Asset Embeddings

Xavier Gabaix Ralph Koijen Robert Richmond Motohiro Yogo

Harvard - Chicago - NYU - Princeton

May 2025

Identifying similar firms

- In economics, we often try to find similar firms or assets.
  - E.g., similar growth rates, expected returns, risk, asset substitution, product markets, ...
- Common practice: Use observable characteristics.
  - E.g., industry definitions, accounting data, ...

Identifying similar firms

- In economics, we often try to find similar firms or assets.
  - E.g., similar growth rates, expected returns, risk, asset substitution, product markets, ...
- Common practice: Use observable characteristics.
  - E.g., industry definitions, accounting data, ...
- Those characteristics may be quite imperfect.
  - Standardized accounting data are an incomplete summary.
    - E.g., number of subscribers at Netflix, ...
  - New economic environments call for creative, new characteristics.
    - E.g., exposure to COVID-19, intangibles or AI.

Identifying similar firms

- In economics, we often try to find similar firms or assets.
  - E.g., similar growth rates, expected returns, risk, asset substitution, product markets, ...
- Common practice: Use observable characteristics.
  - E.g., industry definitions, accounting data, ...
- Those characteristics may be quite imperfect.
  - Standardized accounting data are an incomplete summary.
    - E.g., number of subscribers at Netflix, ...
  - New economic environments call for creative, new characteristics.
    - E.g., exposure to COVID-19, intangibles or AI.
- This paper: Use asset embeddings to measure firm similarity.

What are embeddings?

- Embeddings: Represent data (e.g., words) as vectors in a potentially high-dimensional space: $x_a \in \mathbb{R}^K$.
- Embeddings play a central role in the development of large language models (LLMs).
- In LLMs, embeddings capture the similarity between words and it allows us to do "math with words:"

$$x_{\text{Paris}} - x_{\text{France}} + x_{\text{Spain}} \simeq x_{\text{Madrid}}.$$

What are embeddings?

- Embeddings: Represent data (e.g., words) as vectors in a potentially high-dimensional space: $x_a \in \mathbb{R}^K$.
- Embeddings play a central role in the development of large language models (LLMs).
- In LLMs, embeddings capture the similarity between words and it allows us to do "math with words:"

$$x_{\text{Paris}} - x_{\text{France}} + x_{\text{Spain}} \simeq x_{\text{Madrid}}.$$

- Embedding vectors are learned from (lots of) data (not preselected).
- Despite the success of embedding techniques in these fields, their application in finance and economics largely unexplored.

Ideal data to estimate embeddings?

- We introduce the concept of asset embeddings.
  - A vector representation for each asset, that we learn from data.
- Which data to use?

Ideal data to estimate embeddings?

- We introduce the concept of asset embeddings.
  - A vector representation for each asset, that we learn from data.
- Which data to use?
- Our answer: Just like
  - documents organize words in language modeling,
  - images organize pixels in computer vision,
  - songs organize notes in audio,

  investors organize assets in finance and economics.

Ideal data to estimate embeddings?

- We introduce the concept of asset embeddings.
  - A vector representation for each asset, that we learn from data.
- Which data to use?
- Our answer: Just like
  - documents organize words in language modeling,
  - images organize pixels in computer vision,
  - songs organize notes in audio,

  investors organize assets in finance and economics.
- Theoretically, we show how embeddings can be recovered by "inverting the asset demand system."

WHICH METHOD TO LEARN EMBEDDINGS?

- Which method to use?

Which method to learn embeddings?

- Which method to use?
- Traditional approach: LSA (Latent Semantic Analysis), which is related to PCA/recommender systems.
- The recent ML/AI literature went way beyond that.
  - Context-invariant embeddings: E.g., GloVe and Word2Vec.
  - Embeddings with context: E.g., transformer models (e.g., BERT and GPT).
  - Parameters are estimated using masked language modeling.

INVESTOR EMBEDDINGS

- Even though our focus is on asset embeddings, we obtain investor embeddings as a by-product: $\lambda_i \in \mathbb{R}^K$.
  - Learned vector representations of each investor's "taste for characteristics"

Investor embeddings

- Even though our focus is on asset embeddings, we obtain investor embeddings as a by-product: $\lambda_i \in \mathbb{R}^K$.
  - Learned vector representations of each investor's "taste for characteristics"
- Examples of applications:
  - Classify investors beyond institutional type, size, and activeness.
  - Identify crowded trades.
  - Performance measurement (extending Daniel, Grinblatt, Titman, and Wermers, 1997).

Five main contributions

1. Micro-found the use of holdings data as embeddings data.
2. Three benchmarks to compare asset embedding models.
   - Building on the success of benchmark in AI (e.g., ImageNet).
3. Explore different modeling architectures to learn asset embeddings based on language models.
4. Evaluate benchmarks for asset embeddings, text-based embeddings, and observed characteristics.
5. Use earnings calls data to interpret the embeddings.
   - Extends to any other form of text data (e.g., WSJ articles, analyst reports, ...).

Related literature

- Demand system asset pricing.
  - Frameworks to jointly understand prices, characteristics, and holdings data.
- Machine learning and asset pricing.
  - Use (lots of) observable characteristics and price-based variables to predict future returns and risk.
  - Recent literature explores information in text data.
    - Newspapers, 10-K filings, earnings calls, social media, ...
    - E.g., Hassan et al. '19, Bryzgalova et al. '24, and Bybee '24.
  - See Kelly and Xiu (2023) for a recent review.
- Audio, language, and computer vision models.

Outline

- Inverting the asset demand system: Using holdings data as embeddings data.
- Methods to estimate embeddings.
- Data.
- Benchmarking asset embeddings.
- Empirical results.

Holdings data as embeddings data

- Model the log dollar holdings of investor i in asset (i.e. stock) a as

$$h_{ia} = c_i^h + (1 - \zeta_i) p_a + \nu_{ia},$$

where ζ_i is the demand elasticity and ν_{ia} a stock-specific demand shifter.

Holdings data as embeddings data

- Model the log dollar holdings of investor i in asset (i.e. stock) a as

$$h_{ia} = c_i^h + (1 - \zeta_i) p_a + \nu_{ia},$$

  where ζ_i is the demand elasticity and ν_{ia} a stock-specific demand shifter.

- We model the demand shifter as

$$\nu_{ia} = \lambda_i^{\nu\prime} x_a + u_{ia},$$

  which can be micro-founded by (Koijen and Yogo, 2019):
  - Investors having mean-variance demand.
  - Returns follow a factor model.
  - Expected returns and factor loadings are affine in x_a.

Holdings data as embeddings data

- A log-linear approximation to the market clearing condition implies that the log price of asset a is:

$$p_a = c^p + \frac{1}{\zeta_S}\lambda_S^{\nu\prime} x_a + u_{Sa},$$

with $y_S \equiv \sum_i S_i^a y_i$.

Holdings data as embeddings data

- A log-linear approximation to the market clearing condition implies that the log price of asset a is:

$$p_a = c^p + \frac{1}{\zeta_S}\lambda_S^{\nu\prime} x_a + u_{Sa},$$

with $y_S \equiv \sum_i S_i^a y_i$.

- If we substitute the price back into the demand equation:

$$h_{ia} = \phi_i^h + \phi_a^h + \lambda_i' x_a + \epsilon_{ia},$$

where λ_i are the investor embeddings.

- We can also estimate the model in terms of rebalancing.

Methods to extract embeddings

- We consider the following embedding models:
  1. Recommender systems.
  2. Shallow neural networks: Word2Vec.
  3. Models with attention: Transformer models.
     - We build on the BERT architecture and specialize it to holdings data.

Recommender systems

- Recommender systems, with $\theta = (x_a, \lambda_i, \delta_i, \delta_a)$,

$$\min_{\theta} \frac{1}{IA} \sum_{i,a} (h_{ia} - \delta_i - \delta_a - \lambda_i' x_a)^2 + \frac{\xi}{IK} \sum_i \lambda_i' \lambda_i + \frac{\xi}{AK} \sum_a x_a' x_a,$$

  where
  - h_{ia}: Log holdings.
  - x_a: Asset embeddings.
  - λ_{iq}: Investor embeddings.
- Analogous to LSA in the NLP literature.[1]

[1] Dumais, Furnas, Landauer, and Deerwester (1988).

Implementations of recommender systems

- To understand how to best extract information from holdings, we consider five variants:
  1. Binary, $\mathbb{I}_{H_{ia}>0}$.
  2. Percentile ranks of H_{ia} with missing values set to zero.
  3. h_{ia} with missing values set to zero.
  4. h_{ia} with missing values set to the smallest active position.
  5. h_{ia} using only the non-missing values.

Word2Vec

- General approach to estimate language models, such as Word2Vec,[2]
  - Task: Guess masked words.
    - E.g. "Please pass me the _____ and pepper".
  - Use a context window to maximize the probability of a missing word given the context info:

$$\mathbb{P}(w_a \mid w_c) = \frac{\exp(x_a' x_c)}{\sum_b \exp(x_b' x_c)}.$$

[2] Mikolov, Sutskever, Chen, Corrado, Dean (2013a, b).

Word2Vec

- General approach to estimate language models, such as Word2Vec,[2]
  - Task: Guess masked words.
    - E.g. "Please pass me the _____ and pepper".
  - Use a context window to maximize the probability of a missing word given the context info:

$$\mathbb{P}(w_a \mid w_c) = \frac{\exp(x_a' x_c)}{\sum_b \exp(x_b' x_c)}.$$

- Estimation using holdings data:
  - Sentences $\Rightarrow$ Investors.
  - Words $\Rightarrow$ Assets.
  - Objective: Guess masked assets (cross entropy).

[2] Mikolov, Sutskever, Chen, Corrado, Dean (2013a, b).

Masked asset modeling

- Example: The ARKK ETF in July 2023:

Holdings Data - ARKK
As of 07/07/2023

ARK INVEST

ARKK
ARK Innovation ETF

	Company	Ticker	CUSIP	Shares	Market Value ($)	Weight (%)
1	TESLA INC	**TSLA**	88160R101	3,496,872	$967,024,982.88	12.43%
2	COINBASE GLOBAL INC -CLASS A	**COIN**	19260Q107	7,945,138	$620,515,277.80	7.98%
3	ROKU INC	**ROKU**	77543R102	8,865,426	$546,110,241.60	7.02%
4	~~ZOOM VIDEO COMMUNICATIONS-A~~	~~**ZM**~~	~~98980L101~~	~~8,258,591~~	~~$534,248,251.79~~	~~6.87%~~
5	UIPATH INC - CLASS A	**PATH**	90364P105	28,152,366	$463,106,420.70	5.95%
6	BLOCK INC	**SQ**	852234103	7,069,493	$456,759,942.73	5.87%
7	EXACT SCIENCES CORP	**EXAS**	30063P105	4,031,264	$368,739,718.08	4.74%
8	UNITY SOFTWARE INC	**U**	91332U101	8,350,868	$338,627,697.40	4.35%
9	SHOPIFY INC - CLASS A	**SHOP**	82509L107	5,430,238	$335,751,615.54	4.32%
10	DRAFTKINGS INC-CL A	**DKNG UW**	26142V105	12,035,607	$303,658,364.61	3.90%

Context and self-attention: A simple example

- So far, we have one x_a per asset, say, Apple, with no context.
- How does attention[3] work?

[3] Vaswani, Shazeer, Parmar, Uszkoreit, Jones, Gomez, Kaiser, Polosukhin (2017).

Context and self-attention: A simple example

- So far, we have one x_a per asset, say, Apple, with no context.
- How does attention[3] work?

1. $\mathcal{H}_i$: Stocks in the portfolio of manager i.
2. For stock $a \in \mathcal{H}_i$, compute a similarity score with the other stocks $b \in \mathcal{H}_i$

$$\sigma_{ab} = x_a' x_b.$$

   x_a: Query.
   x_b: Key.

3. Compute the contextualized embedding, x_a^i,

$$x_a^i = \sum_{b \in \mathcal{N}_i} \frac{e^{\sigma_{ab}}}{\sum_{c \in \mathcal{N}_i} e^{\sigma_{ac}}} x_b.$$

   x_b: Value.

[3] Vaswani, Shazeer, Parmar, Uszkoreit, Jones, Gomez, Kaiser, Polosukhin (2017).

Self-attention: Example

- Suppose
$$x_a = \begin{bmatrix} x_{a1} \\ x_{a2} \\ x_{a3} \end{bmatrix},$$
where x_{aj} are sub-vectors capturing a firm's industry, reliance on external finance, and supply-chain risk.
- In each quarter, different parts of the embedding vector may be relevant depending on which stocks are held/traded together.
- Similarly, depending on the problem you are studying, you can construct controls depending on what features of firms are relevant in the context of your sample.

Generalizing attention: Transformers

- Transformer models generalize this idea.
  - Query: $q_a = W^Q x_a$.
  - Key: $k_a = W^K x_a$.
  - Value: $v_a = W^V x_a$.
- The contextualized embedding is then computed as

$$x_a^i = \sum_{b \in \mathcal{N}_i} \frac{e^{\sigma_{ab}}}{\sum_{c \in \mathcal{N}_i} e^{\sigma_{ac}}} v_b, \qquad \sigma_{ab} = q_a' k_b.$$

- The matrices W_Q, W_K, and W_V are learned from (lots of) data and determine which aspects of the context are important.

Generalizing attention: Transformers

- Transformer models generalize this idea.
  - Query: $q_a = W^Q x_a$.
  - Key: $k_a = W^K x_a$.
  - Value: $v_a = W^V x_a$.
- The contextualized embedding is then computed as

$$x_a^i = \sum_{b \in \mathcal{N}_i} \frac{e^{\sigma_{ab}}}{\sum_{c \in \mathcal{N}_i} e^{\sigma_{ac}}} v_b, \qquad \sigma_{ab} = q_a' k_b.$$

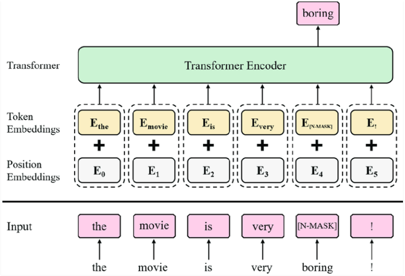
- The matrices W_Q, W_K, and W_V are learned from (lots of) data and determine which aspects of the context are important.
- Features of the full model
  - Stack multiple attention layers with multi-headed attention.
  - Add a feed-forward layer between each self-attention layer:

$$FF(x) = \max(0, xW_1 + b_1)W_2 + b_2,$$

    where the dimensionality of the inner layer $\gg \dim(x)$.
  - Add position embeddings.

BERT: Masked language modeling

- A prime example in NLP is BERT (Bidirectional Encoder Representations from Transformers).
- The model is trained via masked language modeling.



- We estimate a version of a transformer model based on the BERT architecture.
- We then estimate asset embeddings by training a sentence transformer on odd-even pairs:
  - Ownership shares ⇒ Asset embeddings.
  - Portfolio shares ⇒ Investor embeddings.

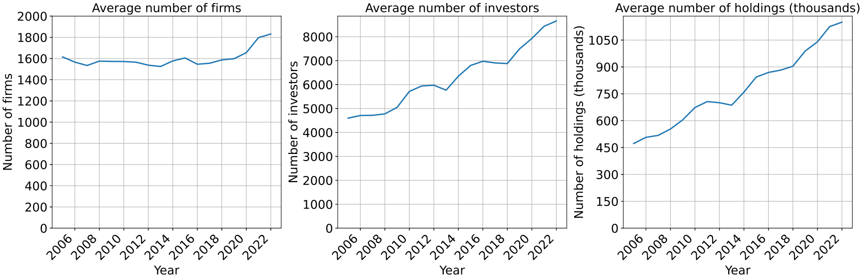
Data

- Holdings data from FactSet:
  - Hedge funds, mutual funds, ETFs, closed-end funds, variable annuity funds.
- Sample construction:
  - 2005.Q1 - 2022.Q4.
  - Remove nano and micro caps.
  - Keep investors (stocks) with at least 20 positions (investors).
- Accounting data and stock returns from CRSP / Compustat, using the Jensen, Kelly, and Pedersen (2023) construction.
- Earnings calls data from FactSet.

Representing firms: The competitors

- Observed characteristics:
  - Market cap, book-to-market, asset growth, profitability, beta, momentum.
- Holdings-based embeddings.
- LLM-based embeddings from Cohere and OpenAI.
  - Cohere:
    - Model: `embed-english-v3.0`.
    - Reduce the dimensionality using UMAP.
  - OpenAI:
    - Model: `text-embedding-3-large`.
    - Download the embeddings for the appropriate size.

Number of firms, funds, and investors



- Main takeaway:
  - The number of holdings per firm steadily increased over time.

Benchmarking models of asset embeddings

- Benchmark competitions identify the best performing models in AI and give metrics for success.
  - E.g. ImageNet to measure improvement in performance in computer vision tasks.
- Resembles the current practice of matching macro-finance moments, pricing the 25 Fama-French portfolios, ...
- However, our cross-sectional benchmarks can discriminate between models using a single quarter of data.

Three benchmarks

1. Predicting relative valuations.
   - Decompose $m_a = \beta_0 + \beta_1 b_{at} + m_a^{\perp}$.
   - Estimate $m_a^{\perp} = \gamma_0 + \gamma_1' x_a + \epsilon_a$ on 80% of the sample.
   - Evaluate using the R^2 on the remaining 20%.

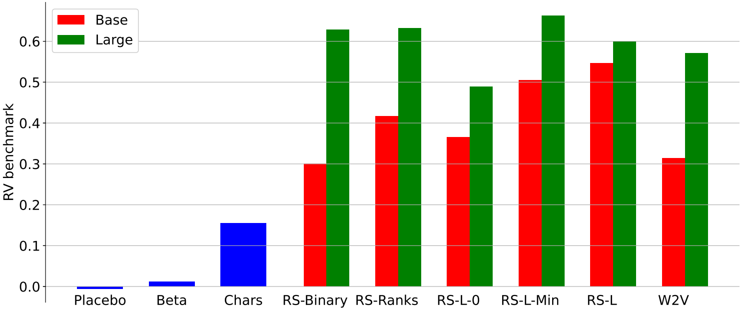
Three benchmarks

1. Predicting relative valuations.
   - Decompose $m_a = \beta_0 + \beta_1 b_{at} + m_a^{\perp}$.
   - Estimate $m_a^{\perp} = \gamma_0 + \gamma_1' x_a + \epsilon_a$ on 80% of the sample.
   - Evaluate using the R^2 on the remaining 20%.
2. Explaining comovement.
   - Estimate $r_{am} = c_m + x_{a,q-1}' f_m + \epsilon_{am}$ on 80% of the sample.
   - Evaluate using the R^2 on the remaining 20%.

Three benchmarks

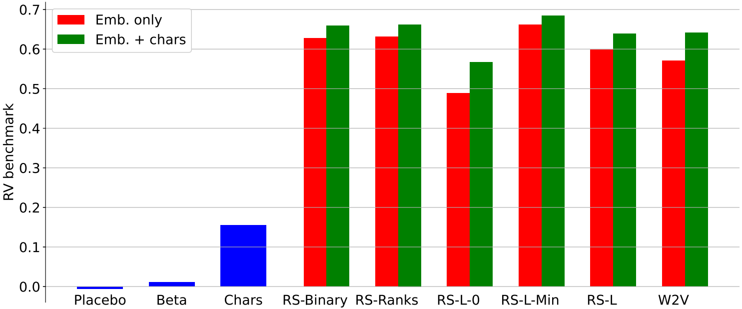
1. Predicting relative valuations.
   - Decompose $m_a = \beta_0 + \beta_1 b_{at} + m_a^{\perp}$.
   - Estimate $m_a^{\perp} = \gamma_0 + \gamma_1' x_a + \epsilon_a$ on 80% of the sample.
   - Evaluate using the R^2 on the remaining 20%.
2. Explaining comovement.
   - Estimate $r_{am} = c_m + x_{a,q-1}' f_m + \epsilon_{am}$ on 80% of the sample.
   - Evaluate using the R^2 on the remaining 20%.
3. Asset similarity in managed portfolios.
   - Mask the second position of a fund.
   - Estimate the probability of the identity of the second holding using embeddings/characteristics.

Benchmark 1: Predicting relative valuations



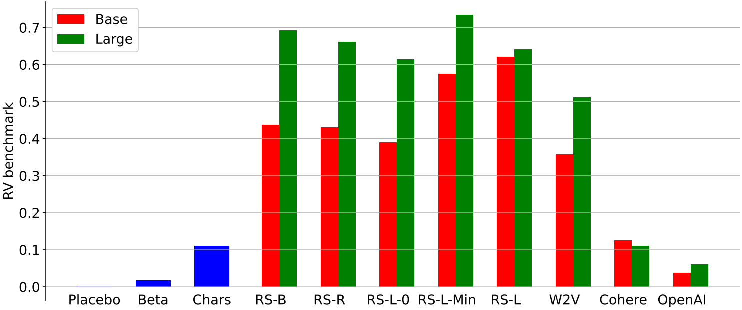
- Main takeaways:
  - Holdings-based asset embeddings perform well relative to characteristics.
  - High-dimensional models perform significantly better.

Combining embeddings and characteristics



- Main takeaway:
  - Adding characteristics to asset embeddings does not improve the benchmark much.

Text-based embeddings



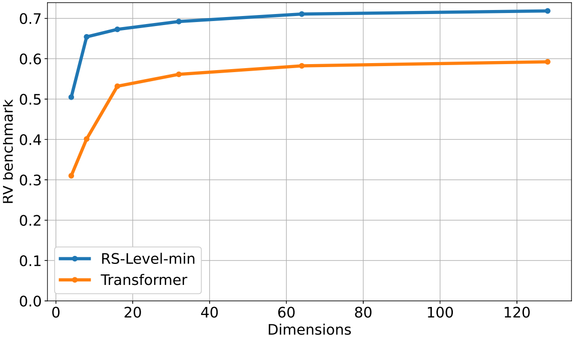
- Main takeaway:
  - Text-based asset embeddings do not perform well.

Understanding text-based embeddings

- Using OpenAI's text-based embeddings, we search for the most similar firms (using cosine similarity).
- OpenAI's embeddings mix economic and semantic similarity.

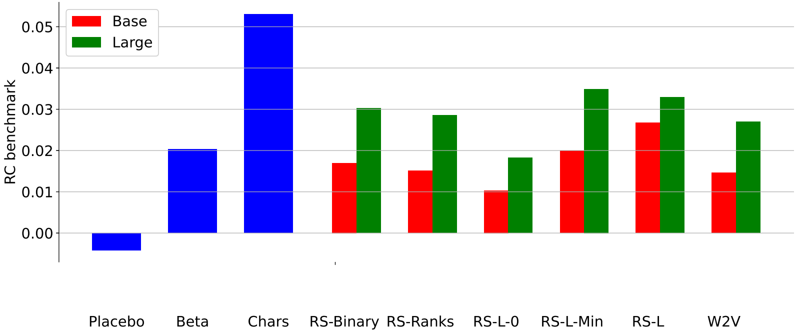
	Similar Firms as predicted by OpenAI		
Input company	Apple Inc	Citigroup Inc	Walmart Inc
Rank 1	Appian Corp	Citizens Financial Group Inc	Walgreens Boots
Rank 2	Adobe Inc	Goldman Sachs Group Inc	Home Depot Inc
Rank 3	Interdigital Inc	American International Group Inc	Murphy Usa Inc
Rank 4	Microsoft Corp	Comerica Inc	Amazon Com Inc
Rank 5	Gopro Inc	Cigna Corp New	Qurate Retail Inc
Rank 6	Netapp Inc	Capital One Financial Corp	Big Lots Inc
Rank 7	Intel Corp	Caci International Inc	Burlington Stores
Rank 8	Alphabet Inc	Capital City Bank Group	Dollar Tree Inc
Rank 9	Autodesk Inc	C N O Financial Group Inc	Nordstrom Inc
Rank 10	Appfolio Inc	Jpmorgan Chase & Co	Kohls Corp

High-dimensional embeddings

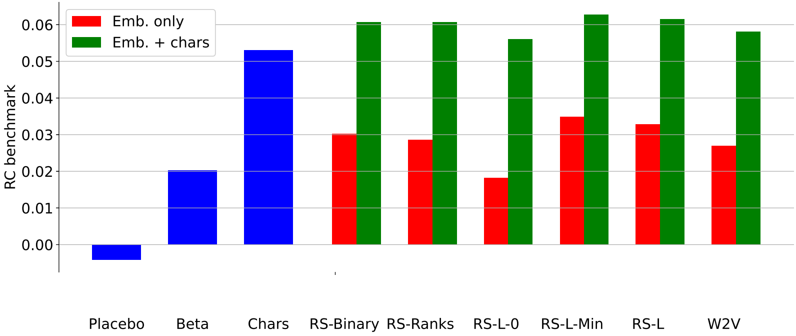


- Main takeaways:
  - High-dimensional models perform particularly well.
  - The transformer model performs well, but outperformed by the simpler recommender system.

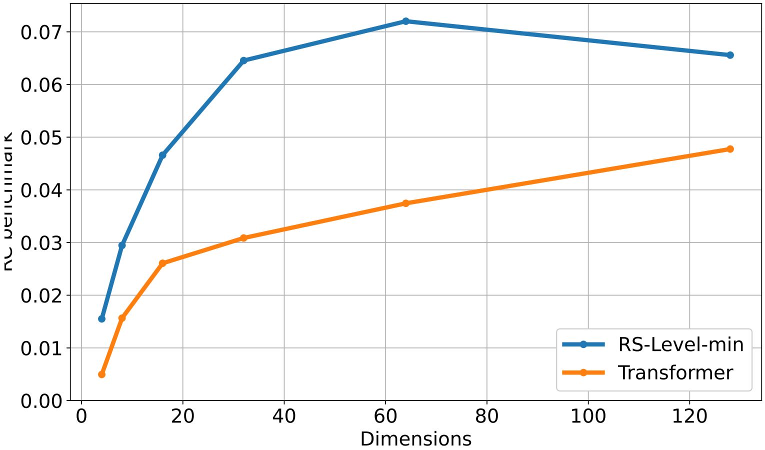
Benchmark 2: Explaining comovement



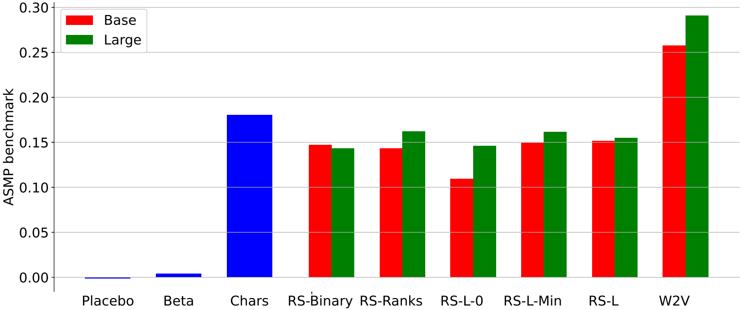
Combining embeddings and characteristics



High-dimensional embeddings

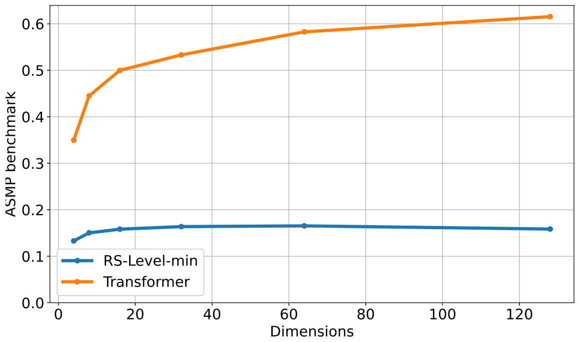


Benchmark 3: Asset similarity



- Main takeaway:
  - Word2Vec performs significantly better than recommender systems and observed characteristics.

High-dimensional embeddings



- Main takeaway:
  - The transformer model performs better than the simpler recommender system and Word2Vec on this benchmark.

Illustration for ARKK in 2022.Q4

Rank	Actual holding	Predicted holding for position 3
1	Zoom Video Communications Inc	Alphabet Inc
2	Exact Sciences Corp	Amazon Com Inc
3	[MASK]	Apple Inc
4	Roku Inc	Servicenow Inc
5	Block Inc	Adobe Inc
6	UiPath Inc	Microsoft Corp
7	Teladoc Health Inc	Advanced Micro Devices Inc
8	Twilio Inc	**Tesla Inc**
9	Beam Therapeutics Inc	Visa Inc
10	Unity Software Inc	Netflix Inc

Asset and investor similarity

- We can use the transformer model to generate similar investors and assets.
- This can be used to create generative portfolios without return data.

Apple Inc	Citigroup Inc	DFA US Small Cap Value ETF	AQR Arbitrage LLC
Tesla Inc	Altria Group Inc	Acclivity Small Cap Value	BCK Capital Management LP
Costco	Exxon Mobil Corp	Undiscovered Mgrs Behavioral Value	Water Island Capital LLC
Amazon Com Inc	AIG Inc	SEI - Small Cap Value	VIA AM SICAV - Alternative-Liquid
Microsoft Corp	Wells Fargo & Co New	SBL Series Q (Small Cap Value)	GAMCO International SICAV - Merger Arbitrage
Nike Inc	General Motors Co	Guggenheim Small Cap Value	Yakira Capital Management, Inc.
Alphabet Inc	Valero Energy Corp New	MassMutual Small Company Value	GDL
Nvidia Corp	Gilead Sciences Inc	MML Small Company Value	Pentwater Capital Management LP
Adobe Inc	Goldman Sachs Group Inc	PF Small Cap Value	Lyxor Newcits IRL - Tiedemann Arbitrage Strategy
Disney Walt Co	Bank Of America Corp	MML Small/Mid Cap Value	Gabelli & Co. Investment Advisers, Inc.

Interpreting asset and investor embeddings

- Asset embeddings yield clusters of stocks.
- We use OpenAI's `GPT-4o` model to summarize the earnings calls of groups of firms and identify
  - Main common risks.
  - Main growth opportunities.
  - ...
- To avoid generic risks, we can add a group of firms (sampled across industries) as a reference point.
- The same logic applies to investor embeddings using, e.g., information in fund prospectuses, analyst reports, et cetera.

Interpretability: 2019Q4

Based on the analysis of the earnings call transcripts for the companies listed, three common, significant risks shared by these companies are:

1. **Commodity Price Volatility:**
   - **Occidental Petroleum Corp., Marathon Petroleum Corp., EOG Resources, Inc., Pioneer Natural Resources Co., ConocoPhillips, Phillips 66, and Valero Energy Corp.**: These companies are heavily exposed to fluctuations in oil and gas prices, which can significantly impact their revenues and profitability. For instance, Occidental Petroleum and ConocoPhillips discussed the impact of oil price volatility on their financial performance and strategic decisions. Similarly, Valero Energy and Phillips 66 highlighted how refining margins and crude differentials, which are influenced by global oil prices, affect their earnings. The uncertainty in commodity prices can lead to unpredictable cash flows and necessitate adjustments in capital expenditure and operational strategies.
2. **Regulatory and Environmental Risks:**
   - **The Boeing Co., American International Group, Inc., and Valero Energy Corp.**: These companies face significant regulatory and environmental risks that can affect their operations and financial performance. Boeing, for example, is dealing with regulatory scrutiny and safety concerns related to the 737 MAX, which has led to production halts and reputational damage. American International Group (AIG) is subject to regulatory changes in the insurance industry, which can impact its product offerings and profitability. Valero Energy and other energy companies are also navigating environmental regulations, such as those related to emissions and renewable fuels, which can lead to increased compliance costs and operational adjustments.
3. **Market and Economic Uncertainty:**
   - **Discover Financial Services, The Boeing Co., and American International Group, Inc.**: These companies are exposed to broader market and economic uncertainties that can affect consumer demand, investment decisions, and financial stability. Discover Financial Services, for instance, is sensitive to changes in consumer credit behavior and economic conditions that influence loan growth and credit quality. Boeing's performance is tied to global air travel demand, which can be impacted by economic downturns and geopolitical tensions. AIG's insurance business is also influenced by economic cycles, which affect underwriting results and investment income. These uncertainties require companies to be agile in their strategic planning and risk management practices.

Overall, these risks highlight the importance of strategic flexibility, robust risk management, and proactive regulatory compliance for these companies to navigate the challenges in their respective industries.

INTERPRETABILITY: 2020Q2

The earnings call transcripts for the companies listed above reveal several common risks that these companies faced during the second quarter of 2020. Here are three of the most important shared risks:

1. **Impact of COVID-19 on Demand and Operations:**
   - **Occidental Petroleum Corp., Marathon Petroleum Corp., and Valero Energy Corp.:** These companies highlighted the significant decline in demand for oil and refined products due to the COVID-19 pandemic. Occidental Petroleum noted the unprecedented drop in oil prices and demand, which led to operational adjustments and financial measures to preserve liquidity. Marathon Petroleum and Valero Energy also discussed the impact of reduced travel and economic activity on gasoline and jet fuel demand, leading to lower refinery utilization rates.
   - **The Boeing Co.:** Boeing faced a sharp decline in demand for commercial aircraft as airlines reduced operations due to travel restrictions and decreased passenger traffic. The company had to adjust production rates and manage supply chain disruptions.
   - **Discover Financial Services and American International Group, Inc.:** These financial companies experienced challenges related to the economic downturn caused by the pandemic. Discover Financial Services noted increased credit risk and the need for higher loan loss provisions, while AIG faced potential claims related to business interruption and other insurance lines.
2. **Financial Liquidity and Capital Management:**
   - **ConocoPhillips and EOG Resources, Inc.:** Both companies emphasized the importance of maintaining strong liquidity and financial flexibility in response to the volatile oil market. ConocoPhillips highlighted its strong balance sheet and liquidity position, while EOG Resources focused on cost reductions and capital discipline to preserve cash flow.
   - **Phillips 66 and Pioneer Natural Resources Co.:** These companies took measures to conserve cash and maintain liquidity, such as reducing capital expenditures and suspending share repurchases. Phillips 66 secured additional credit facilities to enhance its financial position, while Pioneer Natural Resources emphasized its strong balance sheet and cost-cutting efforts.
   - **American International Group, Inc.:** AIG discussed its focus on liquidity and capital strength, highlighting its actions to manage financial resources prudently during the crisis.
3. **Supply Chain and Operational Disruptions:**
   - **The Boeing Co.:** Boeing faced significant supply chain disruptions due to the pandemic, affecting its production schedules and delivery timelines. The company had to work closely with suppliers to manage these challenges and ensure business continuity.
   - **Valero Energy Corp. and Marathon Petroleum Corp.:** These companies experienced operational disruptions as they adjusted refinery operations to match reduced demand. Valero Energy discussed the need to balance supply with demand to avoid inventory build-up, while Marathon Petroleum highlighted the impact of lower utilization rates on its operations.
   - **Occidental Petroleum Corp. and ConocoPhillips:** Both companies had to navigate supply chain challenges related to oilfield services and equipment availability, as well as manage production curtailments in response to market conditions.

Overall, these companies faced significant risks related to the COVID-19 pandemic's impact on demand, financial liquidity, and supply chain disruptions. Each company took specific actions to mitigate these risks and adapt to the rapidly changing environment.

Conclusions

- Recent advances in AI/ML can be applied to economics and finance via asset embeddings.
- We provide a micro foundation for using holdings data.
- We adjust methods that have been successful in related areas (e.g., NLP, vision, ...) to economics:
  - Recommender systems, Word2Vec, transformer models.
- Other asset classes: Fixed income.
  - Use embeddings to improve on ratings and distance to default to explain yields, yield volatility, and default.
  - An opportunity to redesign the architecture of fixed income markets.
- In progress:
  - Generate stress scenarios by simulating investor and asset embeddings, combined with an asset demand system (Koijen and Yogo, 2019).