AI-POWERED (FINANCE) SCHOLARSHIP

Robert Novy-Marx Mihail Z. Velikov (NBER, 2025)

Present by Li Ziming

Motivation

- Inherent tension between idealized scientific method and practical discovery.
  - First develop theories, generate predictions, and then test against data.
  - HARKing (Hypothesizing After Results are Known)
- Emergence of LLMs transform the contradiction to technological capability.
  - Improve scientific reasoning and prediction.
  - Automatically generate research ideas, conduct experiments and test hypotheses.
  - How about complete **entire academic process**?

Research Questions

- (1) How LLMs used to automatically generating academic finance papers?
- (2) What kind of threat will it pose to traditional academic norms?

Contribution

- Contribute to literature comparing AI-generated results and economic theories.
  - Prior literature: data mining has predictability comparable traditional peer review (Chen, Lopez-Lira, and Zimmermann, 2024).
  - Extend: LLM can rapidly produce coherent theoretical explanations for mined results.
- Contribute to literature on application of LLM in hypotheses generation.
  - Prior literature: reason and predict (van Inwegen et al., 2023); generate hypotheses (Manning et al., 2024); conduct experiments and produce papers (Lu et al., 2024).
  - Extend: demonstrate a **complete pipeline** for automated academic research production in finance, from hypothesis generation through full paper creation.

Data-Driven Signal Construction

- Return predictors discovery
  - Initial dataset: 31,460 accounting variables and their temporal difference.
  - Implement a series of data-quality and sufficiency **filters**.
  - Implement a series of increasingly stringent statistical validation **tests**.

Filter	# of Signals	% of Filtered
Initial set	31460	
Exclude redundant signals	29315	
Require 30 stocks	25852	
Require data until 12/2023	19834	
Require 360 months	17074	100.0%
Panel B: Cumulative significance criteria		
+ $\lvert t_{\hat{r}_{(\text{decile, name, EW})}} \rvert > 1.96$	7,102	41.6%
+ $\lvert t_{\hat{r}_{(\text{quintile, name, EW})}} \rvert > 1.96$	1,249	7.3%
+ $\lvert t_{\hat{r}_{(\text{quintile, NYSE, EW})}} \rvert > 1.96$	808	4.7%
+ $\lvert t_{\hat{r}_{(\text{quintile, NYSE, VW})}} \rvert > 1.96$	640	3.7%
+ $\lvert t_{\hat{\alpha}_{(\text{quintile, NYSE, VW})}} \rvert > 1.96$	183	1.1%
+ $\lvert t_{\text{Assay, Close Span}} \rvert > 1.96$	96	0.6%

AI-Driven Paper Generation

- Systematic Signal Naming (GPT-3.5-turbo)
  - For each validated signal, assign a descriptive and academically credible name.

Signal Naming Prompt

Create a descriptive and short name, as well as an acronym for a financial signal where the signal type is '`[signal_type]`'. Avoid using the words ratio and difference in the name. The acronym should only include capital letters.

Signal Naming Prompt

It is the ratio of '`[numer_full]`' to '`[denom_full]`'.

AI-Driven Paper Generation

- Content Generation and Structuring (Claude 3.5-Sonnet)
  - Generate each section of paper following **structured prompts**.

Finance Paper Introduction Prompt

Write an introduction for a finance academic paper discussing the signal `'[signal_name]'` that predicts stock returns. Please follow these detailed guidelines:

- 1. Motivation (2 paragraphs, ~200 words total):
  - Open with a broad statement about market efficiency or asset pricing.
  - Identify the specific gap or puzzle in the literature.
  - Use active voice and declarative statements.
- 2. Hypothesis Development (3 paragraphs, ~300 words total):
  - Present economic mechanisms linking the signal to returns.
  - Draw on established theoretical frameworks.
  - Build logical arguments step by step.
  - Support each claim with citations to foundational papers in LaTeX format.
- 3. Results Summary (3 paragraphs, ~300 words total):
  - Lead with the strongest statistical finding.
  - Present results in order of importance.
  - Use precise statistical language.
  - Include economic significance.
  - Mirror exactly the terminology used in the results section.
- 4. Contribution (3 paragraphs, ~300 words total):
  - Position relative to 3-4 most closely related papers.
  - Cite papers from: *Journal of Finance, Journal of Financial Economics, Review of Financial Studies, Journal of Accounting Research, Journal of Accounting and Economics.*
  - Highlight methodological innovations.
  - Emphasize novel findings.
  - End with broader implications.
- Use active voice (e.g., "We find" instead of "It is found").
- Maintain formal academic tone.
- Include 2-3 citations per paragraph on average in LaTeX format.

AI-Driven Paper Generation

- Content Generation and Structuring (Claude 3.5-Sonnet)
  - Generate each section of paper following **structured prompts**.

Data Section Prompt

The data section should include a description of the construction of the signal, `'[signal_name]'`, which is constructed as the ratio of COMPUSTAT variable `'[numer]'` and COMPUSTAT variable `'[denom]'`.

Our study investigates the predictive power of a financial signal derived from accounting data for cross-sectional returns, focusing specifically on the ratio of current assets to earnings before interest, taxes, depreciation, and amortization (EBITDA). We obtain accounting and financial data from COMPUSTAT, covering firm-level observations for publicly traded companies. To construct our signal, we use COMPUSTAT's item `ACT` for current assets and item `EBITDA` for earnings.

Conclusion Section Prompt

Write a conclusion for a financial research paper analyzing the signal `'[signal_name]'` in predicting stock returns. Summarize the key findings of the analysis, discussing the significance of the signal in terms of predictive power and practical implications. Conclude with suggestions for future research and limitations of this study. The conclusion should be based on the following abstract:

`'[abstract_text]'`

AI-Driven Paper Generation

- Document Assembly, Formatting, and Quality Assurance
  - Programmatically insert AI-generated sections into a standardized LaTeX template.
  - Create three versions contain different hypotheses and economic explanations.
  - Output: fully formed, academically styled PDF, suitable for journal submission.

Numerator	Denominator	Signal	Title	Citation
ACT	EBITDA	ratio	Operating Liquidity Margin and the Cross Section of Stock Returns	Harking (2024a)
AM	EBITDA	ratio	Intangibles-to-EBITDA and the Cross Section of Stock Returns	Harking (2024b)
AOLOCH	DPACT	ratio	Net Asset Impact to Depreciation and the Cross Section of Stock Returns	Harking (2024c)
AOLOCH	XINT	ratio	Growth Impact Efficiency Metric and the Cross Section of Stock Returns	Harking (2024d)
CH	EBITDA	ratio	Profitable Liquidity Score and the Cross Section of Stock Returns	Harking (2024e)
CH	EBIT	ratio	Cash Earnings Proportion and the Cross Section of Stock Returns	Harking (2024f)
CH	OIADP	ratio	Cash Profitability Index and the Cross Section of Stock Returns	Harking (2024g)
CAPS	XSGA	ratio	Efficiency of Expense Allocation and the Cross Section of Stock Returns	Harking (2024h)
ACT	NOPIO	diff	Asset Income Spread and the Cross Section of Stock Returns	Harking (2024i)
AQC	ACT	diff	Acquisitions Efficiency Ratio and the Cross Section of Stock Returns	Harking (2024j)

Results

- Efficiency of AI-Powered Research
  - Data mining and validation take about a day of computation time.
  - Final paper generation takes minutes.
- Quality of Generated Content
  - Signal names and acronyms are descriptive and creative.
  - Generated introduction, data section and conclusion similar to academic papers.
- Further Content Evaluation
  - Citation accuracy: fictitious references when citing specific or recent work.
  - Expression alignment: hypotheses, empirical results, description in each section.

Potential and Risks of AI-powered Scholarship

- Potential Benefits
  - Democratize research production: reduce barriers to entry.
  - Faster eliminate market inefficiency: quickly generate and test multiple hypotheses.
- Risks and Challenges
  - Change notion: rapid iteration between empirical findings and hypothesis development.
  - Increase p-hacking: theoretical justifications for any statistically significant pattern.
  - Cause academic arbitrage: boost citation through automated paper generation.
  - Overwhelm peer review process: determine scientific contribution become challenging.

New Ideas

- Can it pass AIGC test?
- Current limitations of AI-powered research
  - Citation hallucination, ambiguous theoretical framing, rote replication
  - How to deal with these limitations?
- Prompt engineering
  - The quality of content depends heavily on prompt, slight modifications to prompts can produce vastly different narratives.
  - How to design effective instructions for LLMs?