58
H-L	0.83***	7.15	0.84***	6.75	0.54***	4.89	0.02	0.07	0.39***	1.76	0.53***	2.84	0.59***	2.92
Turnover	690%		667%		619%		123%		341%		660%		499%	
	I5/R20		I20/R20		I60/R20		MOM/R20		STR/R20		WSTR/R20		TREND/R20	
	Ret	SR	Ret	SR	Ret	SR	Ret	SR	Ret	SR	Ret	SR	Ret	SR
Low	-0.00	-0.03	-0.02	-0.12	-0.02	-0.07	0.07	0.20	0.05	0.23	0.01	0.03	-0.01	-0.05
High	0.21	1.09	0.18	1.04	0.14	0.99	0.14	0.74	0.16	0.51	0.19	0.66	0.21	0.76
H-L	0.22***	2.35	0.21***	2.16	0.16***	1.29	0.07	0.25	0.11**	0.55	0.18***	1.23	0.22***	1.39
Turnover	175%		173%		155%		63%		168%		167%		140%	
	I5/R60		I20/R60		I60/R60		MOM/R60		STR/R60		WSTR/R60		TREND/R60	
	Ret	SR	Ret	SR	Ret	SR	Ret	SR	Ret	SR	Ret	SR	Ret	SR
Low	0.07	0.31	0.08	0.32	0.08	0.27	0.11	0.25	0.07	0.27	0.06	0.22	0.09	0.23
High	0.16	0.77	0.13	0.74	0.15	0.88	0.13	0.57	0.14	0.40	0.16	0.48	0.16	0.55
H-L	0.09***	1.30	0.05	0.37	0.07*	0.43	0.02	0.06	0.07	0.34	0.10***	0.65	0.07*	0.38
Turnover	59%		59%		58%		37%		56%		56%		51%	

Note: "Ix/Ry": the model uses x-day images to predict subsequent y-day holding period returns.

"Predicting Returns with Text Data" **Ke, Kelly, and Xiu (2019, WP)**







"Expected Returns and Large Language Models" Chen, Kelly, and Xiu (2023, WP)

- LLMs are developed by a specialized team to estimate a highly parameterized general purpose language model using vast amounts of text from diverse sources and themes.

Model	Company	Release Date	# of Parameters
BERT	Google	Oct 2018	110M(Base)/340M(Large)
GPT-2	OpenAI	Feb 2019	1.5B
RoBERTa	Facebook	Jul 2019	125M(Base)/355M(Large)
Turing-NLG	Microsoft	Feb 2020	17B
GPT-3	OpenAI	Jun 2020	175B
Wu Dao 2.0	BAAI	May 2021	1.75T
Jurassic-1	AI21	Aug 2021	178B
GLaM	DeepMind	Dec 2021	280B
Exaone	LG	Dec 2021	300B
Megatron-Turing NLG	Microsoft/Nvidia	Oct 2021	530B
OPT	Meta AI	May 2022	from 125M to 175B
GPT-3.5 (ChatGPT)	OpenAI	Nov 2022	175B
LLaMA	Meta	Feb 2023	7B, 13B, 33B, and 65B
GPT-4	OpenAI	Mar 2023	1T(?)
LLaMA 2	Meta	July 2023	7B, 13B, and 70B

Economic Motivation

- News text stands as a treasure trove of alternative data, ripe with insights that can illuminate the complexities of stock return behaviors. It's not just about the news—it's about uncovering the hidden narratives that drive market movements.
- While the market efficiently processes information, it doesn't operate at lightning speed. The impact of news on stock prices varies in both velocity and magnitude, introducing intriguing dynamics in how information dissemination affects market equilibrium.
- The inefficiency in incorporating news into stock prices breeds a short-term momentum effect. This phenomenon, where prices gradually adjust to new information, presents unique opportunities for investment strategies.

Context > Words > Past Returns

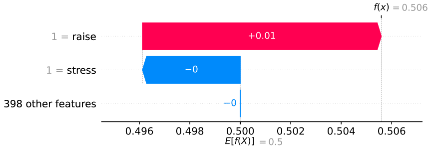
	Stocks with news						ChatGPT					
	EW			VW			EW			VW		
	Long	Short	L-S	Long	Short	L-S	Long	Short	L-S	Long	Short	L-S
Ret	0.35	0.06	0.29	0.29	0.08	0.20	0.38	-0.16	0.54	0.22	0.04	0.18
Std	0.27	0.23	0.18	0.30	0.23	0.24	0.21	0.22	0.11	0.19	0.22	0.11
SR	1.29	0.25	1.58	0.95	0.37	0.83	1.86	-0.71	5.03	1.13	0.20	1.58
	LLaMA2						LLaMA					
	EW			VW			EW			VW		
	Long	Short	L-S	Long	Short	L-S	Long	Short	L-S	Long	Short	L-S
Ret	0.40	-0.12	0.52	0.21	0.07	0.14	0.37	-0.08	0.45	0.21	0.09	0.12
Std	0.21	0.23	0.12	0.20	0.21	0.12	0.21	0.22	0.11	0.20	0.22	0.12
SR	1.88	-0.53	4.43	1.06	0.34	1.19	1.79	-0.34	4.00	1.06	0.42	1.00
	RoBERTa						BERT					
	EW			VW			EW			VW		
	Long	Short	L-S	Long	Short	L-S	Long	Short	L-S	Long	Short	L-S
Ret	0.39	-0.11	0.49	0.24	0.08	0.16	0.38	-0.07	0.44	0.20	0.06	0.14
Std	0.21	0.22	0.11	0.20	0.22	0.12	0.21	0.22	0.11	0.19	0.21	0.11
SR	1.84	-0.48	4.46	1.22	0.36	1.38	1.80	-0.31	4.13	1.01	0.28	1.22
	SESTM						LMMD					
	EW			VW			EW			VW		
	Long	Short	L-S	Long	Short	L-S	Long	Short	L-S	Long	Short	L-S
Ret	0.42	-0.03	0.45	0.30	0.05	0.26	0.26	-0.01	0.28	0.17	0.09	0.07
Std	0.26	0.22	0.17	0.27	0.23	0.21	0.20	0.22	0.10	0.18	0.21	0.10
SR	1.63	-0.12	2.61	1.10	0.21	1.21	1.30	-0.06	2.79	0.90	0.44	0.71

Why LLMs Disagree with Word-based Methods?

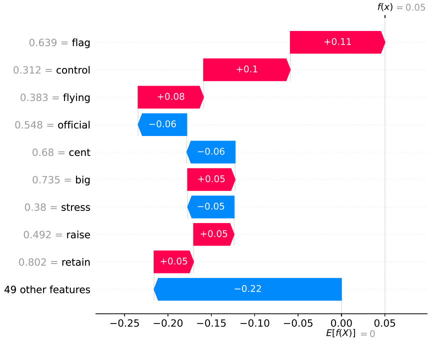
LLAMA2

Brussels has warned British Airways owner IAG ICAG.L that its favoured strategy to allow it to continue flying freely in and around Europe in the event of a nodeal Brexit will not work, the Financial Times reported on Tuesday.After Brexit, European carriers will have to show they are more than 50 per cent EUowned and controlled to retain flying rights in the bloc, the FT said.IAG, which also owns the Spanish flag carrier Iberia, is registered in Spain but headquartered in the United Kingdom and has diverse global shareholders.The FT said part of IAG's strategy to retain both EU and UK operating rights is to stress that its important individual airlines are domestically owned through a series of trusts rather than being part of the bigger a high proportion of nonEU investors.The FT quoted an unnamed senior EU official as saying, "For IAG, I can't see how it can be a solution."Concerns have been raised with IAG over its postBrexit ownership structure, the FT quoted a second Brussels official familiar with the conversations as saying.IAG was not immediately available.

BOW



W2V



Negation Portfolio

	ChatGPT						LLAMA2					
	W/O Negation Words			W/ Negation Words			W/O Negation Words			W/ Negation Words		
	Long	Short	L-S	Long	Short	L-S	Long	Short	L-S	Long	Short	L-S
Ret	0.40	-0.15	0.56	0.43	-0.23	0.66	0.35	-0.07	0.42	0.48	-0.22	0.70
Std	0.21	0.24	0.13	0.21	0.25	0.17	0.21	0.24	0.13	0.22	0.25	0.17
SR	1.96	-0.64	4.34	2.05	-0.90	3.98	1.70	-0.28	3.29	2.21	-0.87	4.18
	LLAMA						RoBERTa					
	W/O Negation Words			W/ Negation Words			W/O Negation Words			W/ Negation Words		
	Long	Short	L-S	Long	Short	L-S	Long	Short	L-S	Long	Short	L-S
Ret	0.36	-0.06	0.43	0.50	-0.21	0.71	0.34	-0.07	0.41	0.51	-0.19	0.70
Std	0.21	0.24	0.13	0.22	0.25	0.17	0.21	0.24	0.13	0.22	0.24	0.16
SR	1.74	-0.27	3.34	2.32	-0.82	4.23	1.64	-0.30	3.14	2.37	-0.76	4.35
	BERT						SESTM					
	W/O Negation Words			W/ Negation Words			W/O Negation Words			W/ Negation Words		
	Long	Short	L-S	Long	Short	L-S	Long	Short	L-S	Long	Short	L-S
Ret	0.33	-0.03	0.36	0.45	-0.11	0.56	0.33	-0.05	0.38	0.38	-0.01	0.40
Std	0.21	0.23	0.12	0.22	0.25	0.17	0.21	0.24	0.13	0.22	0.25	0.15
SR	1.56	-0.14	2.94	2.06	-0.45	3.37	1.57	-0.22	3.00	1.78	-0.05	2.58
	Word2vec						LMMD					
	W/O Negation Words			W/ Negation Words			W/O Negation Words			W/ Negation Words		
	Long	Short	L-S	Long	Short	L-S	Long	Short	L-S	Long	Short	L-S
Ret	0.28	-0.04	0.32	0.32	-0.01	0.33	0.26	-0.03	0.28	0.29	0.04	0.25
Std	0.21	0.23	0.12	0.22	0.24	0.15	0.21	0.24	0.12	0.21	0.24	0.15
SR	1.35	-0.18	2.71	1.49	-0.02	2.21	1.25	-0.11	2.31	1.35	0.17	1.66

Complexity: Article Headline vs Article Body

	ChatGPT						LLAMA2					
	Headline			Body			Headline			Body		
	Long	Short	L-S	Long	Short	L-S	Long	Short	L-S	Long	Short	L-S
Ret	0.33	-0.09	0.42	0.34	-0.14	0.48	0.32	-0.04	0.36	0.35	-0.10	0.45
Std	0.20	0.22	0.10	0.20	0.22	0.10	0.20	0.22	0.10	0.20	0.23	0.11
SR	1.65	-0.41	4.12	1.71	-0.62	4.62	1.58	-0.19	3.51	1.75	-0.43	4.16
	RoBERTa						BERT					
	Headline			Body			Headline			Body		
	Long	Short	L-S	Long	Short	L-S	Long	Short	L-S	Long	Short	L-S
Ret	0.35	-0.00	0.35	0.33	-0.06	0.39	0.31	-0.02	0.33	0.32	-0.04	0.36
Std	0.21	0.22	0.10	0.20	0.22	0.10	0.20	0.21	0.10	0.20	0.22	0.10
SR	1.69	-0.02	3.47	1.62	-0.29	3.75	1.55	-0.09	3.48	1.59	-0.19	3.60
	SESTM						Word2vec					
	Headline			Body			Headline			Body		
	Long	Short	L-S	Long	Short	L-S	Long	Short	L-S	Long	Short	L-S
Ret	0.38	-0.02	0.40	0.31	-0.03	0.34	0.35	-0.01	0.37	0.29	-0.01	0.30
Std	0.21	0.22	0.10	0.20	0.22	0.10	0.21	0.23	0.11	0.21	0.22	0.10
SR	1.84	-0.07	4.05	1.53	-0.14	3.43	1.70	-0.06	3.32	1.41	-0.05	3.06

Beyond Sentiment Analysis

- Economic text is constantly generated by human writers striving to understand and make predictions about economic phenomena.
- This ever-growing body of text represents a valuable source of alternative data, filled with narratives that shed light on the complex behavior of financial markets.
  e.g., "Business News and Business Cycles", **Bybee, Kelly, Manela, and Xiu (JF, forthcoming)**
- LLMs provide more sophisticated text representations, capturing syntactic and semantic nuances conveyed through word order and contextual relationships.

Asset Pricing "Anomalies"

"Taming the Factor Zoo: A Test of New Factors", **Feng, Giglio, and Xiu (JF, 2020)**

- Historically, factors in the "zoo" have been benchmarked against FF3.
  - Not surprising to observe factor proliferation if we have been rediscovering the same anomalies over and over: we have not updated the benchmark to incorporate new discoveries
  - E.g., Seasonality (Heston and Sadka (JFE, 2008)) has correlations 0.63 with Carhart Momentum and 0.78 with 6M Momentum (Jegadeesh and Titman (JF, 1993))
    FF3 Alpha: 27 (2.06) FF4 Alpha: -9 (-0.88)
- This paper adopts double machine learning Chernozhukov et al. (EJ, 2018) to determine if a factor is new or redundant, controlling all existing factors
  - A natural application for advances in machine learning and model selection techniques.
  - Chronological evaluation of factors reveals significantly fewer factors.

Risk Premia, Alphas, and Weak Factors

"Asset Pricing with Omitted Factors", **Giglio and Xiu (JPE, 2021)**

- Risk premia estimation and inference, without specifying an asset pricing model.

"Thousands of Alpha Tests", **Giglio, Liao, and Xiu (RFS, 2021)**

- A toolkit for addressing false discoveries in alpha tests, robust against missing data and omitted variable bias.

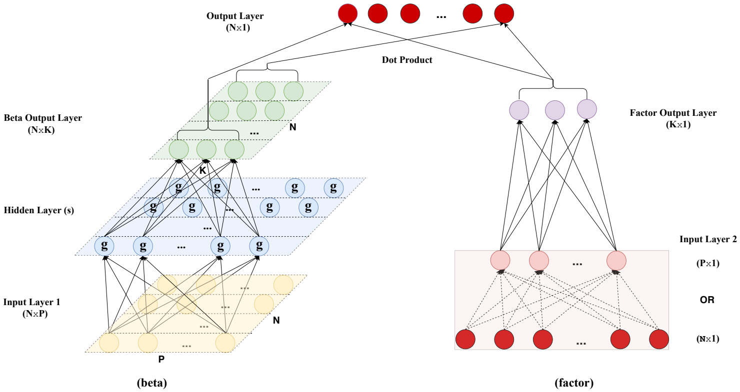
"Test Assets and Weak Factors" **Giglio, Xiu, and Zhang (JF, forthcoming)**

- A new approach that serves a dual purpose: first, it provides a well-founded basis for the selection of test assets, and second, it leverages the selection to mitigate the bias in risk premium estimation.

Risk-Return Tradeoffs

"Autoencoder Asset Pricing Models", **Gu, Kelly, and Xiu (J of Econometrics, 2023)**

- The first deep learning model of individual equity returns that effectively disentangles expected returns into components of risk compensation and minimal mispricing.

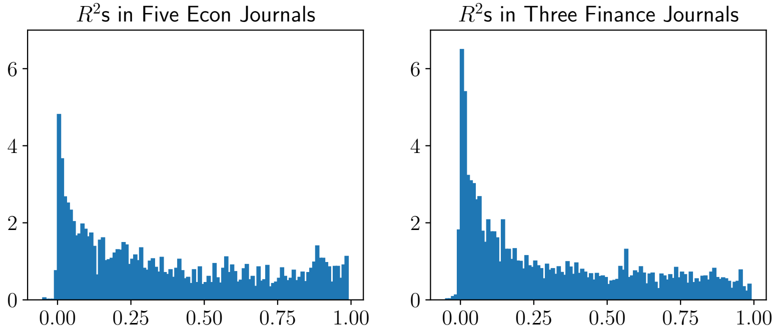


The Limits of the Machines

"The Statistical Limit of Arbitrage", **Da, Nagel, and Xiu (WP, 2023)**

- Economic agents (arbitrageurs) are often assumed to have rational expectations (Lucas (1978)), so that the effect of learning disappears.
- In practice, rational expectations are seldom a precise model of reality. This raises the question: **what are the economic consequences of learning?**
- The necessity for arbitrageurs to learn imposes a limit on the gains achievable through machine-learned arbitrage strategies.
  - The challenge of learning about *a large number of* alphas induces a limit to arbitrage due to statistical learning.
  - Pricing errors (alphas) $\neq$ Arbitrage opportunities
  - This arbitrage profit is much lower than what investors might anticipate.

"Can Machines Learn Weak Signals?", **Shen and Xiu (WP, 2024)**

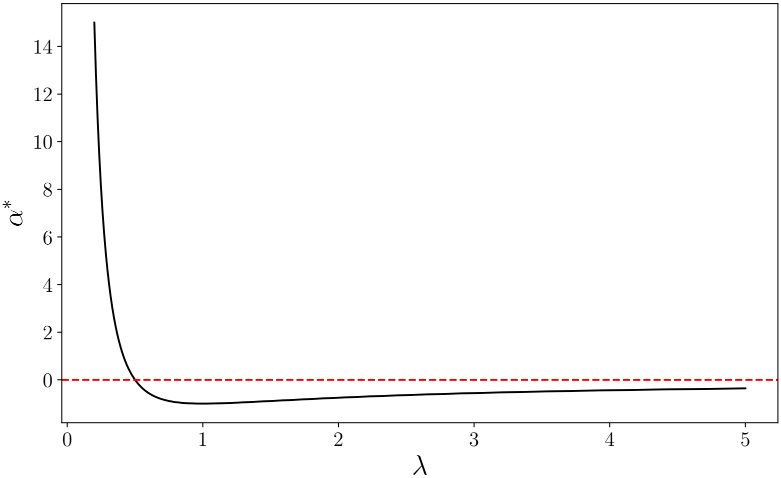


- When Econ and Fin adopted ML from CS and Stat, there were concerns about its efficacy given the low signal-to-noise ratio.

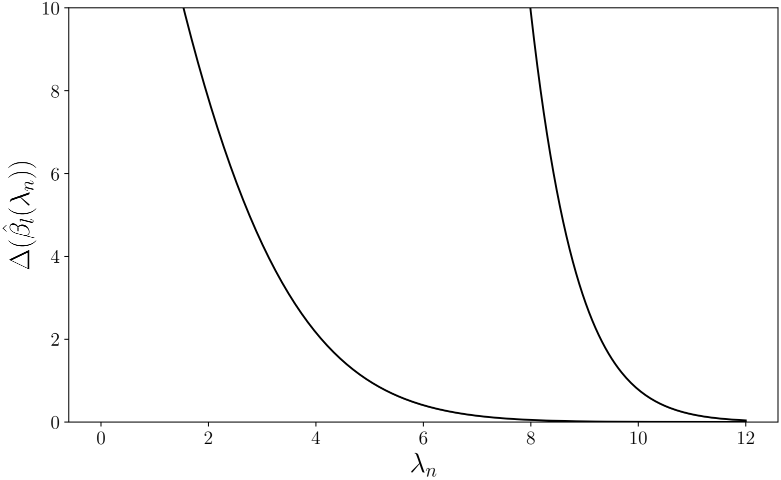
Machine Learning

- Machine learning methods have proved effective in mitigating overfitting and discerning true signals from fake ones when the true signals are strong.
- These methods employ regularization techniques, such as penalizing ℓ_1 or ℓ_2 norms of model parameters, to achieve this.
- A pivotal question arises: Can machines learn weak signals?
  - Only if they can, can they outperform the (naive) baseline zero-estimator!
  - Yet, both Lasso and Ridge nest Zero as a special case.

Optimal Ridge Beats Zero!



Optimal Lasso is Zero!



Empirical Evidence

Dataset constructed by Welch and Goyal (RFS, 2008).

- US market (S&P 500) return via 16 lagged financial and macroeconomic indicators
- Sample: 74 annual time-series observations from 1948 to 2021

	Ridge	Lasso	OLS/Ridgeless	RF	GBRT	NN(ℓ_2)	NN(ℓ_1)
R^2_{oos}	0.8%	-12.19%	-81.08%	-1.08%	-14.21%	1.41%	-10.31%

Dataset constructed by Gu, Kelly and Xiu (RFS, 2020).

- US Individual Equity returns via 920 covariates, including characteristics, macroeconomic predictors and their interactions
- Sample: Monthly returns from CRSP for all firms from $1957 - 2021$

	Ridge	Lasso	OLS/Ridgeless	RF	GBRT	NN(ℓ_2)	NN(ℓ_1)
R^2_{oos}	0.19%	0.10%	-1.25%	0.08%	-0.30%	0.26%	0.14%