Discussion of Gabaix, Koijen, Richmond, and Yogo (2025)

“Asset Embeddings”

Discussant: Sangmin Simon Oh (Columbia Business School)

Five-Star Conference 2025

Much of Asset Pricing Uses Embeddings

- **Firm characteristics:** Manually engineered low-dimensional representation

- **Characteristics + Macro Variables:** Context-dependent embeddings

- **Latent states x_t in affine term structure models:** embeddings of the macro-finance environment

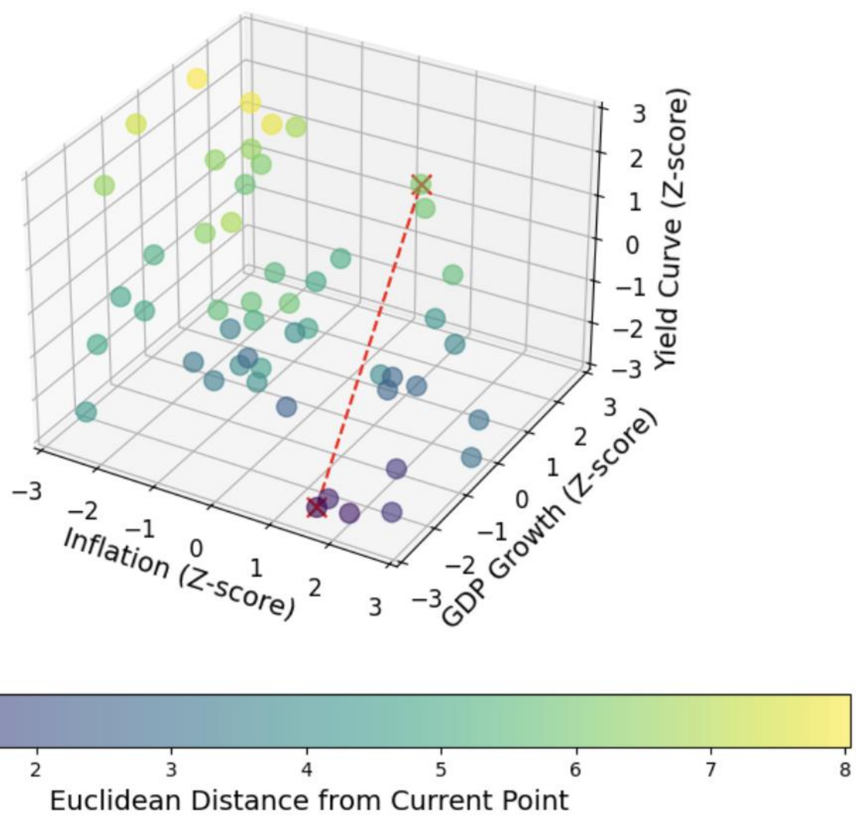
- **Option-implied risk-neutral densities:** embedding of the market's beliefs

- **Investor characteristics** (e.g. style tilts, mandates, horizon): Embedding of an institutional identity

⋮

More Recent Examples of Embeddings

Identifying similar historical moments:



Source: Verdad Research (2024), “Analogous Market Moments”

More Recent Examples of Embeddings

Identifying similar historical moments:

Haven't We Seen This Before? Return Predictions from 200 Years of News

AJ Chen Gerard Hoberg Miao Ben Zhang *

June 2025

Abstract

We postulate that our historical record has become adequately long and informative that newly arriving economic states often resemble historical states. Building on this insight, we develop a framework to predict future economic outcomes using the average of the realized outcomes that follow highly similar historical states. Using 210 million newspaper articles from 1815 to 2021, we identify historically similar months for each focal month and construct a predictor of aggregate U.S. stock returns, "SeenItRet". SeenItRet strongly forecasts future market-wide stock returns up to two years ahead, with an annualized impact of 4–7% for a one standard deviation shift. Our framework is general and also predicts real economic outcomes, including recessions, inflation, and patenting activity. A virtue of our approach is its use of economic principles to reduce the high dimensionality of the underlying state space to an ex-ante measurable and intuitive unidimensional predictor. Our model performs better when historical states are more similar to the focal state, and it offers interpretable economic insights by highlighting the specific themes that drive its predictions.

More Recent Examples of Embeddings

Compress vast information into tractable latent representation:

Who Owns What?
A Factor Model for Direct Stockholding*

Vimal Balasubramaniam[†] John Y. Campbell[§]
Tarun Ramadorai[‡] Benjamin Ranish[¶]

February 8, 2021

Abstract

We build a cross-sectional factor model for investors' direct stockholdings, by analogy with standard time-series factor models for stock returns. We estimate the model using data from almost 10 million retail accounts in the Indian stock market. We find that stock characteristics such as firm age and share price have strong investor clienteles associated with them. Similarly, account attributes such as account age, account size, and extreme underdiversification (holding a single stock) are associated with particular characteristic preferences. Coheld stocks tend to have higher return correlations, suggestive of the importance of clientele effects in the stock market.

Recap

Question: Can we construct, using holdings and tools from AI/ML, low-dimensional representation of assets that can explain valuations, returns, and portfolio choices?

Methodology

- Motivate asset embeddings using a simple micro-foundation
- Estimate latent representation inspired by (1) recommender systems and (2) transformers
- Evaluate embeddings using three benchmarks and interpret them using text data

Main Findings

- Holdings-based embeddings encode rich economic structure, outperforming characteristic-based and text-based models
- Different embedding architectures specialize in different economic tasks

Deep, paradigm-shifting paper with multiple core contributions in key areas!

- Advocates a pragmatic approach building on authors' earlier efforts in DSAP

Plan for Discussion:

1. Three Ways to Read the Paper
2. Comment 1. Set of Benchmarks and Candidate Models
3. Comment 2. Can text-based models do better?

Three Ways to Read the Paper

(1) Paper in Context: AI/ML in Finance

Three Clusters of Papers:

1. **Information extraction**, in which existing AI/ML tools can be used to extract information from sources that are traditionally not easy to parse

The Costs of Housing Regulation: Evidence from Generative Regulatory Measurement

Alexander W. Bartik, Arpit Gupta, and Daniel Milo*

August 19, 2025

Abstract

We introduce "generative regulatory measurement," using Large Language Models to interpret administrative documents with 96% accuracy in binary classification and 0.87 correlation for continuous questions. Our analysis of U.S. zoning regulations reveals four facts: (1) Housing regulations are multidimensional with two main principal components. (2) The first principal component represents *value capture* in high housing demand areas. (3) The second principal component associates with *exclusionary zoning*, increasing housing costs and socioeconomic exclusion. (4) Zoning follows a monocentric pattern with regional variations and is especially strict in Northeast suburbs. We develop a model of municipal regulatory choice consistent with these facts.

(1) Paper in Context: AI/ML in Finance

Three Clusters of Papers:

1. **Information extraction**, in which existing AI/ML tools can be used to extract information from sources that are traditionally not easy to parse

CEO Stress, Aging, and Death

MARK BORGSCHULTE, MARIUS GUENZEL, CANYAO LIU, and ULRIKE MALMENDIER*

ABSTRACT

We assess the long-term effects of managerial stress on aging and mortality. Using a difference-in-differences design, we apply neural network–based machine-learning techniques to CEOs' facial images and show that exposure to industry distress shocks during the Great Recession produces visible signs of aging. We estimate a one-year increase in "apparent" age. Moreover, using data on CEOs since the mid-1970s, we estimate a 1.1-year decrease in life expectancy after an industry distress shock, but a two-year increase when antitakeover laws insulate CEOs from market discipline. The estimated health costs are significant, both in absolute terms and relative to other health risks.

(1) Paper in Context: AI/ML in Finance

Three Clusters of Papers:

1. **Information extraction**, in which existing AI/ML tools can be used to extract information from sources that are traditionally not easy to parse
2. **Modelling complex relationships**, where non-linearities can be flexibly integrated

Empirical Asset Pricing via Machine Learning*

Shihao Gu
Booth School of Business, University of Chicago

Bryan Kelly
Yale University, AQR Capital Management, and NBER

Dacheng Xiu
Booth School of Business, University of Chicago

We perform a comparative analysis of machine learning methods for the canonical problem of empirical asset pricing: measuring asset risk premiums. We demonstrate large economic gains to investors using machine learning forecasts, in some cases doubling the performance of leading regression-based strategies from the literature. We identify the best-performing methods (trees and neural networks) and trace their predictive gains to allowing nonlinear predictor interactions missed by other methods. All methods agree on the same set of dominant predictive signals, a set that includes variations on momentum, liquidity, and volatility. (*JEL* C52, C55, C58, G0, G1, G17)

(1) Paper in Context: AI/ML in Finance

Three Clusters of Papers:

1. **Information extraction**, in which existing AI/ML tools can be used to extract information from sources that are traditionally not easy to parse
2. **Modelling complex relationships**, where non-linearities can be flexibly integrated

Machine-learning the skill of mutual fund managers ☆

Ron Kaniel [a,b,c], Zihan Lin [d], Markus Pelger [e], Stijn Van Nieuwerburgh [f,*]

[a] *Simon School of Business, Rochester University, 300 Wilson Blvd, Rochester, NY 14620 USA*
[b] *Fanhai International School of Finance, Fudan, Shanghai 200001, China*
[c] *Arison School of Business, Reichman University, Herzliya 4610101, Israel*
[d] *Institute for Computational and Mathematical Engineering, Stanford University, 475 Via Ortega, Stanford, CA 94305, USA*
[e] *Department of Management Science and Engineering, Stanford University, 475 Via Ortega, Stanford, CA 94305, USA*
[f] *Columbia Business School, 665 West 130 Street, New York, NY 10027l, USA*

ARTICLE INFO

Article history:
Received 10 April 2023
Revised 27 July 2023
Accepted 31 July 2023
Available online 11 August 2023

JEL classification:
G11
G12
G17
G23
C45

ABSTRACT

We show, using machine learning, that fund characteristics can consistently differentiate high from low-performing mutual funds, before and after fees. The outperformance persists for more than three years. Fund momentum and fund flow are the most important predictors of future risk-adjusted fund performance, while characteristics of the stocks that funds hold are not predictive. Returns of predictive long-short portfolios are higher following a period of high sentiment. Our estimation with neural networks enables us to uncover novel and substantial interaction effects between sentiment and both fund flow and fund momentum.

© 2023 Elsevier B.V. All rights reserved.

(1) Paper in Context: AI/ML in Finance

Three Clusters of Papers:

1. **Information extraction**, in which existing AI/ML tools can be used to extract information from sources that are traditionally not easy to parse
2. **Modelling complex relationships**, where non-linearities can be flexibly integrated
3. **Fair Benchmarks**, where AI/ML-generated forecasts can be considered "rational" or "fair" benchmarks to which human judgments can then be evaluated

Belief Distortions and Macroeconomic Fluctuations†

By Francesco Bianchi, Sydney C. Ludvigson, and Sai Ma*

This paper combines a data-rich environment with a machine learning algorithm to provide new estimates of time-varying systematic expectational errors ("belief distortions") embedded in survey responses. We find sizable distortions even for professional forecasters, with all respondent-types overweighting the implicit judgmental component of their forecasts relative to what can be learned from publicly available information. Forecasts of inflation and GDP growth oscillate between optimism and pessimism by large margins, with belief distortions evolving dynamically in response to cyclical shocks. The results suggest that artificial intelligence algorithms can be productively deployed to correct errors in human judgment and improve predictive accuracy. (*JEL* C45, D83, E23, E27, E31, E32, E37)

(1) Paper in Context: AI/ML in Finance

Three Clusters of Papers:

1. **Information extraction**, in which existing AI/ML tools can be used to extract information from sources that are traditionally not easy to parse
2. **Modelling complex relationships**, where non-linearities can be flexibly integrated
3. **Fair Benchmarks**, where AI/ML-generated forecasts can be considered “rational” or “fair” benchmarks to which human judgments can then be evaluated

Predictably Unequal? The Effects of Machine Learning on Credit Markets

ANDREAS FUSTER, PAUL GOLDSMITH-PINKHAM, TARUN RAMADORAI, and ANSGAR WALTHER

ABSTRACT

Innovations in statistical technology in functions including credit-screening have raised concerns about distributional impacts across categories such as race. Theoretically, distributional effects of better statistical technology can come from greater flexibility to uncover structural relationships or from triangulation of otherwise excluded characteristics. Using data on U.S. mortgages, we predict default using traditional and machine learning models. We find that Black and Hispanic borrowers are disproportionately less likely to gain from the introduction of machine learning. In a simple equilibrium credit market model, machine learning increases disparity in rates between and within groups, with these changes attributable primarily to greater flexibility.

(1) Paper in Context: AI/ML in Finance

This paper joins a fourth category ("**intellectual transfer**") – a relatively smaller set of papers that borrows core ideas from AI/ML and applies them to economics and finance.

Some examples:

High-Frequency Expectations from Asset Prices: A Machine Learning Approach*

Aditya Chaudhry† Sangmin S. Oh‡

September 16, 2020. Comments welcome.

Abstract

We propose a novel reinforcement learning approach to extract high-frequency aggregate growth expectations from asset prices. While much expectations-based research in macroeconomics and finance relies on low-frequency surveys, the multitude of events that pass between survey dates renders identification of causal effects on expectations difficult. Our method allows us to construct a daily time-series of the cross-sectional mean of a panel of GDP growth forecasts. The high-frequency nature of our series enables clean identification in event studies. In particular, we use our estimated daily growth expectations series to test the "Fed information effect" and find little evidence to support its existence. Extensions of our framework can obtain daily expectations series of any macroeconomic variable for which a low-frequency panel of forecasts is available. In this way, our method provides a sharp empirical tool to advance understanding of how expectations are formed.

AI-Powered Trading, Algorithmic Collusion, and Price Efficiency

Winston Wei Dou Itay Goldstein Yan Ji *

July 14, 2025

Abstract

The integration of algorithmic trading with reinforcement learning, termed AI-powered trading, is transforming financial markets. Alongside the benefits, it raises concerns for collusion. This study first develops a model to explore the possibility of collusion among informed speculators in a theoretical environment. We then conduct simulation experiments, replacing the speculators in the model with informed AI speculators who trade based on reinforcement-learning algorithms. We show that they autonomously sustain collusive supra-competitive profits without agreement, communication, or intent. Such collusion undermines competition and market efficiency. We demonstrate that two separate mechanisms are underlying this collusion and characterize when each one arises.

This Paper:

- Show that hierarchical relationship between words and paragraphs in language models is directly analogous to relationship between assets and portfolios,
- Operationalize this insight in the form of asset embeddings.

(2) Paper in Context: R^2 in Asset Pricing

Link to podcast

THE JOURNAL OF FINANCE • VOL. XLIII, NO. 2 • JULY 1988

R^2

RICHARD ROLL*

ABSTRACT

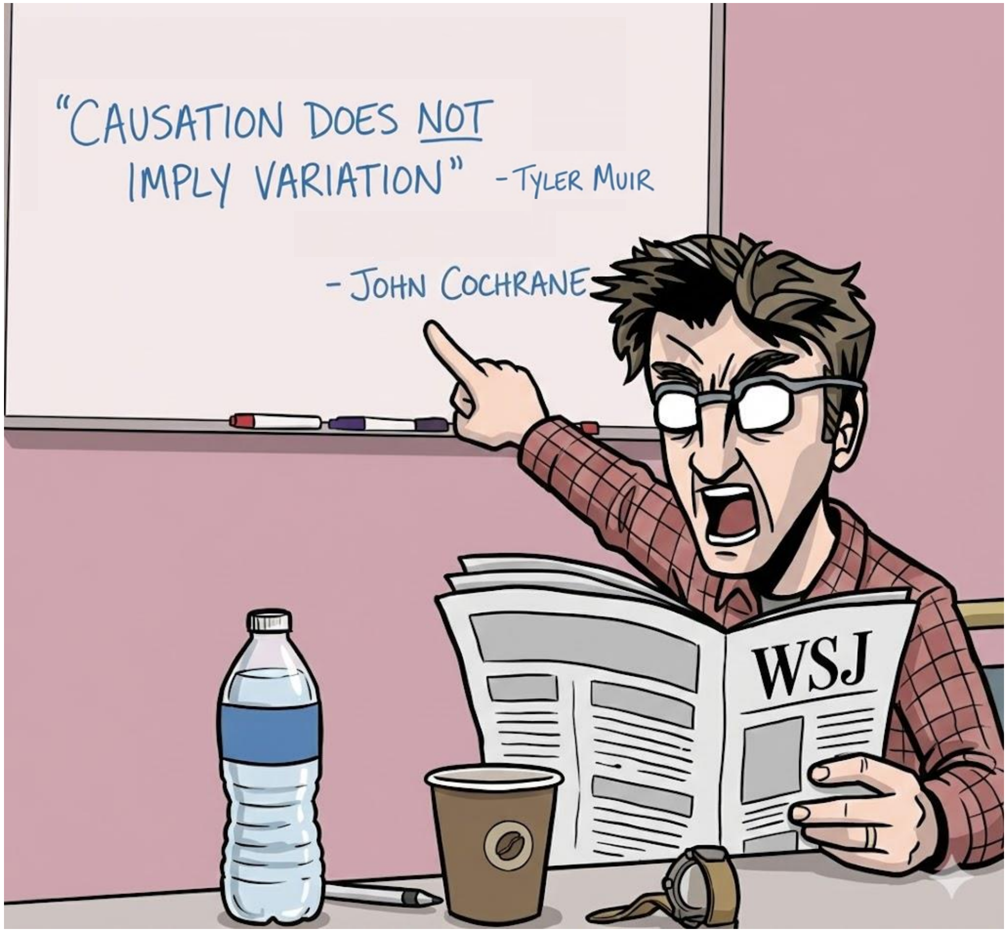
Even with hindsight, the ability to explain stock price changes is modest. R^2s were calculated for the returns of large stocks as explained by systematic economic influences, by the returns on other stocks in the same industry, and by public firm-specific news events. The average adjusted R^2 is only about .35 with monthly data and .20 with daily data. There is little relation between explanatory power and either the firm's size or its industry. There is little improvement in R^2 from eliminating *all* dates surrounding news reports in the financial press. However, the sample kurtosis is quite different when such news events are eliminated, thereby revealing a mixture of return distributions. Non-news dates also indicate the presence of a distributional mixture, perhaps due to traders acting on private information.

(2) Paper in Context: R^2 in Asset Pricing

The Grumpy Economist

Causation Does not Imply Variation

JOHN H. COCHRANE
NOV 09, 2025



(2) Paper in Context: R^2 in Asset Pricing

Asset pricing has a rich history of quantifying asset price variation by going beyond structural equilibrium models:

- Variance Decomposition and Present Value Identities
  - Campbell and Shiller (1988), Campbell (1991), Vuolteenaho (2002)
  - Shiller (1981), LeRoy and Porter (1981)
- Reduced-Form Factor Models and Statistical Approaches
  - Ross (1976)
  - Fama and French (1992, 1993, 2015), Carhart (1997)
  - Kozak, Nagel, and Santosh (2018, 2020), Kelly, Pruitt, and Su (2020)
- Dynamic Term Structure Models
  - Vasicek (1977), Cox, Ingersoll, and Ross (1985)
  - Litterman and Scheinkman (1991)
- Demand System Asset Pricing (DSAP)

This Paper:

Proof of concept on how we can use tools from AI/ML to systematically extract useful information from portfolio holdings to explain the "variance" in P, ΔP, and Q

(3) Paper raises deep questions for academic research

Paper also pushes the reader to confront several foundational questions about how to approach research in asset pricing.

What constitutes a good model?

- If useful representations can be learned, then model quality maybe hinges less on parametric elegance and more on whether the representation captures economically meaningful structure

How should we evaluate models?

- Out-of-sample performance on benchmarks that we deem relevant for asset pricing

Which information is valuable to the econometrician?

- Holdings as a “sufficient statistic” for investor beliefs, constraints, and preferences

What is the goal of our profession?

- Develop empirically successful models that recover economically meaningful structure hidden in data and generate reliable predictions about key asset pricing quantities

This Paper:

A friendly invitation to revisit the criteria by which we judge models, evidence, and progress in our profession.

Comment 1. Set of Benchmarks and Candidate Models

Which benchmarks?

Authors introduce three benchmarks – an evaluation suite for asset embeddings

1. Valuations: Do embeddings explain valuations in the cross-section?

    Prediction of M/B using embeddings in the cross-section

2. Returns: Do embeddings explain realized returns in the cross-section?

    Prediction of stock returns using embeddings in the cross-section

3. Holdings: Can embeddings infer missing assets in portfolios?

    Prediction of the missing asset identity using embeddings

Excellent starting point to think about the set of benchmarks. Authors also write:

> We could imagine a quarterly competition to evaluate new models on these benchmarks, just like the ImageNet Large Scale Visual Recognition Challenge

This mirrors the practice in AI!

- Stable, repeatable tasks against which new models can be compared real-time

Complementary to traditional asset pricing benchmarks

Table I

The Equity Term Structure: Stylized Facts versus Theory

The equity term premium $E_t[TP_t] = E_t[r_{t+1}^{long} - r_{t+1}^{short}]$ is the conditional expected annual return to long-maturity equity minus the annual return to short maturity equity. The cyclicality of the equity term premia is measured by linear projections of the realized term premium on the ex ante dividend-price ratio of the market portfolio. The cyclicality of yield spreads is similarly measured as the linear projection of yield spreads on the contemporaneous dividend-price ratio. The habit model refers to the Campbell and Cochrane (1999) model. The long-run risk model refers to the Bansal and Yaron (2004) model.

	Average Slope	Cyclicality of Term Premia	Cyclicality of Yield Spread
Paper	Binsbergen, Brandt, and Koijen (2012)	This Paper	Binsbergen et al. (2013)
Data Measured as	$TP = E[r_{t+1}^{long} - r_{t+1}^{short}]$	$TP_t = \beta_0 + \beta_1 D_t/P_t$	$YS_t = \theta_0 + \theta_1 D_t/P_t + e_t$
Result	Downward sloping $E[TP] < 0$	Countercyclical $\beta_1 > 0$	Procyclical $\theta_1 < 0$
Theories			
Habit	Upward	Countercyclical	Countercyclical
Long-run risk	Upward	Countercyclical	Procyclical
Lettau and Wacther (2007)	Downward	Procyclical	Procyclical
Gabaix (2012)	Flat	Constant	Procyclical
Hasler and Marfe (2016)	Downward	Procyclical	Procyclical
Ai et al. (2018)	Downward	Procyclical	Procyclical
This paper	Downward	Countercyclical	Procyclical

Gormsen (2021), “Time Variation of the Equity Term Structure”

Benchmarking the Benchmarks in AI

Benchmark	Description		Gemini 3 Pro	Gemini 2.5 Pro	Claude Sonnet 4.5	GPT-5.1
Humanity's Last Exam	Academic reasoning	No tools	**37.5%**	21.6%	13.7%	26.5%
		With search and code execution	**45.8%**	—	—	—
ARC-AGI-2	Visual reasoning puzzles	ARC Prize Verified	**31.1%**	4.9%	13.6%	17.6%
GPQA Diamond	Scientific knowledge	No tools	**91.9%**	86.4%	83.4%	88.1%
AIME 2025	Mathematics	No tools	**95.0%**	88.0%	87.0%	94.0%
		With code execution	**100%**	—	**100%**	—
MathArena Apex	Challenging Math Contest problems		**23.4%**	0.5%	1.6%	1.0%
MMMU-Pro	Multimodal understanding and reasoning		**81.0%**	68.0%	68.0%	76.0%
ScreenSpot-Pro	Screen understanding		**72.7%**	11.4%	36.2%	3.5%
CharXiv Reasoning	Information synthesis from complex charts		**81.4%**	69.6%	68.5%	69.5%
OmniDocBench 1.5	OCR	Overall Edit Distance, lower is better	**0.115**	0.145	0.145	0.147
Video-MMMU	Knowledge acquisition from videos		**87.6%**	83.6%	77.8%	80.4%
LiveCodeBench Pro	Competitive coding problems from Codeforces, ICPC, and IOI	Elo Rating, higher is better	**2,439**	1,775	1,418	2,243
Terminal-Bench 2.0	Agentic terminal coding	Terminus-2 agent	**54.2%**	32.6%	42.8%	47.6%
SWE-Bench Verified	Agentic coding	Single attempt	76.2%	59.6%	**77.2%**	76.3%
τ2-bench	Agentic tool use		**85.4%**	54.9%	84.7%	80.2%
Vending-Bench 2	Long-horizon agentic tasks	Net worth (mean), higher is better	**$5,478.16**	$573.64	$3,838.74	$1,473.43
FACTS Benchmark Suite	Held out internal grounding, parametric, MM, and search retrieval benchmarks		**70.5%**	63.4%	50.4%	50.8%
SimpleQA Verified	Parametric knowledge		**72.1%**	54.5%	29.3%	34.9%
MMMLU	Multilingual Q&A		**91.8%**	89.5%	89.1%	91.0%
Global PIQA	Commonsense reasoning across 100 Languages and Cultures		**93.4%**	91.5%	90.1%	90.9%
MRCR v2 (8-needle)	Long context performance	128k (average)	**77.0%**	58.0%	47.1%	61.6%
		1M (pointwise)	**26.3%**	16.4%	not supported	not supported

For details on our evaluation methodology please see deepmind.google/models/evals-methodology/gemini-3-pro

How should we choose the benchmarks?

Benchmarks in the AI community evolve to meet user needs

Coding benchmarks for developers

Instruction-following benchmarks for everyday users

Math benchmarks for mathematicians

Who are the “users” of asset embeddings, and what benchmarks should serve them?

1. Industry Participants

- Example: Predict replacement assets in tax loss harvesting

2. Regulators

- Example: Predict “runs” and flight-to-safety (e.g. March 2020, UK LDI)

3. Academics

- Example: Return co-movement on identified “policy days” (e.g. FOMC, CPI)

Suggestion #1. High-level categorization of the benchmarks based on user needs

- Would provide the reader a more structured way to think about next generation of benchmarks

Which candidate models?

Authors evaluate broadly four categories of models:

1. Characteristics-based Models: (1) Beta, (2) Beta + three/five characteristics
2. Recommender Systems with varying information sets and assumptions about shorts
    - Binary information sets
    - Percentile ranks
    - Log dollar holdings
3. Word2Vec ordered by investors’ portfolio weights
4. Transformers trained on (1) portfolio shares and (2) ownership shares

Which candidate models?

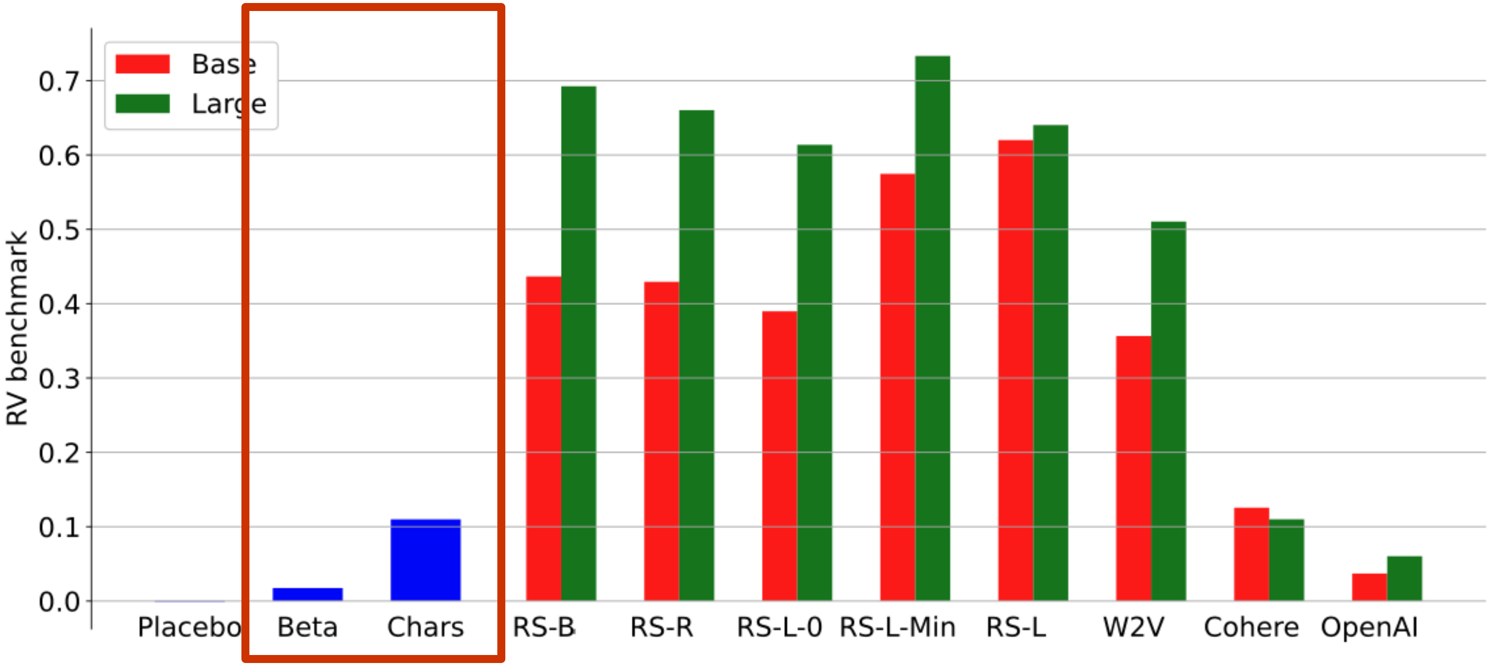


Figure 7. Text-Based Embeddings in the Relative Valuation Benchmark. The dimension of the text-based embeddings from Cohere and OpenAI are reduced to match the base (four-dimensional) and large (ten-dimensional) models of asset embeddings. The sample period is 2022.Q4.

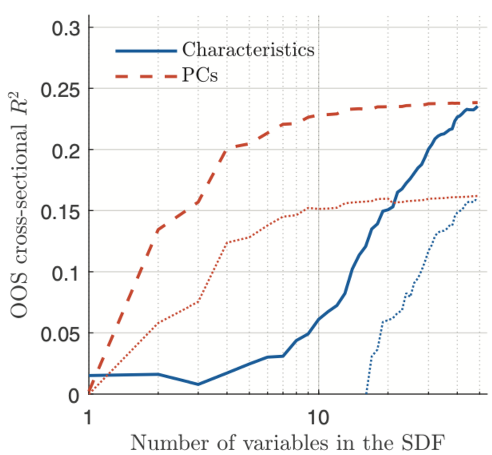
Which candidate models?

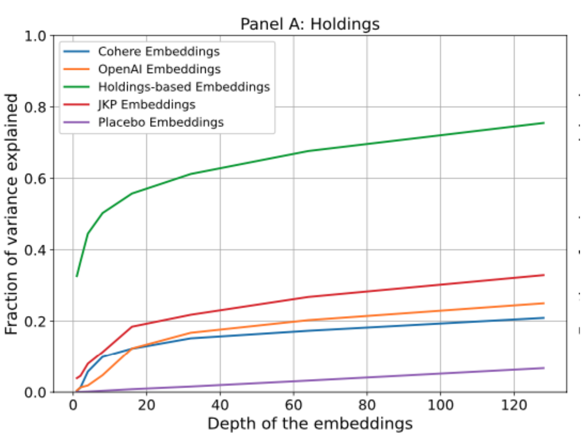
Authors evaluate broadly four categories of models:

1. Characteristics-based Models: (1) Beta, (2) Beta + three/five characteristics
2. Recommender Systems with varying information sets and assumptions about shorts
   - Binary information sets
   - Percentile ranks
   - Log dollar holdings
3. Word2Vec ordered by investors' portfolio weights
4. Transformers trained on (1) portfolio shares and (2) ownership shares

The current set of models omits another candidate model: PCs of firm characteristics

- Kozak, Nagel, and Santosh (2020): The first few PCs of characteristics capture a large share of cross-sectional variation in expected returns





Which candidate models?

Authors evaluate broadly four categories of models:

1. Characteristics-based Models: (1) Beta, (2) Beta + three/five characteristics
2. Recommender Systems with varying information sets and assumptions about shorts
   - Binary information sets
   - Percentile ranks
   - Log dollar holdings
3. Word2Vec ordered by investors' portfolio weights
4. Transformers trained on (1) portfolio shares and (2) ownership shares

The current set of models omits another candidate model: PCs of firm characteristics

- Kozak, Nagel, and Santosh (2020): The first few PCs of characteristics capture a large share of cross-sectional variation in expected returns

Suggestion #2. Add a KNS2020 Sparse PC baseline to the list of evaluated models

- Would be instructive to see how well KNS PCs perform given it's the best linear compression of anomaly portfolios

Which candidate models?

Authors evaluate broadly four categories of models:

1. Characteristics-based Models: (1) Beta, (2) Beta + three/five characteristics
2. Recommender Systems with varying information sets and assumptions about shorts
   - Binary information sets
   - Percentile ranks
   - Log dollar holdings
3. Word2Vec ordered by investors' portfolio weights
4. Transformers trained on (1) portfolio shares and (2) ownership shares

Also, note that different models use different amounts of economic information.

- The dimensionality of embeddings is fixed, but the amount of information is different
- This is intentional – the whole point is to highlight the wealth of information in holdings!
- Still useful to compare the value of (i) holdings vs. (ii) accounting vs. (iii) news (text)

Question. Is there a way to compute the "value of information" in this framework?

If I have a $1, should I use it to buy information on (1) NVIDIA's data center capex, (2) how much Korean investors hold NVIDIA, or (3) Jensen Huang's exclusive interview?

- Baley and Veldkamp (2025) could provide a useful starting point.

Comment 2. Can Text-based Models Do Better?

Text-based Embeddings

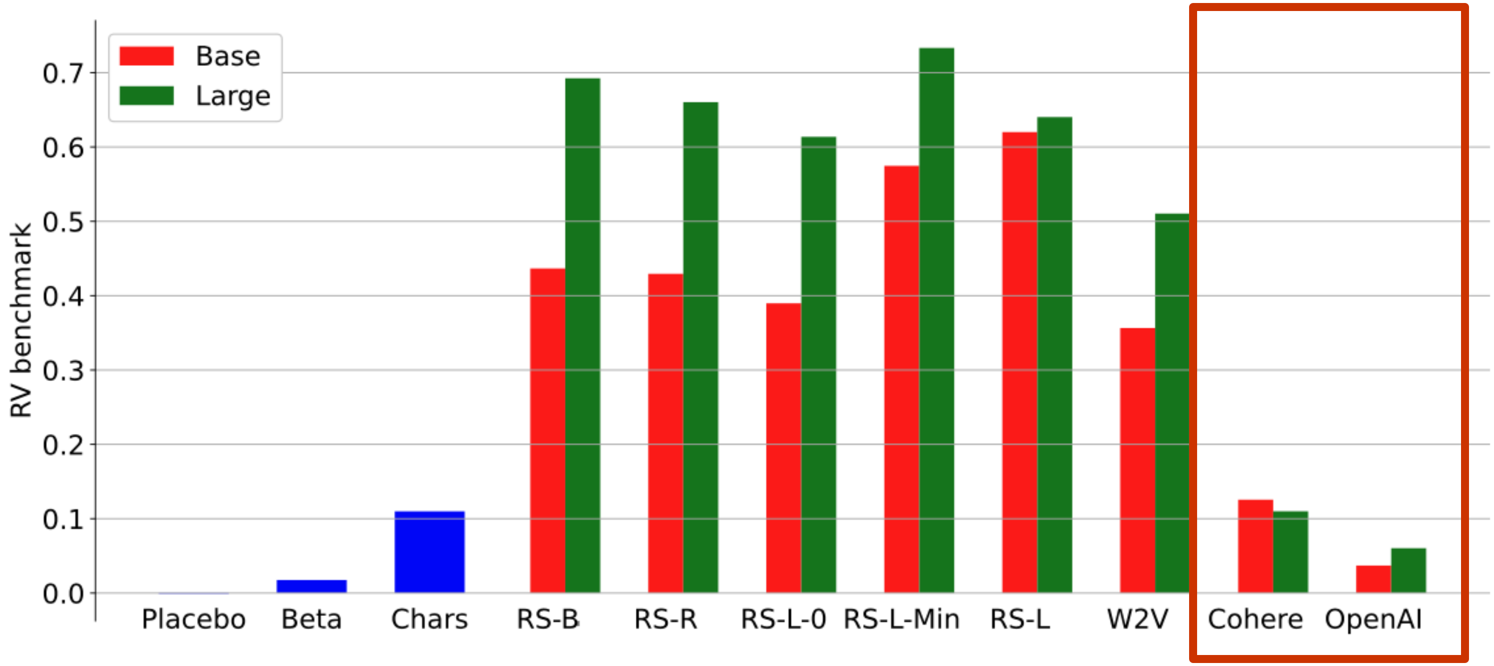


Figure 7. Text-Based Embeddings in the Relative Valuation Benchmark. The dimension of the text-based embeddings from Cohere and OpenAI are reduced to match the base (four-dimensional) and large (ten-dimensional) models of asset embeddings. The sample period is 2022.Q4.

Text-based Embeddings

"NVIDIA CORPORATION"

↓

Cohere, OpenAI, …

Embedding Model

↓

1024-dimensional embedding

[0.122, -0.357, 0.486, 0.279, -0.577, -0.993, 0.407….]

↓

4-dimensional embedding

[0.356, 0.481, -0.374, 0.985]

↓

Benchmark Tasks

- Predict valuations
- Predict returns
- Predict holdings

…

An Alternate Approach

"NVIDIA CORPORATION"

↓

Large Language Model

↓

1024-dimensional embedding

[0.122, -0.357, 0.486, 0.279, -0.577, -0.993, 0.407….]

↓

4-dimensional embedding

[0.356, 0.481, -0.374, 0.985]

↓

Benchmark Tasks

- Predict valuations
- Predict returns
- Predict holdings

…

An Alternate Approach

"NVIDIA CORPORATION"

↓

Large Language Model

↓

Benchmark Tasks

- Predict valuations
- Predict returns
- Predict holdings

…

An Alternate Approach

Investor description

Pershing Square Capital Management is known for concentrated, long-term strategy specializing in fundamental value investing…

Asset description

NVIDIA Corporation is a global semiconductor and computing company specializing in …

Large Language Model

Benchmark Tasks

- Predict valuations
- Predict returns
- Predict holdings

…

An Alternate Approach

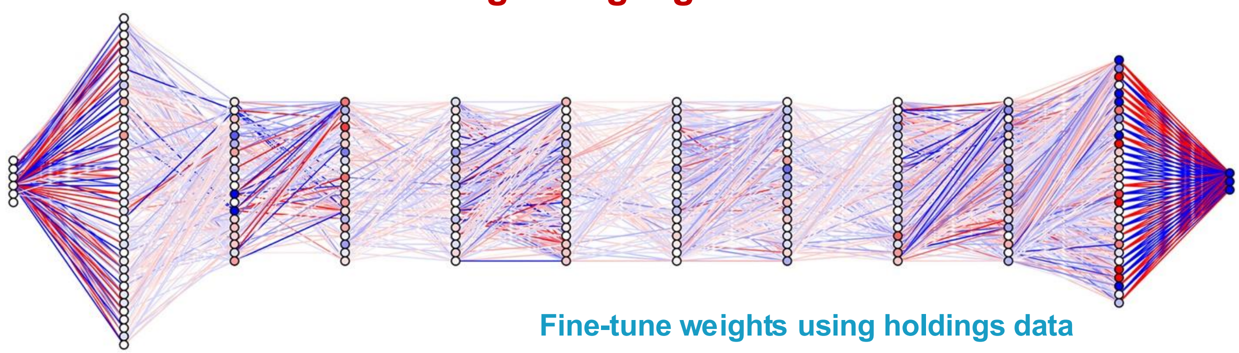
Investor description

Pershing Square Capital Management is known for concentrated, long-term strategy specializing in fundamental value investing…

Asset description

NVIDIA Corporation is a global semiconductor and computing company specializing in …

Large Language Model



Fine-tune weights using holdings data

Benchmark Tasks

- Predict valuations
- Predict returns
- Predict holdings

…

An Alternate Approach

Why might this be useful?

1. A model that "knows" a lot may do better with slightly more context.

- Industry linkages, supply-chain relationships, corporate strategies, risk narratives
- Investor philosophies, sensitivities to events

2. Preliminary evidence of success in predicting market shares (in the context of marketing)

- If LLMs can learn consumer demand, maybe they can learn investor demand

3. Compressing assets into low-dimensional embedding may collapse too many economically relevant distinctions.

Ultimately, we'd like to integrate accounting + holdings + text information.

The question is what is the optimal combination.

Suggestion #3

It would be useful to consider the performance of a text-based transformer (+ holdings-based fine-tuning) that works on text of investor and asset descriptions.

Final Thoughts: When Does the LLM Analogy Break Down?

1. What constitutes a "high-quality" dataset in this setting?
   - In NLP, concerns about web-scale noise led to a curation of corpora

2. What is the "grammar" of portfolio management?
   - Grammar in text provides compositional structure (e.g. subject-verb-object)
   - In language, speakers share syntax; do managers also share syntax?

3. Scope for Reinforcement Learning from Human Feedback (RLHF)
   - RLHF: Model generates output, humans rank or evaluate those outputs, then the model is optimized to prefer outputs that humans judged as better

4. NLP benefits from seeing incorrect grammar ("I goed").
   - Do we have "forbidden pairs" in portfolios?

Final Thoughts

- Deep, paradigm-shifting paper with multiple core contributions in key areas
- A fundamentally pragmatic approach that builds on authors' earlier efforts in DSAP

- **Punchline:** With the right tools, the plethora of information in holdings can be smartly leveraged to make important progress on the most important Qs in asset pricing.

- **A few questions prompted by the paper for the future:**
  - ML embraces the idea that no single model performs best across all benchmarks. Should we do the same?
  - When does the LLM analogy break down?

Everyone should read it…

… or at least discuss it with an LLM.